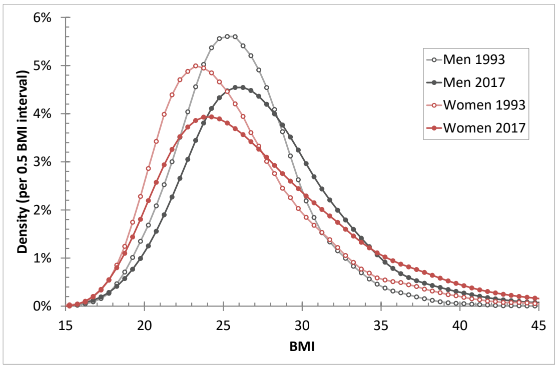
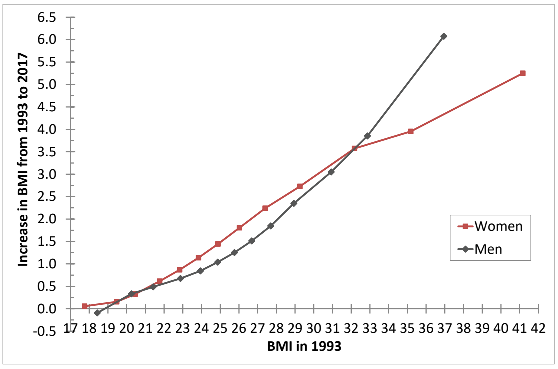
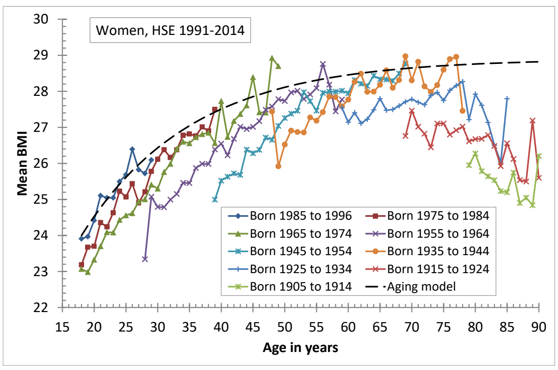
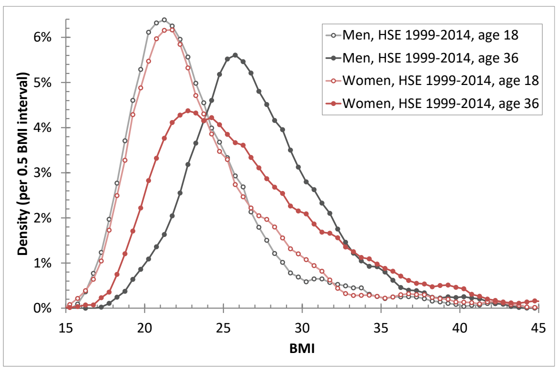
The changes in BMI and obesity prevalence for adults in England since the early 1990s

Roger J. Brooks









The changes in BMI and obesity prevalence for adults in England since the early 1990s

Roger J. Brooks

Lancaster University Management School:
Management Science Working Paper Series No. 2023:1

Department of Management Science
Lancaster University Management School
Lancaster
LA1 4YX
UK

General enquiries: msci-admin@lancaster.ac.uk
LUMS home page: www.lancaster.ac.uk/lums

About the author

Roger Brooks did a mathematics degree at Oxford University in the 1980s. He then worked as a Chartered Accountant before completing an MSc and a PhD in Operational Research at Birmingham University. After the PhD he worked for two years on the CLIVARA climate change and agriculture project at the Long Ashton Research Station in Bristol. From 1998 until 2021 he was a lecturer in the Department of Management Science at the Lancaster University Management School (LUMS). He is currently a visiting researcher in the Department of Management Science at LUMS.

Roger's general areas of research and expertise are data analytics and mathematical modelling. Past projects have included applications in health, sport, and business. His PhD and a lot of his subsequent research have been in computer simulation and he has co-authored a textbook with Professor Stewart Robinson on this topic. Other related research interests include agent-based modelling and complex adaptive systems. He has taught modules on simulation, spreadsheet modelling, operational research, statistics, accountancy, and project management. He has supervised several PhD students and many MSc summer projects, and has also carried out various consultancy assignments. Projects in the health area have included working with NHS England, the Department of Health, North West Air Ambulance, ScHARR, and hospital trusts.

Contact details

Contact details for questions or feedback on the report, or for other topics such as potential future research or consultancy projects:
Email address: brooksrogerj@gmail.com
Linkedin page: https://www.linkedin.com/in/roger-brooks-39a553a/

Report Summary

This summary sets out the main points from the report, including some of the key statistics and charts. A brief abstract is given below. The next section overleaf gives a list of bullet points of the main findings from the whole report. The sections after that in the summary give the most important points for each chapter in turn with some specific results and charts.

The selected charts included in the summary are reproduced from the main report. They keep the figure number that is used in the main report, so as to make it easy to find them in the report. The first number of a figure is the chapter number. A list of all the figures and their page numbers is given on pages xxvi - xxix.

Each chapter in the report also has a list of key points at the start of the chapter.

Abstract

Body mass index (BMI) data for men and for women of age 18 and over from the Health Survey for England (HSE) is analysed to investigate the changes in adult BMI in England since the early 1990s. A method is devised for producing estimated population BMI distributions, and these distributions give much more detailed information than just using the standard BMI categories. Topics include the changes in BMI over time, how BMI varies with age, a model for producing BMI scenarios, calorie values for the BMI changes, and how BMI varies across different categories of deprivation.

General findings include: the way that BMI has increased over time is approximately a linear scaling transformation, which means that BMI has increased much more for those with a higher BMI than those with a lower BMI; BMI increased at a faster rate in the 1990s than since then; the BMI distributions for men and women are very different, but have increased in a similar way over time; within narrow birth cohorts, mean BMI tends to increase with age for all ages with the fastest rate of increase being for the youngest ages; the age 18 distributions are very similar for men and women and are the lowest BMI values with about two thirds in the healthy BMI category; the BMI distribution increases with age in different ways for men and women; model scenarios were able to be produced for the continuation of current conditions and for returning to the lower BMI levels of the early 1990s; the increases in daily calorie amounts calculated for the BMI changes using models from the literature are quite small; BMI tends to be higher for categories of greater deprivation with the differences being particularly large for the data for women.

Main findings from the whole report

- The Health Survey for England (HSE [1]) is an excellent data source for analysing body mass index (BMI). The HSE provides sample data for England for each year since 1991. The BMI units are kg / m^2.
- A method has been devised for producing estimated population BMI distributions from the HSE data. Four years of data are used and this gives good sample sizes. The data is adjusted to have a standard age profile, and is then grouped into narrow BMI intervals of width 0.5 to give a data distribution. The data distribution is smoothed using the Savitzky-Golay method (Savitzky and Golay, 1964; Steinier et al., 1972) to give the final population distribution.
- The smoothing method is easy to apply and works extremely well in smoothing out the irregularities in the data whilst maintaining the shape of the data. The resulting BMI distributions enable detailed comparisons to be made, such as how BMI has changed over time.
- BMI population distributions for ages 18 and over were generated for men and for women for 1993, 1997, 2001, 2005, 2009, 2013, and 2017. For both men and women, BMI increases over time from one distribution to the next, with BMI increasing at a faster rate from 1993 to 2001 than since 2001. The overall changes from 1993 to 2017 are large, and changes in the statistical values for the distributions include: mean BMI increases by 1.58 for men and by 1.71 for women, the percentage in the healthy category reduces by 9.7% for men and by 10.8% for women, the percentage in the severely obese category increases from 0.3% to 2.2% for men and from 1.4% to 4.1% for women.
- For both men and women, the way that the BMI distribution changes over time between successive distributions and also overall from 1993 to 2017 is approximately a linear scaling transformation, and the changes in the distributions can be modelled well using linear scaling transformations. The pattern of the changes is that the start of the left tail of the distribution stays much the same but the rest of the distribution gets more and more stretched out to the right over time. Looking at equivalent percentile values, BMI increases more for higher BMI values than for lower BMI values in approximately a linear relationship. One implication is that the changes over time in the many factors affecting obesity have had a greater effect on those with a high BMI than on those with a low BMI.
- As stated in the previous point, the increases in the percentile values from 1993 to 2017 have approximately a linear relationship with BMI, with small increases in the lower percentiles and very large increases in the higher percentiles. For example, the 90^{th} percentiles have BMI increases of 3.05 for men and 3.57 for women. Converting these values into weight values using average height, they each correspond to weight increases of about 9.3 kg (or about 1½ stone).
- The BMI distributions have a quite different shape for men and women. The distribution for men has a much higher total percentage than the distribution for women of those who are overweight or obese, for each year about 9% higher. The distribution for women has a greater prevalence of high BMI values. Specifically, for the 2017 distributions the frequency percentage for women compared to men is 8.9% higher for BMI < 24.0, 12.8% lower for 24.0 ≤ BMI < 34.5, and 3.9% higher for BMI ≥ 34.5. Hence, the general obesity issue of being overweight or obese is more common and widespread for men, whereas very high BMI values have a greater prevalence for women. Despite the different distribution shapes, the changes over time for men and women are very similar.
- The relationship between BMI and age was examined. It was found to be very important to take account of the birth year to avoid the confounding effect of living in different times and hence different circumstances. The relationship was therefore analysed for narrow birth cohorts. Within the cohorts, mean BMI tends to increase with age for all ages. For the age range considered here of 18+, the fastest increase in BMI is for the youngest ages with the rate of increase reducing as age increases. Older cohorts who were born in earlier times are at lower levels for mean BMI, which is presumably due to very different lifetime experiences through living in a different period of time.

[1] The references for the HSE datasets are listed at the end of the references section in this report.

- BMI distributions were produced for age groups and for single years of age. The pattern of the increases over time for the different age groups is similar to that of the population as a whole, being approximately a linear scaling transformation. The changes in the BMI distributions with age have a different pattern to this and are also different for men and women. The BMI distributions for age 18 are very similar for men and women and have a high proportion in the healthy category – this is about two thirds for the most recent data analysed for 1999-2014. As age increases BMI increases. For men the BMI distribution stays a similar shape for ages of about 24 and above and mainly shifts right along the x-axis as age increases (i.e., a translation). For women the shape of the distribution changes considerably as age increases with some elements of a linear scaling transformation. The result is that the distributions for men and women are quite different shapes for older ages, with the distributions for men being fairly symmetrical and the distributions for women being much more positively skewed. For the older ages, there is a higher percentage of women than men in the healthy category but also a higher percentage of women with very high BMI values.
- A method was developed for modelling BMI and generating BMI distributions for alternative future scenarios. The basis of this is that the BMI distributions for each year of age are approximately a lognormal shape. Scenarios were developed for the continuation of current trends in mean BMI and also for returning to the better BMI distributions of the early 1990s. The latter scenario can be used to give targets to aim for, at each year of age, in order to lower BMI back to 1990s levels.
- A way was found for converting BMI values and BMI changes into calorie values using models from the literature. This gives estimated changes in daily calorie amounts. The scenarios that this was applied to include the changes in the BMI distributions over time from 1993 to 2017 and the changes in BMI with age from 18 to 24. Each of these are large changes in BMI and yet the calculated calorie values are fairly small. For example, from 1993 to 2017, for a BMI of 30 the calorie increase calculated per person for the median activity level is 141 kcal / day for men and 102 kcal / day for women. These are average population values rather than being applicable directly to individuals, and their accuracy depends partly on the validity of the models from the literature.
- BMI was compared for five categories of deprivation and for three categories of height. BMI tends to be higher for greater levels of deprivation. BMI also tends to be slightly lower for categories of taller height. The differences between BMI for the most deprived and the least deprived categories for the data for women are considerable and, by comparison, are greater than the changes in the BMI population distributions from 1993 to 2017. This includes the mean BMI for the most deprived category being 2.42 higher than for the least deprived category and the healthy BMI category percentage being 15.2% lower.
- Given suitable data, distributions can be produced fairly easily using the methods in this report. In particular, the results in the report can be extended in the future as new HSE data becomes available. New analysis could also be done such as looking at the relationship of BMI with other variables, or looking at BMI in other countries if there is similar data to the HSE.
- The BMI distributions produced in this report use 0.5 BMI intervals and give detailed information on the prevalence of different BMI values. The traditional BMI categories (underweight, healthy, overweight, obese I, obese II, severely obese) are useful, but they are limited in corresponding to very wide ranges of body weights. The more detailed BMI distributions derived here and the related analyses have led to insights into how BMI has changed over time, how BMI varies with age, and how BMI changes with other variables. Therefore, a particularly important general finding from this work is that there are great benefits in analysing BMI in detail using narrow intervals.

Aim and scope of the project (Chapter 1)

- The World Health Organisation (WHO, 2021) fact sheet web page definition of obesity is: "Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health." A statistic that is used widely as an indicator of obesity is body mass index (BMI). BMI is calculated as weight in kg (w) divided by height in metres (h) squared (i.e., BMI = w / h^2). The units for all the BMI values in this report are therefore kg / m^2.
- This report analyses the values of BMI in adults, using the age range of 18 and over, from the Health Survey for England (HSE, references given at the end of the report references section). As is usual for BMI and obesity, the data for men and for women are analysed separately since the data distributions for men and women are quite different.
- BMI values are often categorised as: underweight (BMI < 18.5), healthy (18.5 ≤ BMI < 25.0), overweight (25.0 ≤ BMI < 30.0), obese I (30.0 ≤ BMI < 35.0), obese II (35.0 ≤ BMI < 40.0), severely obese (BMI ≥ 40.0). These are broad categories. For example, the healthy category corresponds to a range of weights of 19.9 kg (44 lbs, or 3 stone 2 lbs) for a person of height 1.75 m (the mean height for men), and 17.1 kg (38 lbs, or 2 stone 10 lbs) for a person of height 1.62 m (the mean height for women).
- One of the aims of the work in this report is to analyse BMI in much more detail than the BMI categories, in particular by generating percentage frequency distributions using narrow BMI intervals. This provides greater detail on how the prevalence varies across the complete range of BMI values. It also enables other analysis to be done including more specific comparisons of how two BMI distributions differ, such as for two different years, two different ages, or for men and women. Percentile values are also calculated for the distributions and used in the comparisons.
- Statistics are still given extensively in this report for the BMI categories as these can be very helpful in seeing and understanding some of the general patterns and comparing some aspects of the distributions. They also provide familiar comparisons as they are used widely in other work on obesity.
- The overall objective of the work presented in this report is to identify the patterns of the changes over time in the BMI distributions for men and for women in England since the start of the HSE surveys in 1991. Understanding in detail what has happened is a very important foundation for other work on why BMI has changed and how to improve the situation. It can also give insights into what might happen in the future. The hope is that this work will make a contribution to the wider research and policy efforts to improve health and that these will ultimately result in greater life expectancy and better quality of life in the population.

Obtaining the BMI distributions from the HSE data (Chapters 2 and 3)

- The data used is from the Health Survey for England (HSE). This is an excellent source and gives data since 1991. The HSE dataset references for each year are listed at the end of the references section in this report.
- The approach taken here for selecting valid cases from the HSE follows that used for producing the summary statistics in the Excel files on the NHS Digital website. Checks of the data selected included reproducing some of those statistics and a good level of agreement was obtained. Analysis of missing values and of cases with a high weight gives reassurance on the comparability of the data from the different years.
- A method was devised for producing estimated BMI distributions for men and for women for the population of England from the HSE data. This has the following main steps:
    - Extract the cases with valid BMI values from the HSE data for each year
    - Combine the HSE data in groups of four years to give larger samples
    - Adjust the data so that each group has the same standard weighting by age
    - Calculate relative frequencies for the data using BMI intervals of width 0.5
    - Smooth the frequency values to produce the final estimated population BMI distribution
- The smoothing used was the Savitzky-Golay method (Savitzky and Golay, 1964; Steinier et al., 1972). This is based on fitting local polynomial curves to the data. It is very easy to implement as it just requires applying a weighted moving average. The smoothing was found to be a very suitable and effective method that maintains the general shape of the data well. Also, with the version of the method that was used, the mean, standard deviation, and skewness values for the smoothed distribution are the same as the values for the original data distribution.
- The method was applied first to the HSE data for 1991-2014 to produce population BMI distributions for men and for women for 1993, 1997, 2001, 2005, 2009, and 2013. These results are in Chapter 4. The method can also be applied in the future as further data becomes available. In particular, by the time this report was completed, the 2015-2018 HSE data was available and so the 2017 population distributions were also derived. The results for 2017 are in Chapter 14.

Analysis and comparison of population BMI distributions (Chapter 4)

The population BMI distributions produced by the method in Chapters 2 and 3 are shown below. These are Figure 4.1 for the distributions for men and Figure 4.5 for the distributions for women.

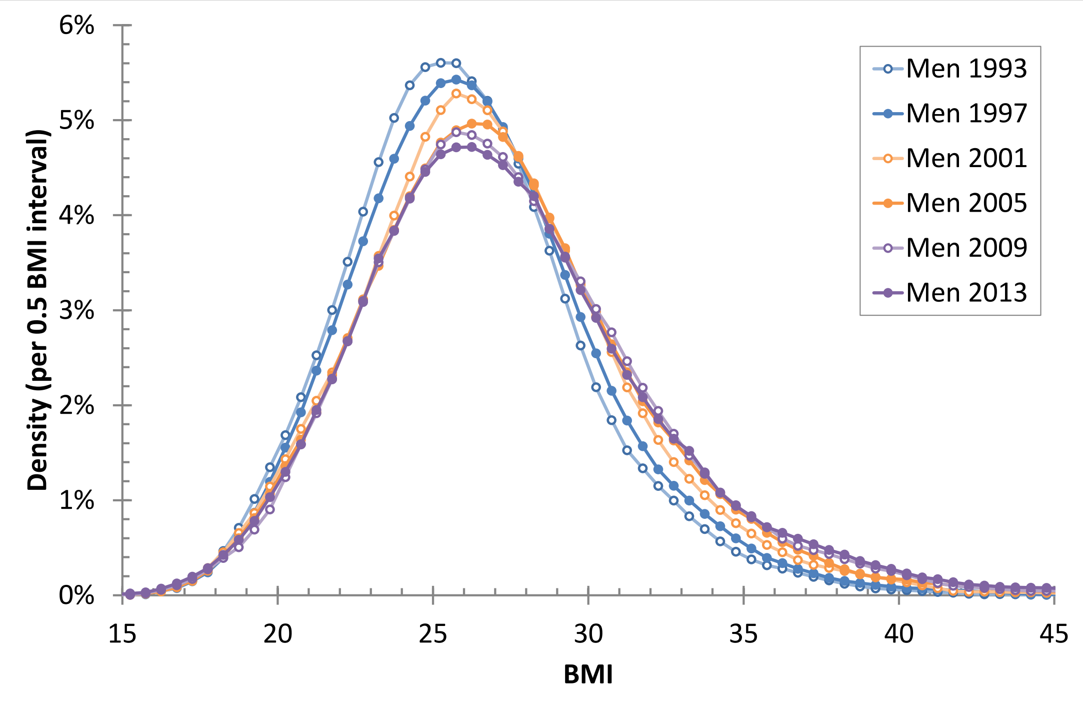


Figure 4.1 BMI population distributions for men for 1993 to 2013.

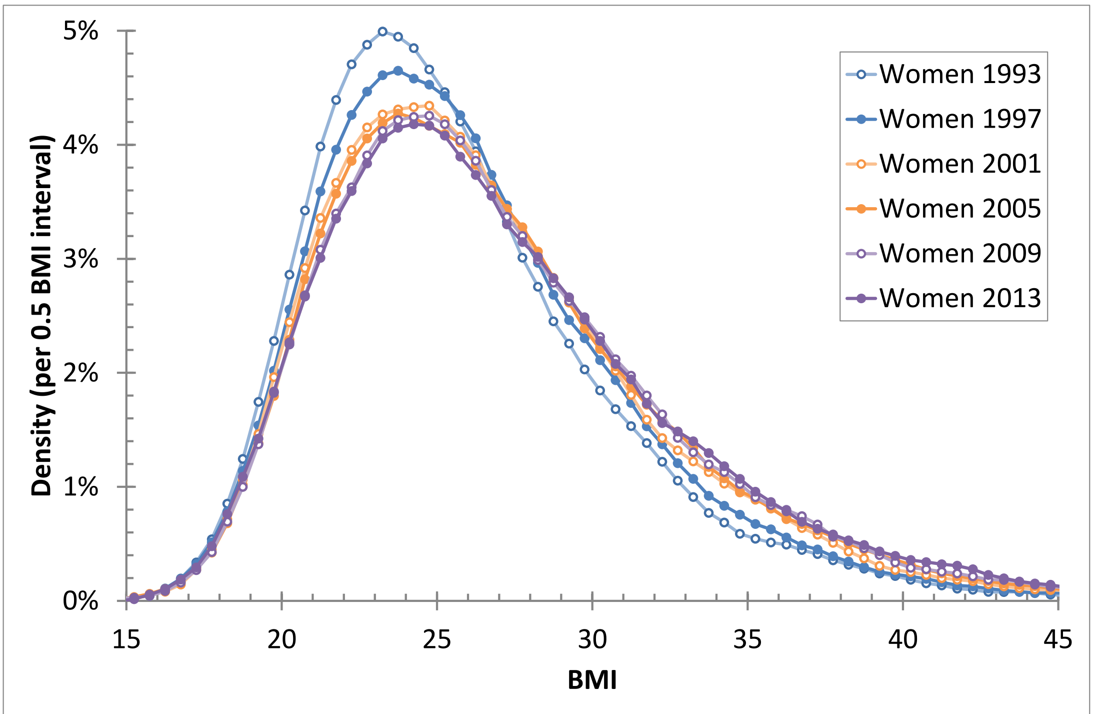


Figure 4.5 BMI population distributions for women for 1993 to 2013.

The distributions for 1993 and 2013 for men and women are shown below in Figure 4.34. The increases in the percentile values are in Figure 4.33. The distributions for men and for women in Figure 4.34 have a very different shape. The overall changes from 1993 to 2013 are considerable and, from the percentile increases in Figure 4.33, are very similar for men and women despite the different shapes of the distributions.

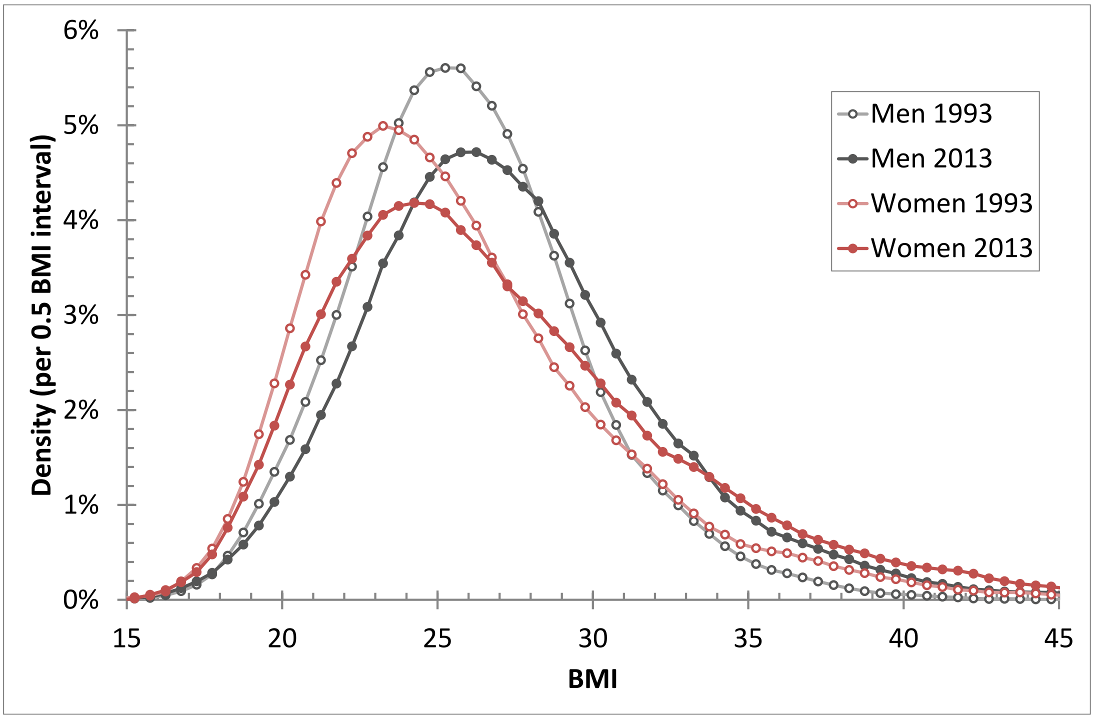


Figure 4.34 BMI distributions for men and women in 1993 and 2013.

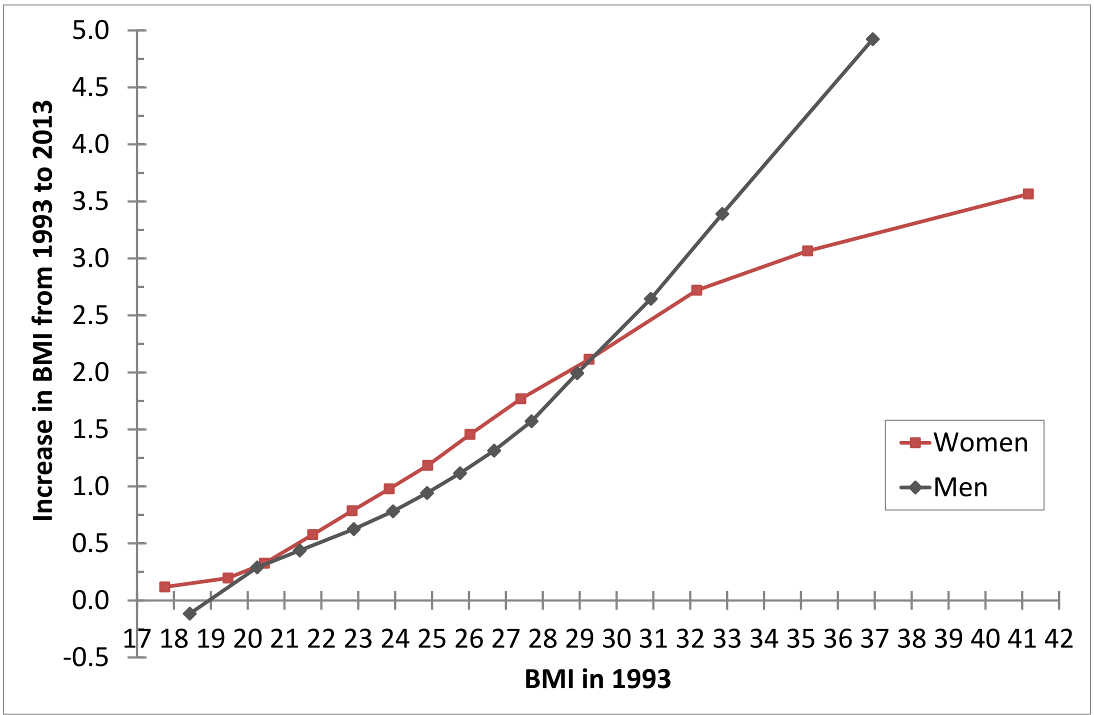


Figure 4.33 Percentile increases from 1993 to 2013 for men and women.

Chapter 4 sets out a wide range of analyses of the BMI distributions including descriptive statistics, BMI category percentages, percentiles, and high BMI values. The main findings are:

- For both men and women, the population BMI distributions are essentially a smooth curve with a single peak. They are positively skewed with a long right tail.
- For both men and women, the BMI distribution changes from one distribution to the next. The pattern of the changes in each case is for the left end of the distribution to stay basically the same and for the rest of the distribution to get more and more stretched out to the right over time. This means a worsening situation for BMI and obesity, but in a particular way where the pattern of the changes looks like a linear scaling transformation. This applies to the changes from each distribution to the next one, and also overall from 1993 to 2013.
- The increases in the percentile values (given from 1993 to 2013 in Figure 4.33 on the previous page) also indicates a linear scaling transformation. The increases are larger for higher BMI values in approximately a linear relationship. The important implication is that the changes in obesity factors over time (discussed in Chapter 6) have tended to affect the BMI of those with a higher BMI much more than those with a lower BMI.
- The overall increases in BMI from 1993 to 2013 are substantial. For men, the mean BMI increases by 1.37 from 26.05 to 27.42, and the healthy category percentage reduces by 9.1% from 40.4% to 31.3%. For women, the mean BMI increases by 1.37 from 25.76 to 27.13 and the healthy category percentage reduces by 9.4% from 49.0% to 39.6%. The changes in these statistics are very similar for men and women. The percentile increases from 1993 to 2013 are also very similar for men and women (Figure 4.33 on the previous page), and have an approximately linear pattern with higher BMI increases for higher 1993 BMI values. In particular, the higher percentiles have large increases in BMI. For example, the 70th percentile BMI increases are 1.57 for men and 1.77 for women, and the 90th percentile increases are 2.65 for men and 2.72 for women. BMI values and increases can be converted into weight values using average height and, for example, these 90th percentile BMI increases correspond to substantial weight increases of 8.10 kg (17.9 lbs) for men and 7.14 kg (15.7 lbs) for women.
- The changes from one distribution to the next for the four year distributions can be compared. For both men and women, the increases in BMI are much greater from 1993 to 1997 and from 1997 to 2001 than between the more recent distributions. BMI therefore increased at a much faster rate between 1993 and 2001 than from 2001 to 2013.
- The shapes of the BMI distributions for men and women are very different. Comparing the two distributions for any of the years, the distribution for women has a lower peak at a smaller BMI value, with higher proportions in most of the left tail and at the end of the right tail. The distributions for women have a higher standard deviation and skewness than those for men. Despite the differences, as already noted above, the changes over time are very similar for men and women.
- The different shapes mean that the distributions for women have higher percentages for low and high BMI values and lower percentages for middle values than the corresponding distributions for men. For 2013, for example, compared to the distribution for men the distribution for women has frequencies that are 9.4% higher for BMI < 24.0, 12.5% lower for 24.0 ≤ BMI < 33.5, and 3.1% higher for BMI ≥ 33.5.
- For the BMI categories, in each of the years the total percentage being overweight or obese is about 9% higher for men than for women. Therefore, in the sense of having a BMI value above the healthy category, the general issue of obesity is more widespread for men than for women. For the highest categories of obese II and severely obese, the percentages are about 3% higher for women. Hence, very high BMI values are more prevalent for women. Whilst there are these differences in prevalence, there still considerable numbers in all the overweight and obese categories for both men and women.

Modelling BMI changes with linear scaling transformations (Chapter 5)

- The results in Chapter 4 indicate that the changes in the BMI distributions over time look like a linear scaling transformation. To test this further, linear scaling transformations were used to model the changes from one distribution to the next for the BMI distributions presented in Chapter 4. The basic approach taken was to apply the scaling to each distribution to model the next distribution four years later as closely as possible. Excel solver was used to find the scaling starting point and the scaling factor that gives the best match.
- The scaling transformations are able to model the changes well and give a close match to the target distributions. This shows that the way that BMI in the population has changed in each four year time interval is similar to a linear scaling transformation.
- The magnitude of the changes over time naturally follows the analysis of the distributions in Chapter 4 with the scaling models from one distribution to the next having much greater increases in BMI between 1993 and 2001 than since 2001. The changes for men and for women are similar.
- The scaling transformations from each distribution to the next one were combined to model the total change over the whole time period from 1993 to 2013. The resulting "combined model" distributions are shown below in Figure 5.12. These work well in modelling the changes and matching the 2013 distributions. The BMI increases for the combined model from 1993 to 2013 are similar to the percentile increases given in Chapter 4 in Figure 4.33 (shown two pages before on page vii). A single linear scaling transformation, and other similar models, were also applied to the changes from 1993 to 2013 and these all give similar results and match the 2013 distributions well.

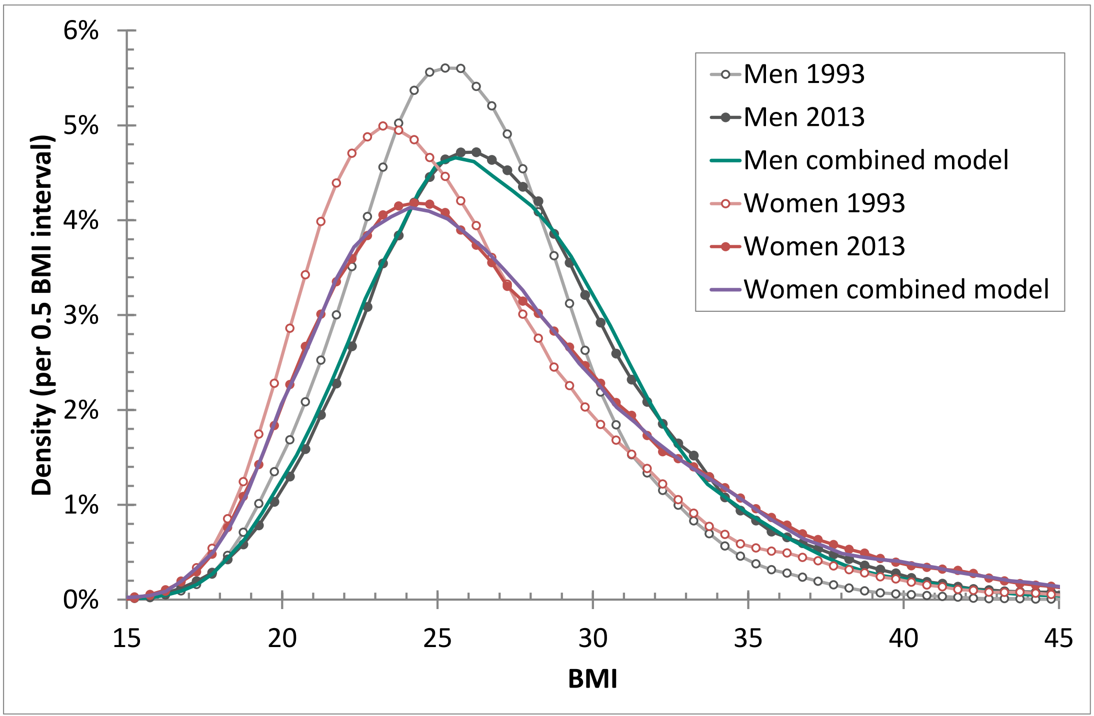


Figure 5.12 Combined scaling models for 1993 to 2013 for men and women.

Interpretation of the changes in the BMI distributions (Chapter 6)

- The changes in the population BMI distributions over time represent how BMI is different in equivalent people. In other words, how BMI is different in people of the same age and similar characteristics but living in a different time.
- The difference in the population at one year compared to another year is therefore that the people in the two populations have lived in different time periods. This results in altered lifestyles and an experience of a different range of conditions over their lives. In particular, the assumption is that the 2013 population would have had the 1993 BMI distribution if they had all lived 20 years earlier.
- Lifestyle factors that affect BMI and have changed over time will result in the population BMI changing over time. Chapter 6 provides a suggested list of many potential factors, including changes in the food industry, work, leisure, income, costs, and society. One implication of the results from Chapters 4 and 5 is that changes in these factors have tended to result in a greater increase in the BMI of those with a higher BMI than the BMI of those with a lower BMI.

Variation of mean BMI with age (Chapter 7)

- The focus of the analysis in Chapters 7-10 is to look in more detail at population BMI by investigating how BMI varies with age. Chapter 7 considers the relationship between mean BMI and age.
- To understand how BMI varies with age it is important to divide the population into narrow birth cohorts. People within the cohorts have lived in similar times, and so the variation of BMI with age is then not confounded by the effect of living in very different conditions.
- The mean BMI values for the cohorts are shown overleaf for men in Figure 7.5 and for women in Figure 7.6. The pattern within each cohort is that BMI tends to increase with age for all ages. Older cohorts are at lower levels, presumably because of quite different experiences earlier in life. The increase of BMI with age is greatest at the youngest ages with the rate of increase tending to get smaller with age.
- An equation for a BMI aging model is proposed based on the data analysis. This is subjective but is based on extending the BMI values for the most recent youngest cohorts using the general shape of the patterns in the older cohorts. This has BMI increasing towards a horizontal asymptote. The aging model curves are shown on Figures 7.5 and 7.6 overleaf as the dashed lines. The hypothesis is that this pattern is the current relationship of mean BMI with age in modern conditions. If this is the case and conditions stay the same then this is the path that the youngest cohorts will follow.
- It is not considered that the BMI aging model inevitably has a particular shape. The variation of BMI with age may well have been quite different in the past. It could be different in the future if the situation changes through health policy interventions or through changes in some of the obesity factors.

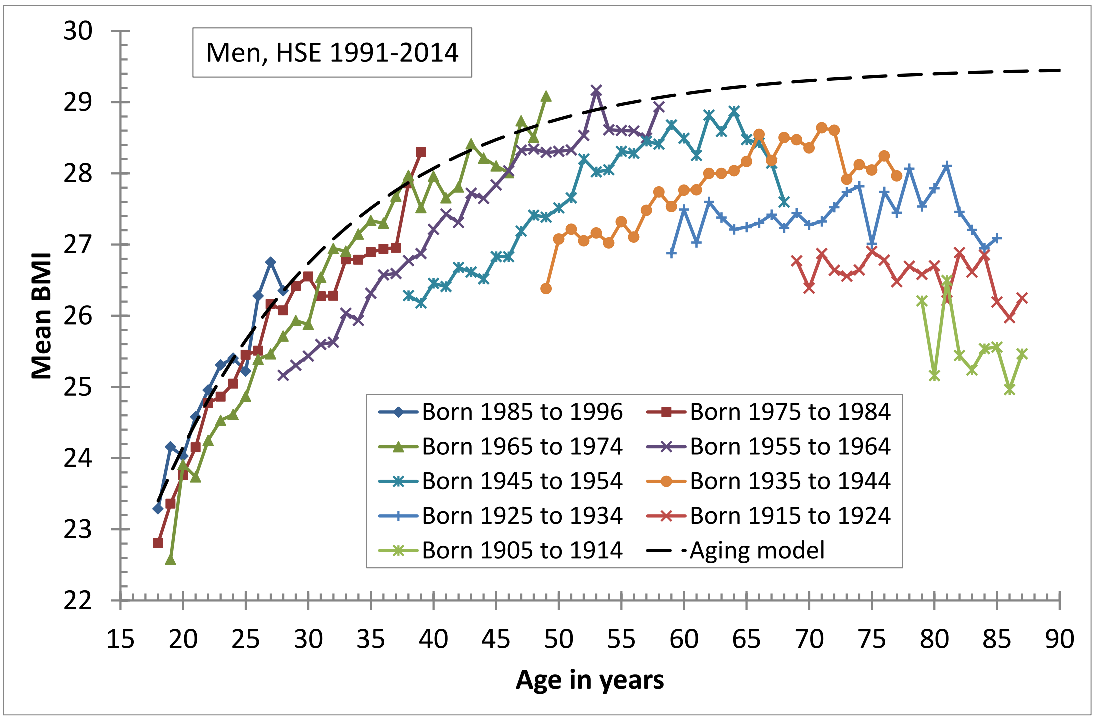


Figure 7.5 Cohort series for mean BMI for men with the aging model curve.

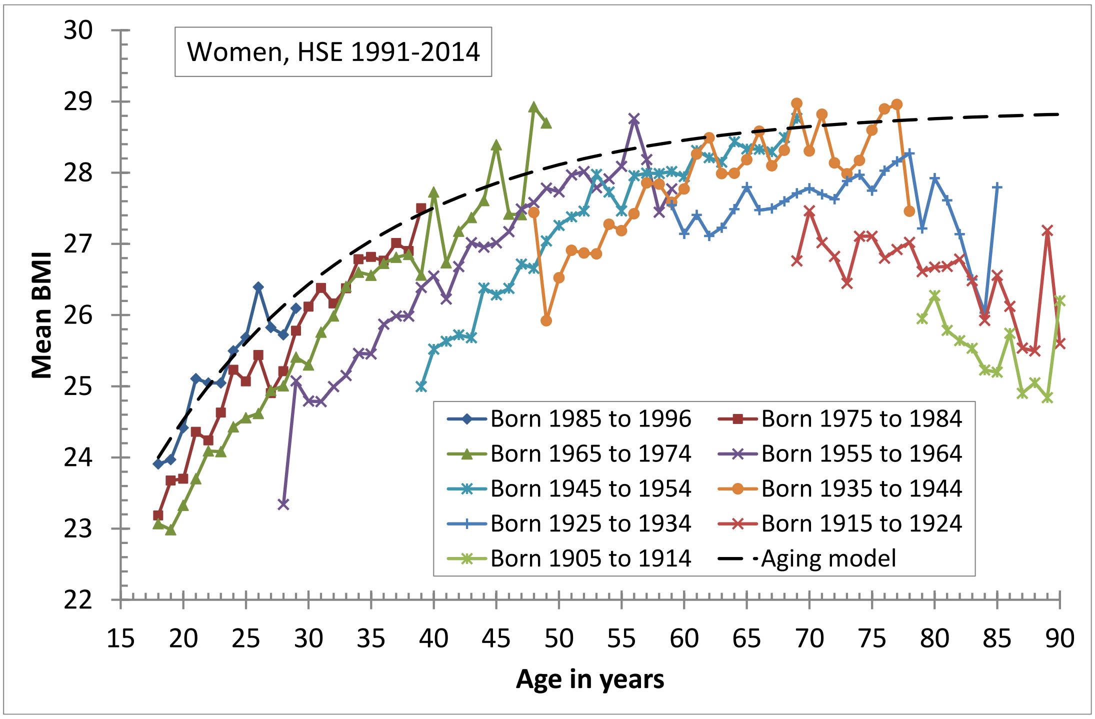


Figure 7.6 Cohort series for mean BMI for women with the aging model curve.

Detailed analysis by age for men and for women (Chapters 8-10)

- BMI distributions were produced for men and for women for age groups of about 10 years and for single years of age using the approach from Chapters 2 and 3. Wider intervals of years of HSE data were used here to give larger sample sizes. The distributions for the age groups were calculated from eight years of HSE data (1991-1998, 1999-2006, 2007-2014), while just two time intervals were used for the single year age distributions (1991-1998, 1999-2014). The distributions were compared over time as to how they change between the different HSE time intervals, and how they change with age (using the data for the most recent HSE time interval).
- The changes in the BMI distributions over time for men and women are approximately a linear scaling transformation. This is consistent with what was found for the population as a whole in Chapter 5. As discussed in Chapter 5, this implies that those with a higher BMI have been affected more than those with a lower BMI by the changes that have occurred in the obesity factors. The finding here is that this applies for most ages.
- For men, the changes in the BMI distributions with age have a different pattern to the changes over time. There is some change in the shape of the distribution from age 18 to 24, with elements of a linear scaling. However, above age 24, the shape stays much the same and BMI increases mainly by the distribution shifting up the x-axis (i.e., a translation). This pattern implies that, apart from the youngest ages, the tendency to increase BMI with age for men is about the same for all BMI levels.
- For women, the changes in the distributions as age increases are mainly a linear scaling and so the shape of the distribution changes. There is also some element of a shift along the x-axis to the right. The tendency is for BMI to increase for all BMI values but with greater increases for higher BMI values. This is a different pattern to the changes with age for men.
- The BMI distributions for age 18 are much healthier than for older adults. There are 67% of men and 64% of women in the healthy BMI category for the age 18 distributions for the 1999-2014 data. The mean values are 23.24 for men and 23.81 for women. The values are even better for the 1991-1998 data: for men 78% in the healthy category and a mean of 22.64, for women 70% in the healthy category and a mean of 23.02. BMI increases rapidly from age 18 with each year of age. This indicates that targeting health messages and interventions at those of about age 18 could be a particularly effective and important strategy.
- The age 18 distributions do have a sizable percentage in the underweight BMI category. For the 1999-2014 data this is 7.1% for men and 6.4% for women. By age 24 these percentages have reduced to 2.5% and 3.9%, respectively. The issue of underweight is not a focus of this report as it concentrates on overweight and obesity, but it could be considered in other work and in health policy.
- The BMI distributions at age 18 are very similar for men and women, with much the same shape and location. The distributions for men and women become quite different with increasing age. In the ages of 30s, 40s and 50s, the distributions for men are more symmetrical than those for women, with a peak further to the right and a lower right tail. These different shapes mean that the healthy percentages are much higher for women than for men. The overweight percentages are higher for men, whereas very high BMI values are more prevalent for women. These effects can be seen in Figure 10.9 overleaf which shows the distributions for ages 18 and 36 for the 1999-2014 data: the distributions for men and women for age 18 are much the same whereas the distributions for age 36 are quite different. For the age 36 distributions in Figure 10.9, compared to the distribution for men the distribution for women has frequencies that are 16.0% higher for BMI < 24.0, 20.1% lower for $24.0 \leq BMI < 33.0$, and 4.1% higher for BMI ≥ 33.0.
- There may be implications for health interventions in the different patterns of the changes with age. A broad strategy may be better for men since BMI tends to increase with age by a similar amount for all BMI values. A more targeted strategy may be better for women since BMI tends to increase with age more for higher BMI values than for lower BMI values.

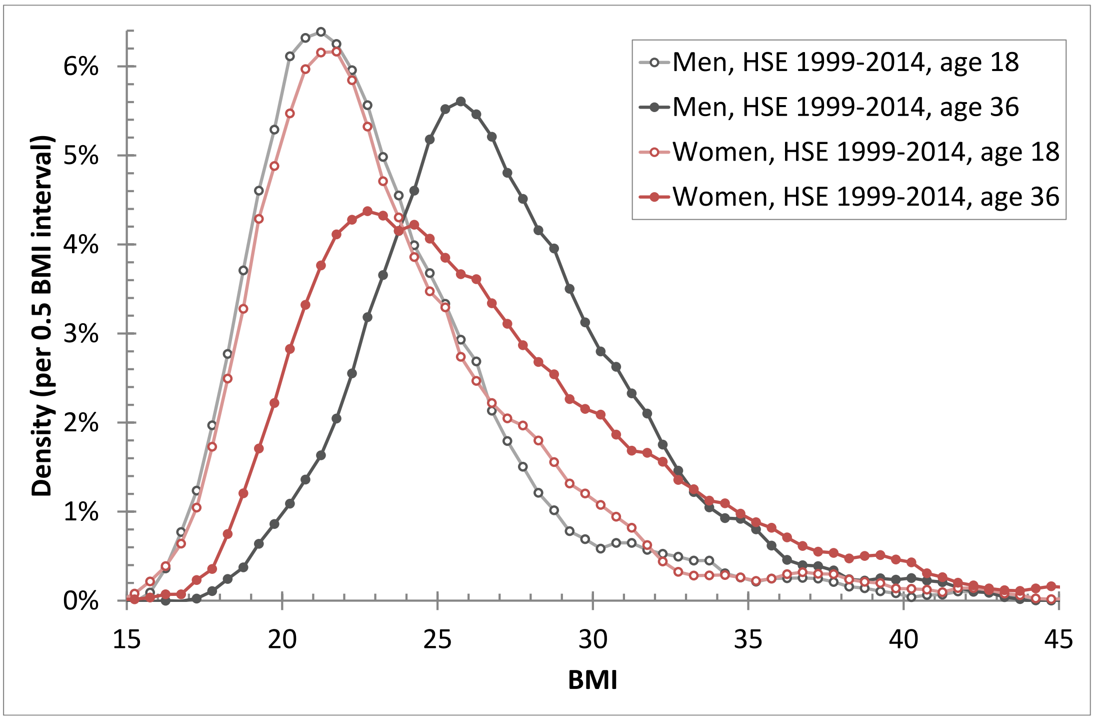


Figure 10.9 BMI distributions for age 18 and 36 for men and women for 1999-2014.

Modelling BMI scenarios (Chapters 11 and 12)

- A method is developed for modelling population BMI scenarios. This uses a BMI distribution for each year of age and combines them together into the BMI distribution for the whole population. The distribution used for each year of age is the lognormal. There is also a small cosine function adjustment for the distribution for men.
- The model requires parameter values for the mean, standard deviation, and minimum value, for each year of age. The version for men needs two further parameter values for the cosine function.
- The model is able to match the past population distributions well. This gives confidence in using the model for generating future scenarios.
- The model was applied to generate two example future scenarios. A scenario is specified by values for the mean, standard deviation, and minimum for each year of age. Parameter values for the cosine adjustment are also required for the model for men. These are scenarios rather than specific forecasts. They apply a given pattern of parameter values, particularly the mean BMI value for each year of age, to produce an overall population BMI distribution. The purpose of the scenarios is to calculate the BMI distribution that would result from a particular set of parameter values, given the assumptions made in the model structure.
- One scenario that was investigated is a "current trend" scenario, which applies the proposed BMI aging model curves for recent modern conditions from Chapter 7 (the dashed lines on Figures 7.5 and 7.6 on page xi). The aging model is an estimate of the mean BMI values if conditions stay the same, and the current trend scenario produces the BMI distributions for this situation. The aging model curves are slightly above the recent actual data used for producing the 2013 BMI distributions. Consequently, the scenario results are slightly worse than the 2013 BMI distributions, with mean BMI values about 0.4 higher and healthy category percentages about 3% lower for both men and women.
- The other scenario that was examined is a "low BMI" scenario. The aim with this one was to apply a suitable mean value curve of the type from Chapter 7 to generate population BMI distributions similar to those for 1993. The scenario therefore considers what is required to get back to the much better BMI levels of 1993. Very similar distributions to 1993 were produced. The mean and standard deviation values for this scenario could therefore be used as target values for each age. The benefits would be considerable. For example, the healthy category percentages in this scenario are 8.5% higher than 2013 for men, and 8.3% higher for women. There are considerable reductions in each of the obese category percentages.

Converting BMI changes into calories (Chapter 13)

- Energy balance equations can be applied to convert BMI values and BMI changes into calories. These equations work on the basis of equilibrium with energy intake equalling energy expenditure. If energy intake changes, the weight of the person changes until a new equilibrium is reached. The Henry (2005) equations, recommended in a 2012 report by The Scientific Advisory Committee for Nutrition (SACN, 2012), are used here. The equations can be applied to any of the BMI values or BMI changes in this report. Calories values are calculated for two examples from earlier chapters in the report. The results are population level average values and so are not applicable for individuals. The validity of the values produced depend, of course, on the how realistic the equations are.
- In the first example in Figure 13.1 below, the calories are calculated for the increase in BMI from 1993 to 2013 using the combined scaling model from Chapter 5. These are given for three levels of physical activity level (PAL). The effect of applying the energy balance equations is to multiply each BMI value by a constant to convert to calories. The general pattern is therefore the same as that of the BMI increases from the scaling model, with the increase in calories being higher for higher starting BMI values. The value used for the constant is different for each PAL level and for men and for women (i.e., it is different for each series in Figure 13.1). For 1993 BMI values between 25 and 30 (values above the healthy range with high 1993 prevalence), the extra calories calculated per person for the median activity level vary from about 50 to 120 kcal / day for men, and from about 40 to 80 kcal / day for women. These are fairly small amounts of calories and yet the changes in the BMI distributions are considerable.

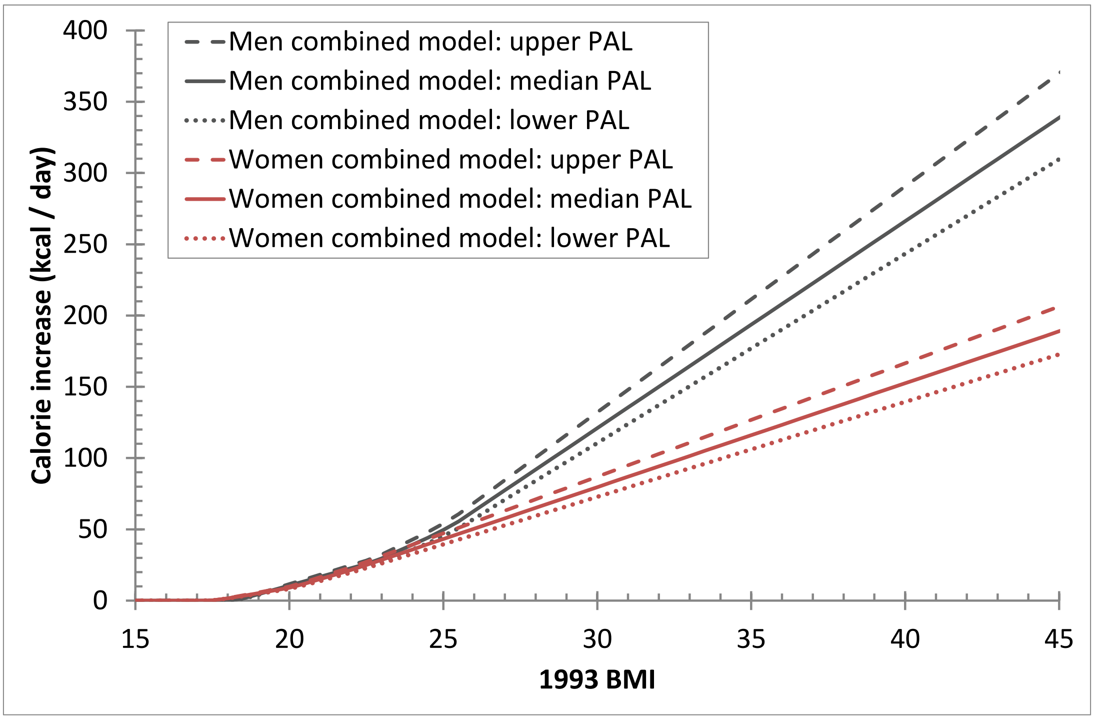


Figure 13.1 Changes in calories for the BMI changes of the combined model from 1993 to 2013

- In the second example, the increase in BMI from age 18 to 24 is converted into calories. This is the age range with the greatest increase in BMI with age. This is based on BMI increases for the percentile values from 5% to 95%. The calories amounts are again reasonably small. Most of the values calculated for men for the median activity level are about 170 kcal / day. The values for women increase with higher initial BMI values, up to a maximum of about 140 kcal / day.

BMI distributions for 2017 (Chapter 14)

- For the methodology in this report the BMI distributions for 2017 require the HSE data for 2015-2018. During the project this data became available, and this enabled the 2017 population distributions to be produced. Some of other analysis from earlier chapters was also updated.
- The general patterns of the BMI distributions for 2017 are consistent with the earlier results. BMI increases from 2013 to 2017, and the increase is approximately a linear scaling transformation. In the analysis of the earlier distributions the BMI increases tend to get smaller over time with the changes from 2009 to 2013 being the smallest change. This does not continue as the increase in BMI from 2013 to 2017 is greater than from 2009 to 2013. For men the increase is very similar to that from 2005 to 2009. The increase is greater for women than for men and is bigger than for each successive change from 2001 to 2013, although not as great as from 1993 to 1997 or 1997 to 2001. There is a considerable increase in the prevalence of very high BMI values for both men and women.
- The overall increase in BMI from 1993 to 2017 is large. Overleaf, Figure 14.13 shows the BMI distributions for 1993 and 2017 and Figure 14.14 shows the percentile increases from 1993 to 2017. The mean BMI increases by 1.58 for men and 1.71 for women. The healthy category percentage reduces by 9.7% for men and by 10.8% for women. Looking at the highest BMI category, the severely obese percentages for men and women in 2017 are 2.2% and 4.1% compared to just 0.3% and 1.4% in 1993 (and 1.7% and 3.4% in 2013). The relationship of the increases in the BMI values of the percentiles against 1993 BMI in Figure 14.14 is approximately linear with small increases in the lower percentiles but very large increases in the higher percentiles. For example, the 90th percentiles have BMI increases of 3.05 for men and 3.57 for women.
- Some of the BMI changes from 1993 to 2017 were converted into weight values using average height. The increases in mean BMI correspond to weight increases of 4.83 kg (10.7 lbs) for men and 4.48 kg (9.9 lbs) for women. The weight values for the percentile increases were also calculated and give large values for the high percentiles. For example, the values for the 90th percentile for men and women are each about 9.3 kg (or about 1½ stone).
- As for the previous distributions, there is a big difference in the shape of the BMI distributions for men and women. Specifically, for the 2017 distributions the frequency percentage for women compared to men is 8.9% higher for BMI < 24.0, 12.8% lower for 24.0 ≤ BMI < 34.5, and 3.9% higher for BMI ≥ 34.5. For the BMI categories, this means a higher percentage of men being overweight or obese (total of 68.2% for men and 59.8% for women), but a higher percentage for women of the high BMI values in the obese II and severely obese categories (total of 7.6% for men and 11.5% for women). Hence, as commented on in Chapter 4, the general issue of being overweight or obese is more widespread for men. The narrower issue of very high BMI values is more prevalent for women.
- The age group analysis shows particularly large increases in BMI for the 18-24 age group for both men and women. This is a concern as to what the future trajectory will be for this age group and subsequent cohorts.
- The 2017 distributions have got closer than expected to the current trend models from Chapter 12. In particular, the distributions for both men and women have a higher severely obese percentage than the model, and this also applies to the obese II category for women. One reason for this is that the current trend model uses standard deviation values based on the 2011-2014 data, but the standard deviation values for the 2015-2018 data are higher than this. Adjusting the standard deviation to 2015-2018 values brings the right tail of the model to a similar or slightly higher level than the data, and this may be a better estimate of future BMI if conditions stay the same as they were during these years.
- The calorie amounts corresponding to the changes in BMI since 1993 from Chapter 13 were updated to use the results for the 2017 distributions rather than the 2013 distributions. The general pattern and the values are similar to those in Chapter 13, but are naturally slightly higher as a result of the

further BMI increases from 2013 to 2017. The calorie increase amounts are still fairly small, indicating than such levels of calorie increase can result in large increases in BMI at a population level. For example, for a BMI of 30 the calorie increase calculated for the median activity level is 141 kcal / day for men and 102 kcal / day for women. These are average long term population values rather than individual values and depend on the accuracy and validity of the energy balance equations used, as discussed in Chapter 13.

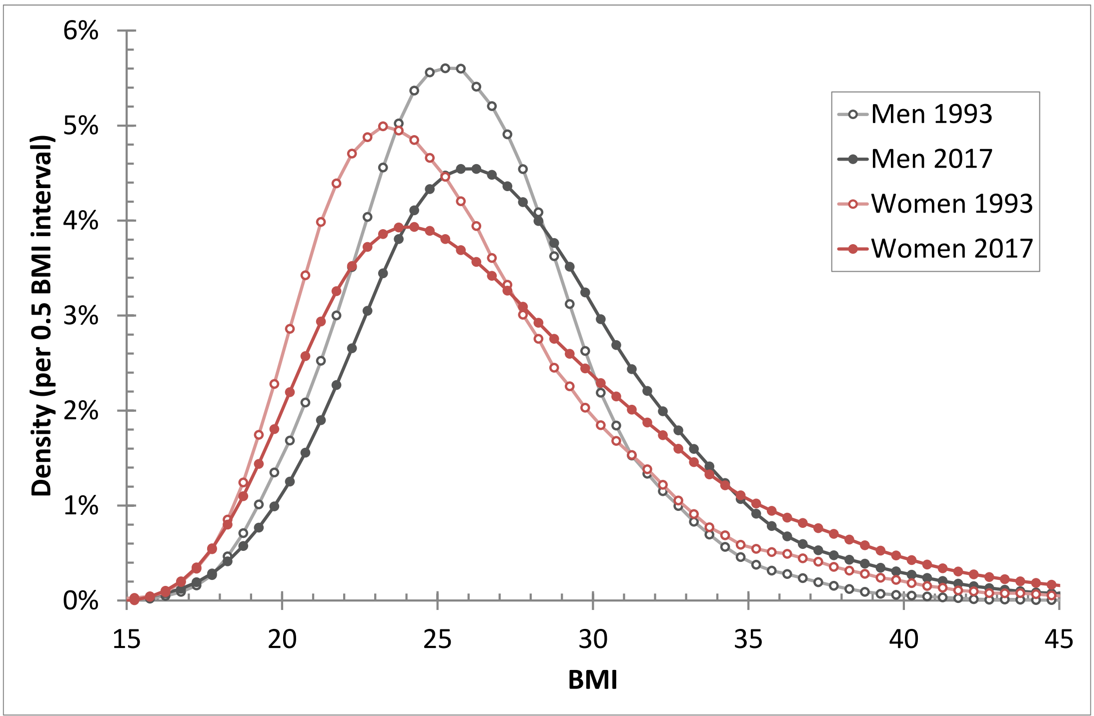


Figure 14.13 BMI distributions for men and women in 1993 and 2017.

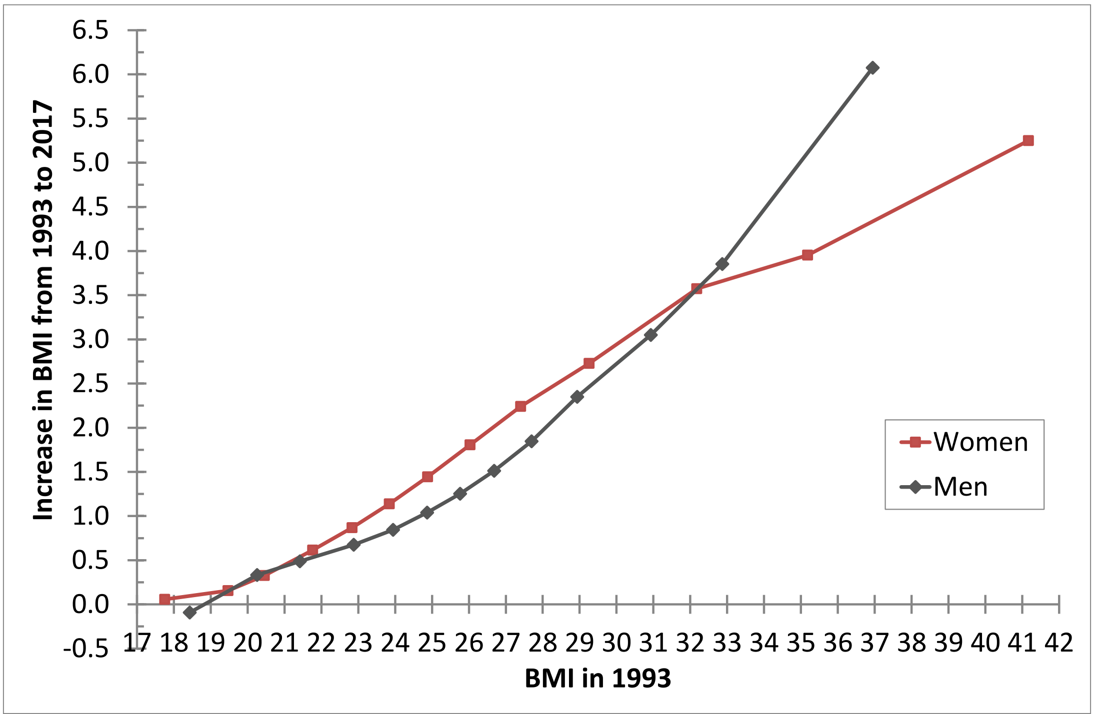


Figure 14.14 Percentile increases from 1993 to 2017 for men and women.

Relationship of BMI with deprivation and with height (Chapter 15)

- The relationship was examined between BMI and different categories of deprivation and different categories of height. The reason for this analysis is partly due to interest in the relationship of BMI with these variables and partly illustrative to show how the methodology in this report can be applied to look at other variables.
- The same approach as in the rest of the report was applied to produce estimated population BMI distributions for each category of the variables using four years of HSE data. Distributions at different points in time were compared, mainly using the changes in the percentile values. For deprivation, the HSE data since 2001 categorises cases by five levels of deprivation. Distributions were produced for 2003 (using 2001-2004 data) and for 2017 (using 2015-2018 data). For height, distributions were produced for 1993 and 2017.
- For 2017, BMI is higher for higher levels of deprivation (i.e., living in more deprived areas) with a larger mean and a greater prevalence of obese values of BMI. The differences between the categories are much greater for women than for men. Comparing the distributions for the most deprived and least deprived categories for 2017, for men the mean BMI is 0.88 higher and the healthy category percentage 3.7% lower, for women the mean BMI is 2.42 higher and the healthy percentage 15.2% lower. These differences for women are very large and, by comparison, are greater than the differences over time for the whole population between 1993 and 2017 in Chapter 14.
- For 2003, there is not much difference between the distributions for men for the deprivation categories. The mean values are all about the same. The distribution for the most deprived category actually has the highest healthy category percentage, although also the highest severely obese category percentage. For women, there are substantial differences in the distributions with the most deprived category having a mean BMI that is 1.35 higher than for the least deprived category and a healthy category percentage that is 10.2% lower. These are substantial differences but not as great as for 2017.
- As found elsewhere in the report, the changes in the percentile values over time from 2003 to 2017 for each of the deprivation categories are approximately a linear pattern with higher increases for higher values of BMI. This indicates that the changes in the distributions over time are approximately a linear scaling transformation. The increases tend to be greater for the higher levels of deprivation and so the differences between the deprivation categories have increased over time.
- For height, the cases were split into three categories. BMI tends to be slightly lower as height increases, with the differences being greater for women than for men. Distributions were produced for 1993 and 2017, and the changes over time from 1993 to 2017 are again approximately a linear scaling pattern.

Contributions and conclusions (Chapter 16)

- The contributions of the work in this report include:
    - Developing a method for deriving population BMI distributions from the HSE sample data.
    - Identifying that the changes over time since the early 1990s are approximately a linear scaling pattern.
    - Comparing the rates of increase in BMI over time.
    - Discussing the interpretation and possible causes of the changes in BMI.
    - Deriving a BMI aging profile from cohort analysis of the data.
    - Showing how the BMI distributions change with age.
    - Modelling BMI scenarios.
    - Converting BMI changes into calorie values.
    - Showing how the method can be applied to the relationship with other variables such as deprivation.
- Implications of the increases in BMI over time include an expected increase in the prevalence of obesity related health conditions. This is likely to result in reduced quality of life for some in the population, and also increased health costs. It could be a factor that causes a significant reduction in average life expectancy.
- There is various future work that could be done. This includes extending the results from the report as new HSE data becomes available. If the same methodology is followed, the next population BMI distributions for 2021 would be based on the HSE data for 2019-2022. However, the HSE survey could not be completed in 2020 because of the Covid-19 pandemic and so there is no data for that year. The 2021 distributions could still be produced from three years of data for 2019, 2021, and 2022 once the 2022 HSE data is available, which will probably be at the end of 2023. Alternatively, the next distribution could be produced from four years of HSE data for 2019 and 2021-2023, or from a different choice of years of HSE data.

Contents

List of Tables

List of Figures

Acknowledgements

This work has been helped by discussions with colleagues at Lancaster University and staff at the Department of Health, Public Health England, and the National Centre for Social Research (NatCen). These include Peter Dick, Vicky Copley, Marianne Scholes, Laura Brown, Shaun Scholes, David Worthington, and Jennifer Logue.

Some of the initial analysis in this report was presented in conference presentations at the Operational Research OR58 Annual Conference in Portsmouth in 2016 and the Operational Research Applied to Health Services (ORAHS) conference in Bath in 2017. These two events had presentations only, with no conference papers.

Chapter 1: Introduction

Key points from this chapter:

- The World Health Organisation (WHO, 2021) definition of obesity is: "Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health."
- A statistic that is used widely as an indicator of obesity is body mass index (BMI). BMI is calculated as weight in kg divided by height in metres squared. The units are therefore kg / m^2. BMI is an indirect method of assessing obesity since it does not measure fat levels directly. It therefore has limitations but it is considered a useful statistic for estimating obesity levels in populations. BMI is used very widely in academic research and in health policy. This report analyses the values of BMI in adults from the Health Survey for England (HSE). The age range used is 18 and over.
- BMI values are often categorised as: underweight (BMI < 18.5), healthy (18.5 ≤ BMI < 25.0), overweight (25.0 ≤ BMI < 30.0), obese I (30.0 ≤ BMI < 35.0), obese II (35.0 ≤ BMI < 40.0), severely obese (BMI ≥ 40.0). These are broad categories. For example, the healthy category corresponds to a range of weights of 19.9 kg (44 lbs, or 3 stone 2 lbs) for a person of height 1.75 m (a typical mean height for a man), and 17.1 kg (38 lbs, or 2 stone 10 lbs) for a person of height 1.62 m (a typical mean height for a woman). Each 1 unit of BMI corresponds to 3.06 kg (6.8 lbs) and 2.62 kg (5.9 lbs) respectively for these two heights.
- One of the aims of the work in this report is to analyse BMI in much more detail than the BMI categories, in particular by generating frequency distributions using narrow BMI intervals of width 0.5. This provides greater detail on how the prevalence varies across the complete range of BMI values. It also enables other analysis to be done including more specific comparisons of how two BMI distributions differ, such as for two different years, two different ages, or for men and women. Percentile values are also calculated for the distributions and used in the comparisons.
- Statistics are still given extensively in this report for the BMI categories as these can be very helpful in seeing and understanding some of the general patterns and comparing some aspects of the distributions. They also provide familiar comparisons as they are used widely in other work on obesity.
- This report focusses just on the analysis of BMI. Data on health conditions and the associations between BMI and health are not considered in the analysis. However, some of the literature on this is discussed in this chapter to give a broader context to the topic of obesity.
- The overall objective of the work presented in this report is to identify the patterns of the changes over time in the BMI distributions for men and for women in England since the start of the HSE surveys in 1991. There are different aspects to the work. One is to look at the pattern and shape of the population BMI distributions and how they have changed over time. This includes finding suitable models of the changes. Another aspect is to look in detail at the relationship of BMI with age, and with this analysis it is important to take into account birth year since people born at different times will have experienced quite different conditions over their lives. Other topics in the report include modelling future scenarios, calories, and the effect of other factors such as deprivation.
- Consistent colour schemes are used in the charts in the report and these are explained in this chapter, along with the main topics covered in the different chapters in the report.

1.1 Background and motivation

Overweight and obesity refers to a person having high levels of body fat. The World Health Organisation (WHO, 2021) fact sheet web page on obesity [2] states that: "Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health."

Overweight and obesity is recognised widely as a serious population health issue. Rates have increased considerably over recent decades, and it is considered a risk factor for several serious health conditions. Increasing levels of being overweight or obese are therefore expected to result in greater prevalence of these health conditions, leading to a lower quality of life and a reduced life expectancy for many people in the population.

Overweight and obesity is a problem that governments and public health bodies in many countries are concerned about, and therefore they are trying to find ways to improve the situation. For example, in July 2020 the U.K. Government Department of Health and Social Care published a policy paper entitled "Tackling obesity: empowering adults and children to live healthier lives" [3]. This states that "Tackling obesity is one of the greatest long-term health challenges this country faces … Obesity is associated with reduced life expectancy. It is a risk factor for a range of chronic diseases, including cardiovascular disease, type 2 diabetes, at least 12 kinds of cancer, liver and respiratory disease, and obesity can impact on mental health. Our country's rates of obesity are storing up future problems for individuals and our NHS." The policy paper also cites evidence that being obese tends to lead to worse outcomes for those who contract coronavirus (Covid-19).

In terms of the scale of the problem, the World Health Organisation (WHO) website [4] says that: "Worldwide, at least 2.8 million people die each year as a result of being overweight or obese, and an estimated 35.8 million (2.3%) of global DALYs are caused by overweight or obesity", where a DALY is a disability-adjusted life year.

In tackling this problem, understanding in detail what has happened in the past is very important. This gives a foundation for work on why rates of being overweight or obese have increased, what might happen in the future, and how to improve the situation. It also provides comparisons and benchmarks to identify and evaluate changes in the future.

The general purpose of the work presented in this report is to provide a detailed picture of the levels of being overweight or obese in past populations. This requires a way of measuring body type and assessing obesity levels, and suitable past data. A very widely used measure of body type is body mass index (BMI), and this is the statistic used in this report. BMI is explained in Section 1.2. For data, there is an excellent annual data source in the form of the Health Survey for England (HSE), which has been produced since 1991. This data source is described in Section 2.1 in Chapter 2, and the references for the HSE datasets are listed at the end of the references section in this report.

More specifically, the aim of the work here is to use the HSE data to provide a thorough analysis of how BMI has changed in the population of men and women in England since the early 1990s. The motivation is that this work will hopefully make a contribution to the wider research and policy efforts to improve health and that these will ultimately result in greater life expectancy and better quality of life in the population

In this chapter, Section 1.2 explains the calculation of BMI, and discusses its interpretation and usage. The analysis work in this report does not consider the incidence or risk of health conditions as it just focuses on how the distribution of BMI has changed in the population. However, some comments are made in Section 1.3 on the literature on BMI and health, although the section is not intended to be a comprehensive literature review. Section 1.4 specifies the objectives of the work and some of the questions addressed. The scope of the work is set out in Section 1.5. Consistent colour schemes have been used as far as possible in the report and these are described in Section 1.6, along with other aspects of the report format. Section 1.7 lists the topics covered in the report chapters.

[2] https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

[3] https://www.gov.uk/government/publications/tackling-obesity-government-strategy/tackling-obesity-empowering-adults-and-children-to-live-healthier-lives

[4] https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3420

1.2 Body mass index (BMI) and other statistics

1.2.1 BMI equation and use of BMI

This report analyses the Health Survey for England (HSE) data on body mass index (BMI). BMI is a statistic that is calculated easily for an individual as weight in kilograms divided by height in metres squared.

Written as an equation, BMI is given by:

$$BMI = \frac{weight\ in\ kg}{(height\ in\ m)^2}$$

BMI is a measure of body type. The basic idea is to give an indication of whether a person has a relatively high or low weight. The weight of a person will be related partly to their height. If a tall person and a shorter person have the same body shape and composition then the tall person will weigh more simply because they are taller. Therefore, height is included in the BMI equation to adjust for this effect with the aim that BMI just reflects body type. An exponent value of two is used for height. Theoretically, another exponent value could be used which would, of course, adjust for height in a different way.

As stated at the start of Section 1.1, the World Health Organisation (WHO) factsheet web page says: “Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health.” This is not a precise definition in that it does not give a specific way of measuring obesity. However, the WHO factsheet web page [5] says in the following sentence: “Body mass index (BMI) is a simple index of weight-for-height that is commonly used to classify overweight and obesity in adults.” For adults the web page then says that “BMI provides the most useful population-level measure of overweight and obesity as it is the same for both sexes and for all ages of adults. However, it should be considered a rough guide because it may not correspond to the same degree of fatness in different individuals.” The purpose of the work in this report is that given in the WHO quote of looking at overweight and obesity for adults in a population, and the statistic used here is BMI.

BMI does not measure obesity directly as it does not include a measure of the amount of fat. BMI just uses the weight of the person and so does not take into account body composition, such as how the weight of a person is divided between different types of material (fat, muscle, bone, etc.), or where in the body the fat tissue is situated. However, people with high BMI have a high weight for their height and so the general assumption is that they will tend to have high levels of body fat. Therefore, BMI gives an indication of whether a person is likely to have excessive fat and therefore whether they are likely to be overweight or obese. In the general subject area of obesity, BMI is used extremely widely in academic research, in discussions in society and the media, and in health policy.

The reason for using a statistic such as BMI for work on obesity is that measuring body composition is very difficult. A Lancet paper on BMI in Asian populations by a WHO expert consultation group (WHO, 2004) includes a discussion of both direct and indirect methods for doing this in the section headed “Validity of body composition methods”. Direct methods are those that make a measurement that can be used to determine some aspect of body composition directly. The example given in the paper is that total body protein = 6·25 × total body nitrogen, where the nitrogen content is obtained by in-vivo neutron activation analysis.

The indirect methods use assumptions or a model and the WHO (2004) paper recommends the chemical four compartment model for comparing groups in a population where body fat = body weight - (minerals + protein + water). The right hand side values can be obtained or estimated from measurements (the paper lists densitometry, deuterium oxide dilution, and dual energy X-ray absorptiometry) and hence body fat can be calculated from the equation. Various body composition models and measurement methods are discussed in Heymsfield and Waki (1991). The WHO (2004) paper has the comment that the techniques for the measurements for the direct and indirect methods are expensive with only a few laboratories being able to do them. I am not sure whether any easier or better methods have been developed since the time of that paper.

[5] https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

By contrast, BMI can be measured with basic equipment of weighing scales and a stadiometer (i.e., a vertical ruler). Fat levels could in principle be estimated from BMI based on statistical models of the relationship between BMI and fat levels from empirical studies, which WHO (2004) describes as a doubly indirect method. However, BMI is generally used as an obesity measure in itself. The assessment by WHO (2004) is that BMI is "an acceptable proxy for thinness and fatness, and has been directly related to health risks and death rates in many populations." The paper also says: "Results of several studies have shown that BMI correlates highly with percentage of body fat and is largely independent of height, enabling an unbiased comparison between short and tall population groups."

As with any simple statistic, on an individual level BMI has its limitations and needs to be interpreted accordingly. For a particular person there might be specific factors that need to be taken into account. Taking an extreme example, a body builder might have a high BMI value because they have a high weight of muscles whilst they actually have very low levels of fat. The same may apply for some other high level athletes in particular sports. At older ages, a lowering of BMI might occur due to weight falling because of reduced muscle mass or lower bone density, rather than reductions in fat levels.

At a population level, changes in the distribution of BMI mean that some physical changes have taken place in the population. In particular, weight and height have changed in some way that alters the distribution of BMI values. In general, BMI is considered to be a useful indicator of obesity. Reference values are defined and used widely to give BMI categories for underweight, healthy, overweight, and obese. These are set out in Section 1.2.4. Consequently, changes in population BMI statistics are commonly interpreted as showing changes in obesity levels in the population.

As already mentioned, an important feature of BMI is that it is quite straightforward to obtain an accurate value as a person's height and weight can be measured easily with basic equipment. This should help in getting measurements on a consistent basis over time. In the case of a survey like the HSE, this helps to get a good response rate and a large data sample. This makes BMI particularly good for producing population level data across different time periods.

Overall, therefore, BMI is considered a very useful statistic, albeit with some limitations, for assessing body type and estimating obesity levels in populations and how these are changing over time.

1.2.2 Other measures of obesity

Other simple physical measurements for indicating obesity have also been suggested. Some studies have argued that alternative measurements to BMI give a stronger indication of the risks for certain health conditions. For example, a World Health Organisation report (WHO, 2011) looks at waist circumference, waist-hip ratio, and waist-height ratio. The WHO document is a meeting report of an expert consultation meeting that took place in December 2008 (report published in 2011). The report considers that the relationship with many health conditions is "convincing" for BMI (discussed further in Section 1.3.2 below), but that some of these other measures provide a stronger relationship for some of the health conditions (Table 4.1 of the report on pages 17-18). The WHO report recommends further work to be done to assess the different measures.

Of course, if several different statistics are all available for a person or a population then they can be considered together in assessing the health risk. For example, the Public Health England adult obesity PowerPoint slide set (PHE obesity adult slide set: England 2020 [6]) includes a table of risks that combines categories for BMI and waist circumference.

This report uses BMI. However, the analysis methodology that is developed and used here could be applied for any other measures given suitable data.

[6] https://www.gov.uk/government/publications/adult-obesity-patterns-and-trends

1.2.3 BMI units and the magnitude of BMI

As stated in Section 1.2.1, the equation for BMI is: $b = w / h^2$ where b is the BMI value, w is the weight in kg, and h is the height in metres. In this report the BMI value is just given as a number, for conciseness. As can be seen from the equation, the units are kg / m^2 and this applies to all the BMI values in this report.

The equation for BMI can be re-arranged to show how weight changes as the BMI value changes, as follows: $w = bh^2$. Therefore, for a given height value, a difference of 1 unit of BMI corresponds to a difference in weight in kg of h^2.

Average height varies slightly with age but, from Table 13.2 in Chapter 13 (values from SACN, 2012), overall mean adult height values are 1.75 m (5 ft 8.9 in) for men and 1.62 m for women (5 ft 3.8 in). For a person of height 1.75 m, 1 unit of BMI corresponds to a difference in weight of $1.75^2 = 3.0625$ kg (6.8 lbs, or just under 0.5 stones). For a person of height 1.62 m the difference in weight for 1 unit of BMI is $1.62^2 = 2.6244$ kg (5.8 lbs). Hence, 1 unit of BMI corresponds to a reasonably large difference in weight. This is discussed further in the next section looking at the BMI categories.

1.2.4 BMI categories and benefits in looking at the detailed BMI distribution

Patterns in population BMI are often reported by the mean or median values, and by BMI category percentages. The usual BMI categories used are underweight (BMI < 18.5), healthy (18.5 ≤ BMI < 25.0), overweight (25.0 ≤ BMI < 30.0), obese I (30.0 ≤ BMI < 35.0), obese II (35.0 ≤ BMI < 40.0), severely obese (BMI ≥ 40.0).

These boundary values come from proposals by WHO expert groups in the 1990s (WHO, 2004). Different category boundaries have sometimes been suggested for certain populations, such as lower values for populations in the Asia-Pacific region (Inoue et al., 2000). Differences in the associations between BMI, levels of fat, and health risks for different populations particularly in Asia are discussed in WHO (2004). That paper considers whether different category boundary values should be used for certain populations. It recommends that the standard values are maintained but that extra values within the categories are added, namely BMI values of 23.0, 27.5, 32.5, and 37.5, and are used "as points for public health action" (recommendations on page 161). The usage of these additional values might vary in different countries. This implies looking at BMI values in a more detailed way than the standard broad categories, which is also one of the main aims of the work in this report.

The categories can be labelled in different ways. The healthy weight category is sometimes labelled as normal weight. The category of BMI ≥ 40 is described as "severely obese", as above, on the Public Health England PowerPoint slides (PHE obesity adult slide set: England 2020 [7]) and on the NHS website [8]. In some other places, such as the Excel tables on the NHS Digital website that give statistics from the HSE it is described as "morbidly obese" [9]. The World Health Organisation just describes these categories as class I, class II, class III [10]. Sometimes the severely obese range can be split into more BMI categories for 40 to <45, 45 to <50, etc. A comparison of journal papers shows some describing BMI ≥ 40 as severely obese and some as morbidly obese. In this report, the terminology in the first paragraph is used with the obese categories labelled as obese I, obese II, severely obese.

One effect of the BMI categories is that they enable a simpler message to be presented for obesity. Most of the above category end points are whole numbers ending in 0 or 5 (25, 30, 35, 40) which makes them easier to remember. The WHO (2004) paper discusses the use of these values, which it terms "cut-off values", with these uses including general policy for the population and, at an individual level, as one criterion for identifying people who might be at particularly high risk of health conditions.

---

[7] https://www.gov.uk/government/publications/adult-obesity-patterns-and-trends

[8] https://www.nhs.uk/conditions/obesity/

[9] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2019/health-survey-for-england-2019-data-tables Excel file link: Health Survey for England, 2019: Overweight and obesity in adults and children data tables

[10] https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations

However, categorising any numerical variable loses some useful information. Health risks for conditions related to obesity are likely to vary continuously with BMI (for example they might increase as BMI increases above a certain value) and so it is really the BMI value that is important rather than the category. For example, there is likely to be a significant difference in health risks in having a BMI of 30.0 or 34.9 even though both are in same category of obese I. Similarly for the pairs of values at the ends of each of the other categories.

One limitation of the BMI categories is that they are quite broad. As discussed in Section 1.2.3, the weight, w (in kg), for a given BMI value, b, and a height value, h (in metres), is given by $w = bh^2$. This means that the overweight, obese I and obese II categories cover a range of weights of $5h^2$ kg, and for the healthy category the range is $6.5h^2$ kg. From Table 13.2 in Chapter 13, overall mean adult height values are 1.75 m (5 ft 8.9 in) for men and 1.62 m for women (5 ft 3.8 in). The category weight values for these heights are given below in Table 1.1 and Table 1.2. The values are given in alternative units of kg, lbs, and stone and lbs (1 stone = 14 lbs), with rounding of the values to the nearest 0.1 kg and 1 lb.

From these calculations, the overweight, obese I and obese II categories have a range of 15.3 kg (34 lbs, or 2 stone 6 lbs) for a person of height 1.75 m, and 13.1 kg (29 lbs, or 2 stone 1 lbs) for a person of height 1.62 m. For the healthy category the range is 19.9 kg (44 lbs, or 3 stone 2 lbs) and 17.1 kg (38 lbs, or 2 stone 10 lbs) for the two heights respectively. These are wide ranges and so each BMI category covers a considerable difference in weights for a given height.

In analysing the BMI values for a population, BMI category percentages give some information on the distribution of the BMI values. However, Tables 1.1 and 1.2 show that the categories are very broad. The work in this report focusses on deriving a much more detailed distribution of BMI for the population of England using narrow BMI intervals of 0.5, rather than the intervals of 5 and 6.5 for the BMI categories. Such BMI intervals of 0.5 represent a range of weights of only 1.53 kg (3.4 lbs) for a height of 1.75 m and 1.31 kg (2.9 lbs) for a height of 1.62 m.

One of the aims of the work in this report is therefore to analyse BMI in much more detail than the BMI categories, in particular by generating frequency distributions using narrow BMI intervals. This provides greater detail on how the prevalence varies across the complete range of BMI values. It also enables other analysis to be done including more specific comparisons of how two BMI distributions differ, such as for two different years, two different ages, or for men and women. Percentile values are also calculated for the distributions and used in the comparisons.

Statistics are still given extensively in this report for the BMI categories as these can be very helpful in seeing and understanding some of the general patterns and comparing some aspects of the distributions. They also provide familiar comparisons as they are used widely in other work on obesity.

Table 1.1 Weights for the BMI category values for a height of 1.75 m (5 ft 8.9 in)

BMI	Weight (kg)	Weight (lbs)	Weight (stones, lbs)
18.5	56.7	125	8, 13
25.0	76.6	169	12, 1
30.0	91.9	203	14, 7
35.0	107.2	236	16, 12
40.0	122.5	270	19, 4

Table 1.2 Weights for the BMI category values for a height of 1.62 m (5 ft 3.8 in)

BMI	Weight (kg)	Weight (lbs)	Weight (stones, lbs)
18.5	48.6	107	7, 9
25.0	65.6	145	10, 5
30.0	78.7	174	12, 6
35.0	91.9	203	14, 7
40.0	105.0	231	16, 7

1.2.5 Existing BMI statistics for the HSE data

As explained further in Chapter 2, the Health Survey for England (HSE) is an annual national health survey that has taken place each year since 1991. This report analyses the BMI values from the HSE data.

Some existing BMI statistics are produced each year for the HSE data. These are published in an Excel file on the NHS Digital website. The statistics are usually made available at the end of the year after the survey year. For example, the 2019 HSE results were published in December 2020 and the website and the link for the statistics file on BMI and obesity are in the footnote [11]. There are actually several Excel files produced each year on different topics from the HSE and the website in the footnote contains links to all the files for 2019. In some earlier years the files are called "Trend tables" (Section 2.5) but this title is not used for the 2019 HSE and the Excel files are called "Data tables". The HSE in 2020 could not be completed because of the Covid-19 pandemic and so there is no HSE data for that year.

For the 2019 HSE, the main summary adult BMI statistics are in Table 3 of the Excel file and give values for each year from 1993 to 2019. The adult statistics are for ages 16 and over. These show the mean and standard error, and the BMI category percentages (Section 1.2.4) for underweight, healthy (labelled as normal in the file), overweight, obese (which in the table is all those with BMI ≥ 30 and so includes the next category of BMI ≥ 40), and severely obese (labelled as morbidly obese and is BMI ≥ 40). Hence the total percentage for obese I and obese II for BMI of 30 to <40 can be obtained from the table as: obese % – morbidly obese %. Separate values are given for men and for women and for all adults. The values are shown for all ages, and for the age categories of 16-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+. The sample sizes ("unweighted bases") and the total weights ("weighted bases") are also shown. Part of the initial analysis done in the work here included reproducing these values as a check on the data extracted and the selection of cases, and this is described in Chapter 2 in Section 2.5.

Various other statistics are given in different tables in the HSE Excel file for 2019. Tables 1 and 2 give statistics on height and weight for each survey year, similar to those for BMI in Table 3. Table 4 is a summary of the BMI values for 2019 from Table 3. Table 5 splits the BMI values for 2019 by geographical region. Table 6 splits the BMI values for 2019 by index of multiple deprivation. There are waist circumference statistics in Tables 7-10, on the same topics as Tables 3-6 for BMI. Table 11 gives statistics for each year for each combination of BMI and waist circumference categories and the corresponding health risk category. Table 12 gives some 2019 BMI statistics for those with diabetes, with Table 13 being the equivalent table for waist circumference. The remaining numbered Tables 14-20 are on the data for children including BMI categories defined by national percentiles (Tables 16-18), and the percentages split by the mother's BMI category (Table 19) and by the father's BMI category (Table 20). There are four additional tables (A1, A2, B1, B2), with Table A1 giving confidence intervals for adult BMI statistics, Table A2 giving confidence intervals for children's obesity statistics, Table B1 giving estimates for the number of people in the population in each BMI category using the Table 3 percentages and the Office of National Statistics population values, and Table B4 giving estimates for the number of children in the population who are overweight or obese.

Public Health England (PHE) (recently replaced by the U.K. Health Security Agency and the Office for Health Improvement and Disparities) produced an adult obesity PowerPoint slide set of statistics from the HSE data with the latest version being for 2020 [12] using the HSE data up to 2018. The charts present analysis on various aspects of obesity using BMI including the trends of obesity over time, and the prevalence of obesity by age, region, income, deprivation, and ethnic group. There is also data presented on perceptions of weight for different BMI categories, waist circumference categories, combined BMI and waist circumference categories, and multiple risk factors (obese BMI plus at least one of smoking, high level of drinking alcohol, lack of fruit and vegetables, and lack of exercise).

[11] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2019/health-survey-for-england-2019-data-tables Excel file link: Health Survey for England, 2019: Overweight and obesity in adults and children data tables

[12] https://www.gov.uk/government/publications/adult-obesity-patterns-and-trends web page link for "PHE obesity adult slide set: England 2020"

Particularly relevant analysis for the methods in this report

An adult obesity PowerPoint slide set was prepared by Public Health England (PHE) in 2016 and this included charts for population BMI distributions for men and for women. The distributions were produced for men and for women at two time points by combining the 2011-2013 HSE data and by combining the 1991-1993 HSE data. The BMI intervals used were of width 0.5 and the ages included were 18 and over. The values were smoothed using a 10 point moving average. Distributions are not included on the latest version of the PHE slide set for 2020 (referenced in the previous paragraph). One of the main aspects of the analysis in this report is to produce distributions of BMI to look at BMI in more detail. The PHE analysis in 2016 was very helpful in showing how a distribution can be generated from the HSE data. The general approach taken here for producing the BMI distributions is similar in using 0.5 BMI intervals and ages 18 and over and smoothing the data. Some of the details of the approach here are different in using four years of data and a different smoothing method.

Another publication that is relevant to the work in this report is the paper by Wardle and Boniface (2008). They used the HSE data and looked at the increases in BMI between 1993-1994 and 2002-2003 using percentile values and the increases in the percentile values over time. Percentile values are used extensively in this report. The paper of Wardle and Boniface (2008) is explained in detail in Section 4.8.4 of this report and the results from the paper are compared in that section with the nearest data used in this report.

1.3 History of BMI and the association with health

1.3.1 Origins of BMI

The history of BMI goes back nearly 200 years to the 1830s. The approximate relationship of weight to height squared as the human body develops was first identified by Adolphe Quetelet in 1832. A short biography of Quetelet and comments on his work are given in Eknoyan (2008). The index of weight divided by height squared was therefore originally called the Quetelet Index.

The term “body mass index” was first used for the standard equation given in Section 1.2.1 of weight, w, in kg divided by height, h, in metres squared by Ancel Keys and co-authors in 1972 (Keys et al., 1972). This paper was reprinted in 2014 with three additional commentaries (Blackburn and Jacobs Jr., 2014, Lukaski, 2014, and Wells, 2014).

Keys et al. (1972) compared the obesity measures of w/h, w/h^2, the ponderal index, and the percentage of standard weight for the given height (using reference values from old life insurance tables). For the “ponderal index” they used $w^{1/3}/h$ (based on the paper comments and the values in Table 1 in the paper, although the Table 1 title seems to have the inverse fraction), and they also discuss some different formulations and variations of it. Using data from studies in several different countries they evaluated the measures by how low the correlation was with height, and by how high the correlation was with measures of skinfold thickness (an estimate of subcutaneous fat levels). They also had a small amount of data from a few subjects on body density. The values for the different measures were fairly similar but, from their results, Keys et al. chose body mass index as their preferred measure. Of course, since then there have been a very large number of papers that have examined or used BMI.

1.3.2 Association of BMI with health conditions

The WHO definition for obesity in Section 1.1 includes the important aspect of fat levels impairing health. Ultimately the main interest with obesity is with the effect on the likelihood of developing health conditions. Therefore, it is perhaps even more important that an obesity measure is related strongly to health risks than to a particular direct measure of fat levels.

There are many papers that have looked at the association of BMI and health conditions. This report focusses only on changes in BMI and does not look at data on health conditions. Therefore, the vast literature on this has not been reviewed in detail or evaluated. Also, not having a background in medicine or epidemiology, this literature is outside my main area of expertise which limits my understanding and the level of interpretation. However, some papers are described in this section to give a few examples of the results in the literature and to give some context to the importance from a population health perspective of looking at BMI data.

Large scale studies

A lot of studies have looked at the relationship between BMI and specific health conditions and have often found a strong association. This section looks at some papers that are based on a review or a combined meta-analysis of many of these studies. These papers are mostly projects with large teams of authors or contributors.

The Global Burden of Disease studies are very large collaborations looking at diseases on a worldwide basis and the impacts of many different risk factors on health. The studies have been done at several different points in time. For the 2010 study, the main paper is Lim et al. (2012) and has 209 authors. The paper lists “diseases and injuries” (also termed “outcomes” in the paper) for many different risk factors. Outcomes relating to a risk factor are included based on four criteria including “sufficient evidence” of a causal effect. For a high body mass index the paper gives the outcomes as being “oesophageal cancer; gallbladder and biliary tract cancer; pancreatic cancer; kidney and other urinary organ cancers; breast cancer; uterine cancer; colon and rectum cancers; diabetes mellitus; IHD; ischaemic stroke; HHD; the aggregate of cardiomyopathy and myocarditis and endocarditis; the aggregate of atrial fibrillation and flutter, PVD, and other CVD; CKD; osteoarthritis; low back pain” (Table 1 item 7.4 in the paper on page 2229).

The most recent Global Burden of Disease study is for 2019 (Murray et al., 2020) and this gives global attributable deaths for high BMI as about 2.5 million both for females and for males (Figures 3A and 3B in the paper on page 1233), with the percentage of disability-adjusted life-years (DALYs) due to high BMI being between 6% and 7% of the total DALYs (Figures 3C and 3D in the paper on page 1234). For the U.K., the estimate from the 2010 study (Murray et al., 2013) is that 8.6% of DALYs are due to high BMI and this is the 3rd highest risk factor behind tobacco smoking and high blood pressure (paper summary on page 997 and Figure 7 in the paper on page 1015, with also a comment on this on the IHME website [13]). The number of U.K. DALYs attributable to high BMI is estimated in Murray et al. (2013) as 807,000 for men and 645,000 for women (Table 3 in the paper on page 1014). These statistics all indicate that high BMI is considered to be a very serious issue for health in the U.K. and worldwide.

Another huge study is that of the Global BMI Mortality Collaboration (Di Angelantonio et al., 2016) with over 500 investigators. This study looked specifically at the relationship between BMI and mortality for adults with ages between 20 and 90. They combined data from 239 prospective studies. They aimed to avoid confounding effects and reverse causality. To do this, their approach was to only include people who had never smoked and to look at mortality data only for five years onwards following the start of the studies (i.e., excluding mortality data for the first five years since people who died within that time may have had a serious health condition at the start of the study that also affected their BMI). This reduced the sample size substantially but still gave nearly 4 million people. One interesting aspect of the analysis is that they used narrower BMI intervals than the BMI categories which they describe as “excessively wide”. This issue has been discussed above in Section 1.2.4, and one of the main aspects of

[13] http://www.healthdata.org/research-article/uk-health-performance-findings-global-burden-disease-study-2010

the work in this report is to look at the distribution for BMI in much finer detail than the standard categories. They used BMI values of 22.5 to <25.0 as the reference range.

The hazard ratios given in the Di Angelantonio et al. (2016) paper are: BMI 15.0 to <18.5: 1.51; BMI 18.5 to <20.0: 1.13: BMI 20.0 to < 22.5: 1.00; BMI 22.5 to <25.0: 1.00; BMI 25.0 to <27.5: 1.07; BMI 27.5 to <30: 1.20; BMI 30.0 to <35.0: 1.45; BMI 35.0 to <40.0: 1.94; BMI 40.0 to <60.0: 2.76. The paper gives 95% confidence intervals and these are quite narrow.

The hazard ratios in these results are therefore greater than 1, indicating increased mortality, for BMI below 20 and above 25. The Di Angelantonio et al. (2016) paper says about the results that "For BMI over 25.0 kg/m², mortality increased approximately log-linearly with BMI" (in the "Findings" section on the first page of the paper). Hence, there was increased mortality for those in the overweight and obese categories with greater BMI tending to increase mortality. The paper includes analysis on different regions, ages, causes of death, and for men and women. The increase in mortality values with BMI was greater for men than for women with hazard ratios per 5 units of BMI of 1.51 for men and 1.30 for women. The paper says: "This analysis has shown that both overweight and obesity (all grades) were associated with increased all-cause mortality. In the BMI range above 25 kg/m² (the upper limit of the WHO's normal range), the relationship of BMI to mortality was strong and positive in every global region we studied (except perhaps south Asia, where numbers of deaths were small), lending support to strategies to combat the entire spectrum of excess adiposity worldwide." (in the section called "Implications of all the available evidence" on page 777).

Aune et al. (2016) carried out a meta-analysis of 230 cohort studies. These had 30.3 million participants and 3.74 million deaths. Similar to the aim of the Di Angelantonio et al. (2016) paper, they wanted to examine and account for the effect of smoking and existing health conditions. Smoking increases health risks but tends to be associated with lower BMI. The onset of some health conditions can cause a lowering of BMI, even before diagnosis of the condition. Hence, these factors can affect the association of BMI and health risks through confounding effects and reverse causality and the paper includes a discussion of these issues.

Aune et al. (2016) analysed the association of BMI and all cause mortality for several categories of subjects including whether they had smoked, whether they were healthy at the time BMI was measured, and also by the duration of the follow-up time used in the study. They calculated summary relative risk per five unit increase in BMI, identified the BMI values with lowest risk, and fitted fractional polynomial curves to the risk values. They included studies with different BMI categories (not just the standard categories) to increase the number of studies they could use. Also, studies with narrow BMI categories give finer detail of the relationship with BMI. They also did analysis by various other subgroups as well as sensitivity analysis.

The relative risk values per five unit BMI increase given in the Aune et al. (2016) paper are: 1.18 for never smokers, 1.21 for healthy never smokers, 1.27 for healthy never smokers with exclusion of early follow-up, and 1.05 for all participants. The healthy term refers to subjects who were "healthy at baseline". Excluding the early years of follow-up after BMI was measured is a way of mitigating the confounding effects of those with undiagnosed disease at the date of the BMI measurement.

In these results in Aune et al. (2016), the relative risk values are much higher for the sub-groups (attempting to account for the various potential confounding factors) than for all participants. This indicates that the confounding factors can have quite a large effect on the data analysis. The 95% confidence intervals are given and are fairly narrow, and even though the value for all participants is quite close to 1 the confidence interval does not include 1 (1.04-1.07).

In Table 2 of the paper the lowest risk values are at BMI of 23 and 24 for never smokers, at 22 and 23 for healthy never smokers, and at 25 for all participants. Aune et al. (2016) also report that in studies of never smokers where the duration of follow-up was at least 20 years the lowest risk values are for BMI of 20 and 22. From Table 2 in the paper, relative to a BMI of 23, the risk value for all participants is slightly below 1 for the overweight category BMI value of 27.5, being 0.98. This is not the case for the never smokers and healthy never smokers where the risk values for a BMI of 27.5 are 1.07 and 1.11, respectively. The 95% confidence intervals are quite wide for these values, being 0.96-1.01, 1.01-1.14, and 1.05-1.18. All of the subject categories have high risk values for low BMI values of 15, 16 and 17.5 in Table 2 of the paper, which are in the underweight range for the standard BMI categories.

Overall, the patterns across the full range of BMI values are described by Aune et al. (2016) as "J shaped" for the never smokers and healthy never smokers, compared to a more "U shaped curve" for all participants. A "U shaped" curve means a wider range of BMI values that have low risk than "J shaped", possibly including values in the overweight category. However, the results for all participants that have a "U shape" are likely to be affected by the confounding factors. These issues are considered further in the next subsection in this report on the topic of the obesity paradox. The paper includes a brief discussion towards the end of the paper of the mechanisms and diseases that could explain the link between higher BMI and increased mortality. The results for the sub-groups that take account of the confounding effects, particularly smoking, lead to the paper conclusion (on the first page of the paper) that "Overweight and obesity is associated with increased risk of all cause mortality and the nadir of the curve was observed at BMI 23-24 among never smokers, 22-23 among healthy never smokers, and 20-22 with longer durations of follow-up."

A paper by the Non-Communicable Diseases Risk Factor Collaboration (NCD-RisC, 2016), which looks at global trends in BMI, states that "High body-mass index (BMI) is an important risk factor for cardiovascular and kidney diseases, diabetes, some cancers, and musculoskeletal disorders." (NCD-RisC, 2016, page 1377). NCD-RisC is a large global network of health scientists [14] and the paper lists hundreds of collaborators.

Lauby-Secretan et al. (2016) is a paper in the New England Journal of Medicine from a working group of the International Agency for Research on Cancer (IARC). The paper updates the assessment of the effect of obesity on cancer from a previous working group in 2002. The review that was done covered more than 1000 studies. It assessed whether there was sufficient evidence of the link between BMI and cancer where (note to Table 2 on page 796): "Sufficient evidence indicates that the International Agency for Research on Cancer Handbook Working Group considers that a preventive relationship has been established between the intervention (in this case, the absence of excess body fatness) and the risk of cancer in humans — that is, a preventive association has been observed in studies in which chance, bias, and confounding could be ruled out with confidence." Table 2 in the paper on page 796 lists 13 cancers of different types or different sites where the evidence is considered sufficient, and gives relative risk values. For 11 of the cancers the relative risk is for the highest BMI categories considered in the studies relative to normal weight, with values varying from 1.1 to 7.1 (most being 1.3, 1.5, or 1.8). For two of the cancers the risk is given as a value is per 5 units of BMI and in both these cases the value is 1.1.

The WHO expert consultation report (WHO, 2011) on different measures of obesity, referred to in Section 1.2.2, describes the relationship with BMI as "convincing" for cardiovascular disease, type 2 diabetes, hypertension, overall mortality ("without mutual adjustment of the anthropometric parameters"), colorectum and breast cancer (WHO, 2011: Table 4.1 on pages 17-18).

A slightly older paper is that of Guh et al. (2009) who reviewed studies looking at incidences of a wide variety of health conditions and conducted a meta-analysis. Most studies used BMI whilst a few used waist circumference. They used the usual categories for BMI and calculated risk ratios for overweight and for obese compared to the normal (healthy) category. They also used categories for normal, overweight and obese when using waist circumference. They found 89 suitable studies covering 18 potential co-morbidities. They found significant results for a wide range of health conditions for increased risk for both overweight and obese. These included type II diabetes, many types of cancer, cardiovascular diseases (all types for obese, and all except congestive heart failure for overweight), asthma, gallbladder disease (for obese), and chronic back pain. They note that the highest risk ratios are for type II diabetes in women using BMI as the obesity measure with values of 3.92 for overweight and 12.41 for obese. The risk ratios for men for type II diabetes using BMI are also high with values of 2.40 and 6.74 – these are higher than for the other conditions for men apart from for osteoarthritis for overweight at 2.76. There were only a few studies for many of the conditions and so some of the 95% confidence intervals for the risk ratios are fairly wide.

Overall, therefore, these large scale study papers indicate that high BMI is considered a risk factor for many serious health conditions.

[14] http://ncdrisc.org/about-us.html

Issues in the association between BMI and health conditions

In looking at the association between BMI and health conditions there are various issues to consider. An extremely important general statistical principle is that an association between any two variables does not prove that there is a causal link. Similarly, lack of association does not necessarily mean that there is no causal relationship. Some aspects that it is very important to consider are potential confounding factors, and the possible direction of a causal relationship.

In the case of BMI and health conditions there are a number of issues that can complicate the analysis of any relationships. One issue is that health conditions will often develop over time and so are likely to be related to the pattern of obesity levels throughout a person's life. Therefore, risks will depend in some way on the lifetime trajectory of values of BMI rather than on just the current value or the value at any one specific time.

Another issue is that there can be interactions with other risk factors. For example, smoking will greatly increase the risk of certain health conditions, but may have a tendency to reduce weight and BMI.

There can also be reverse causality where a health condition may cause an increase or decrease in weight and therefore BMI. Health conditions can lead to changes in BMI through the effects of the condition itself or the treatment for the condition. In particular, some health conditions can cause unintentional weight loss before they are diagnosed. Hence, a lowering of BMI could be the result of a health condition starting to happen. In other words, the undiagnosed health condition causes BMI to be lower than it would be otherwise. On the other hand, it is also possible that a health condition that limits movement and exercise could potentially lead to increases in BMI due to a lack of exercise.

A discussion of reverse causality is given in an editorial article in the journal Circulation by Sattar and Preiss (2017). The paper talks about several examples, including some on obesity. They comment that "reverse causality is more often in play than one might imagine" (page 2370) and they consider some statistical approaches for dealing with it.

It is therefore not straightforward to examine and interpret the links between BMI and health conditions. In particular, using the data in different ways can give different results. As an example, Yu et al. (2018) had BMI data where there was a weight history of the participants from questionnaires over a 16 year time period. The end of this is the baseline after which mortality events are recorded. They analysed associations with mortality for the maximum BMI from the questionnaires and for BMI at baseline (i.e., the end of the period for recording weight). Using maximum BMI, they found hazard ratios above 1 for overweight and obesity categories, which increased with obesity. By contrast, when they calculated results using baseline BMI from the last observations the hazard rate for overweight was slightly below 1 (the obesity paradox, discussed in the next paragraph). This is even though, from the supplementary material, for many of the participants it looks like the maximum BMI was the baseline value. The paper discusses the issue of illness causing weight loss (i.e., reverse causality) and this is the main reason that they used and preferred maximum BMI. They say that "the highest risks of death occurred among those who had experienced substantial decreases in weight, which is most likely reflective of unintentional weight loss caused by apparent or preclinical disease."

One issue that is discussed in many recent papers is the "obesity paradox". This arises from studies that have produced results where health outcomes (such as mortality) are better for the overweight BMI category than for the healthy BMI category. Sometimes, even the data for obese I has better outcomes than the data for the healthy category. For example, a meta-analysis study by Flegal et al. (2013) found hazard ratios of less than 1 for the overweight BMI category compared to the healthy weight category. This indicates lower mortality rates in the data for the overweight group. As mentioned in the last paragraph of this section overleaf, there are many papers that report results with an obesity paradox effect.

This is different to the results above from the later paper of the Global BMI Mortality Collaboration (Di Angelantonio et al., 2016) included in the previous subsection above on "large scale studies". As discussed above they excluded smokers and the first five years of mortality data to try and reduce confounding effects and reverse causality. They examined the effects of adding the adjustments which they describe as "successively stricter precautions against bias" (Table 1 title on page 778). Using the standard BMI categories they found that the hazard ratio for the overweight category changes from below 1 to above 1 going from analysis with no exclusions to the analysis with the most exclusions and adjustments (Table 1 on page 778). This means that the results change from obesity paradox to no obesity paradox. The paper includes some discussion and comparison with the results in Flegal et al. (2013). In the "Implications of all the available evidence" section on page 777, the paper says: "Our results challenge recent suggestions that overweight and moderate obesity are not associated with higher mortality".

The paper of Aune et al. (2016) in the previous subsection also considers whether results with better outcomes for the overweight category could be due to such confounding effects. The paper includes some discussion on the differences compared to the paper of Flegal et al. (2013) in the studies used and in the approach taken. As discussed in the previous subsection, Aune et al. (2016) found a "U shaped" curve for risk for all participants where the risk value for a BMI of 27.5 is 0.98 and so slightly less than for the reference BMI value of 23, although the confidence interval includes 1. By contrast, the values for never smokers and healthy never smokers at a BMI of 27.5 are 1.07 and 1.11, respectively (Table 2 of the paper) and are described as a "J shaped" curve. The paper reports results on many other analyses by subgroup such as by length of follow-up. For all participants they comment that the effect of increasing the follow-up duration is for the curve of the mortality results to alter from a U to a J shape. This comment refers to Table G in the paper supplementary material with the risk values given for all participants for various different follow-up durations. For example, for a BMI of 27.5 the values are 1.08 and 1.06 for a follow-up of 20-<25 years and ≥25 years, compared to risk values ≤ 1 for the shorter follow-up durations. They comment that excluding smokers from the data increases mortality rates for overweight and obese values. They also discuss the results for low BMI values. For a BMI of 20 (low BMI but within the healthy range) for all never smokers the mortality value is 1.10 with a confidence interval of 1.05 – 1.14 (Table E of the supplementary material), and so indicates increased risk. However, this changes to a value of 0.99 for follow-up durations of ≥20 years. They comment on these results and the effect of the confounding factors as follows: "Thus, the increased risk observed with a BMI of 20 in the analysis of all participants and never smokers and the lower risk in overweight people in the analysis of all participants is likely to be caused by confounding by smoking and prediagnostic weight loss." In other words, the results in the paper imply that the measures to account for confounding effects by excluding smokers and early follow-up can have quite a big effect on the findings. This can alter the conclusions for the overweight BMI category and for values at the lower end of the healthy category.

Hence, the results and analysis in Di Angelantonio et al. (2016) and Aune et al. (2016) suggest that confounding factors could sometimes be the cause of obesity paradox results.

There have been many papers that report results with an obesity paradox effect. A wide variety of explanations have been suggested for the obesity paradox. Some papers have suggested causal mechanisms that might result in genuine advantages for those in the overweight or obese I category for a particular health condition. There are also a number of suggested explanations that this is not a causal relationship, such as Di Angelantonio et al. (2016) and Aune et al. (2016). There are a wide range of possible reasons for this including confounding variables (such as smoking), reverse causality, bias in the samples (such as selection bias, survivor bias, or collider bias), limitations of measuring BMI at one point in time (e.g., Yu et al. (2018) discussed above), limitations of BMI as an indicator of obesity, differences in the time when diseases are diagnosed (e.g., obese people being considered more at risk and so perhaps being diagnosed earlier and so getting better outcomes), differences in treatment regimes, and the wide categories of BMI (as discussed in Section 1.2.4). It could be that there is one relationship for the general population between obesity and the risk of developing a condition, but a different relationship for the much smaller group of people who develop the condition between obesity and the outcome of the condition. There could be different explanations for different health conditions. There are hundreds of papers on this issue and no attempt has been made here to review them. It would appear that at present there are many different views on this issue and, as an example, a recent discussion of the different explanations in the context of cardiovascular disease is given by Koliaki et al. (2019) in their paper in Section 4.1 (page 103) and Table 2 (page 104). I am not sure if there is widespread agreement amongst researchers on any particular explanation. There could, of course, be a combination of reasons for the results obtained.

Categories used in the association studies

One aspect of note is that some of the studies looking at the association between BMI and health conditions or mortality use the standard BMI categories. As discussed in Section 1.2.4, each of the categories are broad and cover a wide range of weights. The healthy category is the broadest with a difference in weights of about 3 stone: from Section 1.2.4 the range is 19.9 kg (44 lbs, or 3 stone 2 lbs) and 17.1 kg (38 lbs, or 2 stone 10 lbs) for the mean heights for men and women respectively.

Health risks could vary considerably across each category. For example, the studies often find a substantial increase in the risk of adverse outcomes for the underweight category. Perhaps lower weights within the healthy category could also have a higher risk, affecting the overall data and the risk level for the healthy category. Within any of the standard categories there could be considerable differences in the effects.

If the information is available and there is sufficient data then it would be better to analyse the data in narrower intervals. It would certainly be interesting to see hazard ratios for much narrower BMI categories. This report just looks at BMI on its own without relating it to health conditions, but one of the main aspects of the work here is to use narrow intervals for the BMI frequency distributions.

The Global BMI Mortality Collaboration paper (Di Angelantonio et al., 2016), discussed above in the subsection on large scale studies, is an example of using smaller categories. For example, they found hazard ratios above 1 for a BMI between 18.5 to <20 (i.e., in the healthy range) compared to the reference category of 22.5 to <25.

Another example also discussed in the subsection on large scale studies is Aune et al. (2016), who were able to combine the results from studies with different categories by using a method of fractional polynomial models. This enabled them to use the results from more studies giving them a much larger sample size. As described above, they were also able to specify narrow BMI ranges that had the lowest risk.

Implications of this report for research on health conditions

This report just focusses on BMI values in the population and so is not concerned directly with the health outcomes. However, the results from this report will hopefully be useful for researchers in considering the health implications of obesity. For example, the patterns and changes in BMI identified in the report can potentially be translated into predicted health outcomes using assumptions or models of the relationships between BMI and the health conditions.

As mentioned above, ideally for analysis of the risks associated with BMI we would like to know the BMI of the person throughout their life. The values at one point in time could be quite different to past values, particularly due to aging effects. However, this will not generally be possible as the data will not be available. One aspect investigated in this report in Chapters 7-10 that may provide useful information on this is how BMI generally tends to change with age. This might help the interpretation and comparison of BMI values in studies where BMI is just measured at a single point in time and where this is at different ages for each person. For example, perhaps the percentile in the BMI distribution for the age of the person could be used instead of the actual BMI value. The results in Chapters 7-10 could also allow some estimation or modelling of a person's lifetime BMI values given the BMI value at a particular age.

Other research

There are many other areas of research in obesity and this section mentions a few of these briefly. One important aspect of obesity is interventions to reduce weight and fat levels. Two considerations are whether such interventions are effective and the impact of reductions in weight on health risks. In the latter case this would increase the general understanding of the relationship between obesity and health conditions. Ideally, to give the best evidence, such assessments of interventions or the effect of weight loss should be done with randomised control trials.

Some examples of randomised trials done on the effectiveness of interventions are three studies involving Professor Jebb from Oxford University, as follows. Aveyard et al. (2016) compared different options for general practitioners (GPs) in giving advice to their patients for just 30 seconds. Suitable patients were adults with a BMI of at least 30. The active intervention was to give a referral to a weight management group, compared to the control of giving advice that weight loss would be beneficial. After 12 months the mean reduction in weight was 2.43 kg for the intervention group (940 patients) compared to 1.04 kg for the control (942 patients). They calculated an adjusted mean difference of an extra 1.43 kg of weight reduction for the intervention with a 95% confidence interval of 0.89 to 1.97 kg. This is even though only 40% of the intervention group actually attended the weight management programme. Ahern et al. (2017) looked at three options for the participants in the trial who were adults with a BMI of at least 28. The options were advice and self-help materials, or a weight management programme of either 12 or 52 weeks. The mean weight losses of the three groups after 2 years were 2.30 kg, 3.00 kg, and 4.29 kg, respectively, although the difference between the advice group and the 12 week group was not statistically significant. Astbury et al. (2018) did a trial on a total diet replacement programme that included a very low calorie intake of 810 kcal for the first 8 weeks. The participants for this trial were adults with a BMI of at least 30. After 1 year the group on the diet lost 10.7 kg compared to the control group receiving usual care who lost 3.1 kg. Hence, this is quite large weight losses for the diet group (10.7 kg = 1 stone 10 lbs). The adjusted difference in mean weight loss between the groups was 7.2 kg with a 95% confidence interval of 4.9 to 9.4 kg.

Regarding the outcomes of weight loss interventions, Ma et al. (2017) carried out a meta-analysis of randomised control trials. Low fat diets were part of the interventions for all the trials except one. They used trials where the mean BMI was 30 or more (apart from in post-hoc analysis). For the outcome of all cause mortality they had 34 trials with 21,699 participants. The risk ratio from their analysis of these trials is a value of 0.82 for the combined experimental groups compared to the control groups with a 95% confidence interval of 0.71 to 0.95. Hence, this indicates better outcomes for mortality events for the weight loss groups. The number of trials and hence the amount of data for more specific outcomes was much smaller. The risk ratios that they calculated are below 1 for cardiovascular mortality, cancer mortality, and new events for cardiovascular and cancers but the 95% confidence intervals include 1 in each case.

On the general topic of weight loss trials, Heggie et al. (2020) propose a standard template for reporting the details of how a trial is carried out and the intervention used. The idea is to allow better comparison and evaluation of different trials and to facilitate meta-analysis. They piloted the template on 39 trials.

Interventions at a population level are a matter of health policy, and another research area is to consider how best to apply them. Hazlehurst et al. (2020) summarise and review the clinical pathways developed by NHS England for severe adult obesity. They make some suggestions for changes including more flexibility in the selection of treatments, and just having two main tiers for prevention and treatment. This paper also includes a much more extensive review of the literature on interventions than that given in this section, as part of assessing the evidence for the effectiveness of the different services provided. One issue that they mention with much of the evidence in the literature is a lack of data on outcomes over the long term.

Financial aspects also need to be considered and there will be cost implications of obesity and of weight loss interventions. If obesity results in greater prevalence of health conditions then this will potentially lead to higher costs. Kent at al. (2017) reviewed studies from the literature that looked at the link between BMI and annual health costs using the BMI categories. There were differences between the values in the various studies, but in most cases overweight and obesity were associated with higher costs than the healthy weight category (there were 26 studies giving values for overweight and 28 for obese). The median values were increases in costs of 12% for overweight and 36% for obesity compared to the healthy weight. Based on overweight and obesity prevalence in the U.K., Kent at al. (2017) estimated that this implies that: “overweight and obesity combined are associated with around 12% of adult healthcare expenditure in the UK” (page 874). If this is the case then it is likely that increases in obesity levels will also mean increases in costs either for the individuals or for the healthcare system. Similarly, reductions in obesity levels would reduce costs.

Another aspect of obesity research is to explore different factors that might affect obesity. A large study that produced a map of many factors is the Foresight study from 2007, and this is discussed in Chapter 6 of this report in Section 6.3.1. An example of examining a specific factor is Johnson et al. (2011) who looked at the associations between the BMI of parents and their offspring at adult age.

There are also other analysis and modelling tools that can be used. For example, Tuson’s (2019) PhD thesis applies agent-based simulation modelling to look at how social network effects might be part of the reason for increasing obesity levels.

1.4 Objective

The overall objective of the work presented in this report is to identify the patterns of the changes over time in the BMI distributions for men and for women in England since the start of the HSE surveys in 1991.

As part of this, one of the aims of this report is to provide a detailed record of BMI at different times since the early 1990s. This will hopefully be a useful resource as a long term reference. The results can also potentially be updated as new data becomes available in the future.

There are several different aspects to the work. One is to look at the pattern and shape of the population BMI distributions and how they have changed over time. This includes finding suitable models of the changes. Another aspect is to look in detail at the relationship of BMI with age, taking into account birth year since people born at different times will have experienced quite different conditions over their lives. Other topics in the report include modelling future scenarios, calories, and the effect of other factors such as deprivation.

The specific questions addressed in this work include:

- How has the distribution of population BMI changed over time since the early 1990s?
- What are the differences in the rates of change over this time interval?
- Can the changes be represented by a model and what are the implications of the model?
- How does mean BMI vary with age and how does this alter depending on birth year categories?
- How does the distribution of BMI change as age increases?
- How can the relationships and models identified be used to create scenarios of future BMI?
- Can the BMI changes be translated into daily calorie amounts and how large are these amounts?
- How do the BMI distributions vary by other factors, such as by different categories of deprivation levels?
- How do the results for each of the above questions differ for the data for men and women?

1.5 Scope of the work in looking at adult BMI for men and for women

The scope of the work in this report is to look at adult BMI. This is defined here as ages 18 and over and so only cases with age ≥ 18 were used. Women who are pregnant are also outside the scope since their weight will include the baby and so a valid BMI value cannot be calculated.

The BMI distributions are quite different for men and for women and so the data for each is analysed separately. This is the usual approach in obesity research, although it means twice the work as it generally means doing the same analysis on both data sets with descriptions produced for each set of results! Combined statistics or distributions were not calculated as it is considered that the separate analysis is much more useful and meaningful.

The BMI categories given in Section 1.2.4 are generally used for both men and women and this assumes that the same category values are suitable for both. This implicitly assumes a similar type of relationship for men and women between BMI and the risk of developing health conditions. For example, the range of BMI values that are “healthy” is the same for both men and women. This ideally ought to correspond to BMI values that have the lowest risk of the health conditions.

The main focus of this work is on high BMI values because of the health risks of obesity. There are also potential issues with being underweight. The BMI distributions produced in this report cover the full range of BMI values and so there are results and statistics across all the different BMI categories.

My main area of knowledge and experience is in analytics and modelling. The report therefore concentrates on data analysis, statistics, and mathematical modelling. Some comments on potential explanations and implications are made, but further implications can hopefully be drawn by others with greater expertise in areas such as public health, epidemiology, social policy, biology, medicine, and nutrition.

1.6 Report format and chart colour schemes

Report format and level of detail

The aim of the report is to explain and present the work done clearly. One aspect of this is to provide enough detail so that it is clear how the analysis was done and also so that it can be reproduced. The ability to reproduce or replicate scientific work is a very important part of the scientific method and is a current issue that has been identified and discussed in many areas of science. Reproducing a study is a useful research approach that could be adopted more commonly (Liu and Brooks, 2016). The level of detail here will hopefully be sufficient to enable others to repeat the analysis, or to apply and extend it as new data becomes available over time, or to use the methods on other data.

Another aspect with the report format and presentation is to keep the report readable and to highlight the main points. The analysis done was extensive with over 500 Excel spreadsheets including thousands of charts and so it is difficult to get this balance right.

The report is split into this main report and a separate document of appendices. The results included here in the main report have been selected as those considered to be the most important. This is with the aim of helping the readability of the report. Other results are in the appendices which are quite long but provide additional information for those wanting an additional level of detail.

As explained in Section 1.5, the data for men and for women are analysed separately. In the report, there are usually separate sections for the results for men and for women, although there are also some sections comparing the results.

The obesity issue is considered to be serious for both men and women. If we consider the percentage of people who have a BMI value above the healthy category then this percentage is higher for men. Therefore, in this respect, the general overweight and obesity issue is more widespread and serious for men than for women. The mean BMI value is also higher for men than for women. On the other hand, the severely obese category percentage is higher for women than for men and so the more specific issue of very high BMI values is more prevalent for women. These are just the relative differences and there are many men and women with BMI values in each of these categories.

The analysis process for the data for men and for women is essentially the same and so the report tends to have a structure of giving the same set of results for men and for women. These results are compared, although the patterns of the changes over time are often very similar. For some of the work, the same analysis process was repeated many times when considering the data for different time intervals, different ages, and for men and women. There is therefore some repetition in the report when describing the analysis.

Chart colour schemes

The report includes many charts of the results. Consistent colour schemes have been chosen to try and make these as easy as possible to read and understand. In some cases, the charts have many different series and so distinguishing them was quite difficult.

In general, the colour schemes used are:

- For charts with one series for men and one for women the colours used follow the main colour scheme of the document, which are the main Lancaster University brand colours of grey and red. The colour scheme is: data for men: grey, data for women: red.
- For charts showing the main years analysed of 1993, 1997, 2001, 2005, 2009, 2013, 2017, the colours for both men and women are:
    - 1993 and 1997: blue
    - 2001 and 2005: orange
    - 2009 and 2013: purple
    - Different marker fills are used to distinguish each of the above pairs of years.
    - Chapter 14 has additional analysis on 2017 and the data for this is in gold.
- For charts of the BMI categories the colours are: underweight: blue, healthy: green, overweight: orange, obese I and II: red, severely obese: purple. This is loosely based on traffic lights colours (as also used in some food labelling).
- For the models, the model colours are: model for men: dark green, ,model for women: purple.
- For charts with many series (such as age cohorts), generally the default Excel colour schemes are used. The same colours are used in the charts for men and for women.

The chart legends also have detailed information about each series again to try and help readability. The charts and tables are numbered consecutively within each chapter. So, for example, the charts in Chapter 2 are numbered Figure 2.1, Figure 2.2, Figure 2.3 etc. Lists of all the tables and figures with their page numbers are given before this chapter on pages xxiii – xxvii.

1.7 Report chapters

The main topics in the report chapters are as follows:

- Chapters 2-4: Deriving BMI distributions for men and for women and looking at how the distributions have changed over time. Distributions are produced for 1993, 1997, 2001, 2005, 2009, 2013.
- Chapter 5: Modelling the changes in the BMI distributions using a linear scaling transformation.
- Chapter 6: The interpretation of the changes in BMI and discussion of the factors that affect obesity.
- Chapters 7: Investigating how mean BMI varies with age. It is important to take into account the effect of birth year, which is done through cohort analysis.
- Chapters 8-10: Looking at BMI distributions for different ages and how they vary over time and with age.
- Chapters 11-12: Developing a model for generating BMI distributions for different scenarios.
- Chapter 13: Converting BMI changes into calorie amounts.
- Chapter 14: Updating the results for the latest HSE data for 2015-2018. This includes producing BMI distributions for 2017.
- Chapter 15: The relationship of BMI with category variables for deprivation and for height.
- Chapter 16: A summary of the findings.

There are two main areas of the work. The first area is to derive and analyse the population BMI distributions across the whole HSE data. Four years of HSE data are used, and this produces estimated population distributions for 1993, 1997, 2001, 2005, 2009, and 2013. The distributions are compared and the changes modelled using linear scaling transformations. This is set out in Chapters 2-5 of the report. Chapter 2 describes the selection of cases from the HSE along with some basic descriptive statistics. There is also consideration of possible biases in the data including looking at missing values. Chapter 3 explains the methodology for deriving the BMI distributions. The distributions obtained are then described and compared in Chapter 4. The changes are modelled using scaling transformations and the method and results for this are set out in Chapter 5.

The second main area is to look at the relationship between BMI and age, and again how this has changed over time. An important aspect of this is to look at cohort effects. In other words, how BMI varies with age for people born within a small number of years of each other. This helps to explain how the overall population BMI has changed over time in the way that it has. The general relationships found for mean BMI are explained in Chapter 7. The results from more detailed analysis for men and for women are covered in Chapters 8 and 9 respectively. Chapter 10 compares the BMI distributions for men and women with each other as age increases.

The report includes other related topics. Chapter 6 provides a discussion of the factors affecting BMI and obesity, and in particular how the changes in BMI should be interpreted. The relationship of BMI with age provides a way of modelling BMI and this is described in Chapter 11. This then enables scenarios to be generated and evaluated and some examples are given in Chapter 12. Another topic is to consider what the differences in BMI might correspond to in terms of calories. This is examined in Chapter 13 using models from the literature. The report results were updated during the work using new HSE data that became available and this included producing the BMI distributions for 2017, and these results are given in Chapter 14. The relationship of BMI with categories of deprivation and with height are examined in Chapter 15, partly as examples of how the analysis can be applied to different variables. A conclusion and summary of the findings is in Chapter 16.

Chapter 2: Analysis of the HSE data for each year

Key points from this chapter:

- The extraction of valid cases for each year of the HSE is explained. The approach is similar to that used by NatCen and UCL in producing the Trend Table statistics. One check was to reproduce the Trend Table statistics and good agreement was obtained.
- The descriptive statistics for each year show a steep deterioration in BMI from 1991 to 2001 but much smaller changes since then. The statistics in this chapter are preliminary analysis with more detailed analysis of the BMI distributions and the changes in BMI in subsequent chapters.
- Missing values and cases with a high weight above the limit of the scales were analysed as these are limitations of the data. The analysis in this chapter gives some reassurance that they do not introduce much bias and that the data is comparable over time.

2.1 Health Survey for England (HSE)

2.1.1 HSE background

The Health Survey for England (HSE) is an annual national health survey that has taken place each year since 1991. It is sponsored by NHS Digital (formerly the Health and Social Care Information Centre) [15], and since 1994 the HSE has been carried out by a collaborative team of staff from the National Centre for Social Research (NatCen) and University College London (UCL), Department of Epidemiology and Public Health. The references for the HSE datasets for each year from 1991 to 2019 are listed at the end of the references section in this report.

The HSE is an excellent source of data because it has a good random sampling methodology and the main data is collected by trained interviewers, with some additional more specialist data obtained using a nurse visit. Therefore, the data should be of high quality and representative of the overall population. The data covers an extensive range of health related issues including physical measurements, health conditions, diet, and lifestyle. The height and weight measurements used for the BMI calculation are taken by the interviewers. Information about the survey is given on the UK Data Service website [16] and on the NHS Digital website [17].

The HSE sampling methodology is described in the "Methods and Documentation" document produced each year. For example, the last year of data used in the analysis in this chapter is 2014 and the document for this is: Health Survey for England 2014, Volume 2, Methods and documentation, The Health and Social Care Information Centre, 2015 [18]. The method is described in section 2 of the document as "multi-stage stratified probability sampling". From the description in the document, the general approach is that primary sampling units (PSUs) are created based on postcode areas and these are ordered by the stratifying variables of region, local authority, and a socio-economic criterion (using National Statistics Socio-economic Classification (NS-SEC) groups). The units to be used are selected from the ordered list at fixed intervals from an initial point chosen at random. Households are then selected at random from these units.

[15] https://www.gov.uk/government/organisations/nhs-digital describes NHS Digital as "an executive non-departmental public body, sponsored by the Department of Health and Social Care".
[16] https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000021
[17] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england
[18] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2014 Link to the PDF file is: Health Survey for England 2014: Methods and documentation

There is a core set of questions asked, with additional questions in some years looking at particular issues such as a specific health condition. There is always a main sample, but with a separate boost sample in some years of children, ethnic minorities, or older people.

There have been some changes over the years. The main change has been that a non-response weighting variable was added in 2003. Weightings are calculated at various levels and are described in section 7 of the document for the 2014 HSE (footnote 18 on the previous page). The data used for BMI of height and weight is obtained from the HSE interviews, rather than the nurse visit, and so the "interview weighting" is the relevant weighting. This gives a non-response value taking into account various factors including age group, sex, household type, region, and socio-economic classification. The HSE documentation advice is to use the interview weight for analysis of interview questions.

There is obviously scope for confusion with terminology here. In the remainder of the report "interview weight" or "case weight" will generally be used when referring to the weightings applied to each case to try and account for non-responses, whereas "weight" on its own will generally refer to the physical weight of the person. The meaning should also hopefully be clear from the context of the particular sentence.

Regarding software, the HSE data is provided in SPSS files. The relevant variables used here for this work were copied into Excel spreadsheets and the analysis was done in Excel.

2.1.2 Overview of the HSE analysis for each year

The main purpose of the approach used for the analysis of the HSE data for each year is to produce a data set for men and for women of valid cases. The data sets from this method were then used in the subsequent analysis described in this report, including being combined together in generating the population BMI distributions (Chapter 3).

An outline for the process for selecting the data is in Section 2.2. There are two aspects to this. One is to define and identify valid cases. There are various criteria, and these are set out in Section 2.2.1. The other aspect is to determine the values of the variables. The variables required are age, BMI, and the interview weighting variable and this aspect is discussed further in Section 2.2.2.

Some additional analysis was also done on the data sets for each year. Descriptive statistics were calculated and these results are discussed in Section 2.3. The missing cases were also analysed and this is presented in Section 2.4.

The approach taken for defining valid cases was essentially the same as that used by UCL and NatCen in producing the statistics in their "Trend Table" (Section 2.5). One check that was done was to try and reproduce the statistics in the Trend Table and the results for this are in Section 2.5. Many other checks were done throughout the analysis such as checking totals and calculating statistics on the data in different ways.

One issue with the HSE is the cases where the weight of the person is high and exceeds the weight limit of the weighing scales. Analysis on this is set out in Section 2.6. Some statistics on high BMI values are examined in Section 2.7.

This chapter along with Chapters 3, 4 and 5 gives results for the HSE data for 1991-2014. This was the data available when the work here was started. As discussed in Chapters 3 - 5 the data is combined in groups of four years in producing the BMI distributions. By the time of the final version of this report HSE data for 2015-2018 was available which enabled an additional distribution to be produced for 2017. Updated results for this data are given in Chapter 14. The analysis that can be done for the HSE data for 2015-2018 is a bit more limited as the age of each case is given in wider categories of several years rather than a single year.

2.2 Selecting the data from the HSE

The step by step procedure used for extracting and analysing the HSE data is given in Appendix 2.I. The description in Appendix 2.I is intended as a detailed reference for reproducing the analysis or applying it to new data in the future. This section gives a general description of what was done. The data for men and for women were analysed separately, which is the usual approach for BMI and obesity data (Section 1.5). Hence, each of the steps were carried out separately for the data sets for men and for women.

2.2.1 Valid cases

For each year only data from the main HSE sample was used as the boost samples are not representative of the population. The main criterion used for selecting cases was that they have a valid BMI. These cases are defined in the HSE data by the “bmiok” variable having a value of 1.

It was also decided to exclude any cases with an extremely low BMI value of less than 10 even if the “bmiok” variable has a value of 1. Five such cases with a valid BMI code were excluded. These values are so low that they could be errors. Even if they are true values, they are very extreme values and the analysis here is not concerned with extreme low BMI values. In most years the minimum BMI value in the sample is between 13 and 16, which indicates that values less than 10 are exceptionally low.

Some cases were included that do not have a “valid” BMI value. These are cases where the weight of the person was high and could not be measured directly as it exceeded the limit of the scales used by the interviewer. The weight used in the HSE is then based on a self-reported weight. Such cases have a BMI recorded in the “BMIval” or “BMIval2” variable but not the “BMI” variable in the HSE data. The cases included are those where BMI = -1 (i.e., no BMI value recorded) and BMIval > 0 (for 1997-2011) or BMIval2 > 0 (for 2012-2014). The BMIval variable was added to the HSE data in 1997 and is for estimated weights over 130 kg which was the weight limit of the scales up until 2010. From 2011 the scales had an increased weight limit of 200 kg. In 2012, the BMIval2 variable was added for estimated weights over 200 kg. A value being recorded in these variables indicates that the weight is considered reasonably accurate. It is important to include these cases to get data at the end of the right tail of the BMI distributions. Including these cases also follows what was done for the Trend Table (Section 2.5). Some analysis of these cases is in Section 2.6.

The HSE data set for 1992 has some extra data at the end of the file just consisting of BMI values and no other variables. There are 398 of these values. From observation of these values it was discovered that they are the BMI values from the end of the actual cases immediately above these values and so they are repeated values. Therefore, they should be ignored. As these cases do not have a value for the variables for bmiok, age, or sex they were automatically not included in the valid cases anyway, and so no adjustments were required in the analysis for this data.

2.2.2 Variables for each case

For each valid case, the data needed is the age, the BMI value, and the case weight. The age and the BMI value are just taken directly from the HSE data. The calculation of BMI in the HSE is weight in kg divided by height in metres squared (Section 1.2.1). Hence the units are kg / m^2. The HSE data includes the values for weight and height that are used for BMI. The HSE has the weight value recorded to 1 decimal place of kg and height recorded to 1 decimal place of centimetres (i.e., they are presumably measured or rounded to the nearest 100 grams and the nearest 1 mm). The BMI value is calculated from these values.

An interview weight is also given for each case in the years 2003-2014 (Section 2.1.1). The interview weight values were adjusted to ensure that the mean case weighting value is 1 for the year (explained and discussed further in Section 3.2). Otherwise, the year would have more or less emphasis when several years are combined simply due to the mean case weight being higher or lower than for the other years. The adjustment was done simply by dividing each interview weight value by the mean weighting value for the year. There is no interview weight variable for 1991-2002 and so each case for these years has a case weighting of 1.

One aspect that was apparent from the analysis is that the interview weights vary more in recent years. The standard deviation of the adjusted case weights is higher for each year of 2009-2014 (mean of the standard deviations for these years of: men 0.41, women 0.34) than for any of the years 2003-2008 (mean of the standard deviations for these years of: men 0.28, women 0.22). The maximum case weight values are also much higher after 2008. The reason for this difference in the amount of variation in HSE interview weights is not known. It could be that there is more variation in the types of cases sampled, in the population itself, or that more factors are being taken into account in calculating the interview weights.

For the analysis that is done here the cases are combined together rather than looked at individually, and they are used in producing population distributions and statistics. Some earlier initial analysis produced the BMI distribution for 2013 using the same method as in this report but giving the cases equal weighting. There is very little difference between those results and the results in this report. Some of the analysis in this report was done with different case weightings, including some comparisons using equal weight (Sections 7.2 and 15.2.3) and again there is not much difference in the results. Therefore, for the particular type of work done here in looking a population level data the interview weights appear to make little difference. Hence, there is no particular concern for the results here that the data for 1991-2002 does not have the interview weight variable.

2.3 Descriptive statistics for HSE data for each year

One focus of the work described in this report is to estimate population BMI distributions using batches of four years of HSE data. The initial stage was to extract the HSE data for each year, as explained in Section 2.2, and as part of this some descriptive statistics were calculated. The mean, median, sample standard deviation, sample skewness, and category percentages were calculated for each year. The interview weights were used in these calculations. The statistics calculations are explained in Appendix 2.I (“Statistics” subsection), and the values for each year are given in Appendix 2.II.

Figure 2.1 plots the mean and median values for men and for women for each year, and Figure 2.2 plots the percentages in the healthy BMI category ($18.5 \leq \text{BMI} < 25.0$). There is a general trend of the BMI statistics getting worse from 1991 to 2014 with the mean and median increasing, and the healthy percentage decreasing. Looking at the charts, the main patterns are a steep approximately linear trend from 1991 to 2001, with much smaller changes since then.

The subsequent analysis in this report is more detailed in looking at the complete shapes of the BMI distributions over time and how the distributions change. BMI data distributions were plotted for each individual year but the results are not included in this report as the focus of the work is on generating smoothed population distributions from four years of data, so as to use larger sample sizes and a smoothing method (Chapter 3).

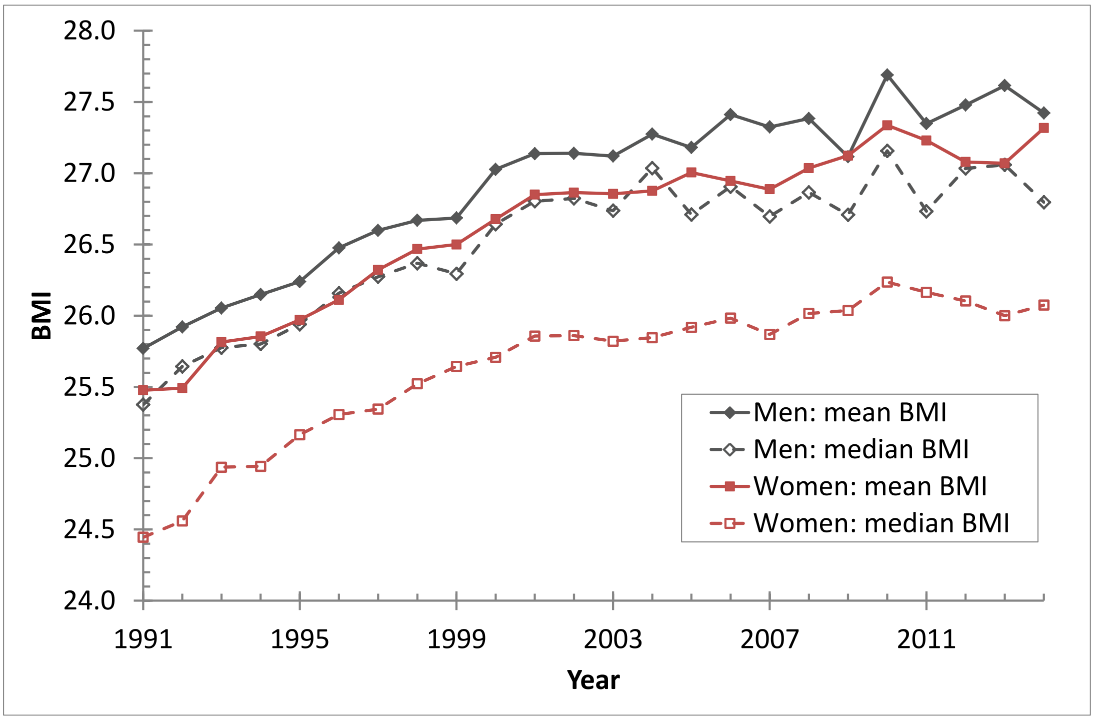


Figure 2.1 Mean and median BMI values for the HSE data for each year.

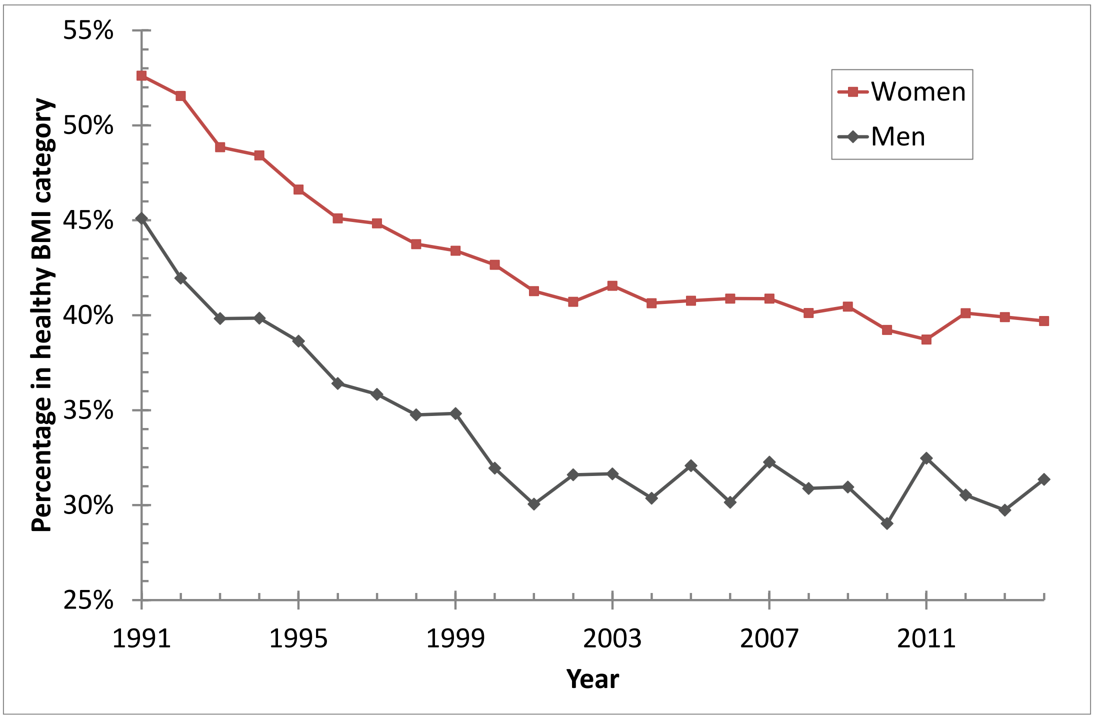


Figure 2.2 Percentage in the healthy category for the HSE data for each year.

2.4 Missing values

There are some missing BMI values in the HSE data. These are mainly cases where height or weight measurements are missing because either the person declined to be measured or it was not possible to do the measurements. For example, the HSE guidelines advise against doing the measurements when the person is unsteady on their feet, which will tend to apply more often for older people. There will be some cases where the weight of the person exceeded the limit for the scales and a suitable estimated weight was not obtained.

There are also some cases where a BMI value is given in the HSE data but is coded as "Not usable" because at least one of the height and weight variables was recorded as "Not usable". The precise reasons for this are not clear but these values are not considered valid cases and are not included in the analysis.

Overall, across all the years of 1991-2014 there are 108,895 possible cases for men (i.e., all cases age 18 and over in the main sample). The number used is 96,938 and so the overall missing percentage is 11.0%. For women there are 132,773 cases of age 18 and over in the main sample. However, these include 2,515 pregnant women who are outside the scope of the work as a valid BMI cannot be calculated for them because their weight includes the weight of the baby (Section 1.5). Hence there are 130,258 possible cases with 113,952 actually used giving a missing percentage of 12.5%.

As discussed in Section 2.1, the later HSE years have an interview weight to adjust for non-responses. However, this only adjusts for complete non-responses where the person did not respond to the survey at all, rather than for missing values for individual questions. Therefore, an analysis was done to look at the characteristics of the missing BMI values to see how they might affect the BMI data. When analysing each HSE annual data set, the number of possible cases and the number of used cases were totalled for each age. From this the missing percentage was calculated for each age. Analysis was then done to compare each year and each age as described below in Sections 2.4.1 and 2.4.2.

2.4.1 Missing values for each year

The missing percentage for each year was calculated as the weighted average of the missing percentages for each age weighted by the standard 2011-2014 population age profile used in this work (explained in Section 3.3). This means that a consistent mix of ages is used for all the years so as to remove any effect of differences in the age profiles for each year.

The results are in Figure 2.3 with the annual missing percentages being between 5% and 18%. There is some variability from year to year but there is a general trend of a slight increase in the percentage of missing values over time.

A noticeable and surprising aspect of Figure 2.3 is the very similar pattern of the two series for men and for women. The correlation between the missing percentages for men and women is extremely high at 0.94, with a scatter chart showing the association in Figure 2.4. This is a very strong pattern. It is not clear what the reason is for this strong association. Perhaps there is an interviewer effect with interviewers in some years being more effective and persuasive than in other years in getting the height and weight measurements from everyone, and so response rates are relatively high or relatively low for both men and women in a particular year. Perhaps there is a household effect with a tendency of either none or all in the household (i.e., both men and women) to agree or disagree to the measurements, with a random effect of more households agreeing to the measurements in some years than others. Perhaps there were slight changes in policy or emphasis from year to year regarding the situations in which measurements should or should not be taken, although I am not aware of any such changes. The one year that is slightly different to the others is 1993 being the only year with more missing values for men than women.

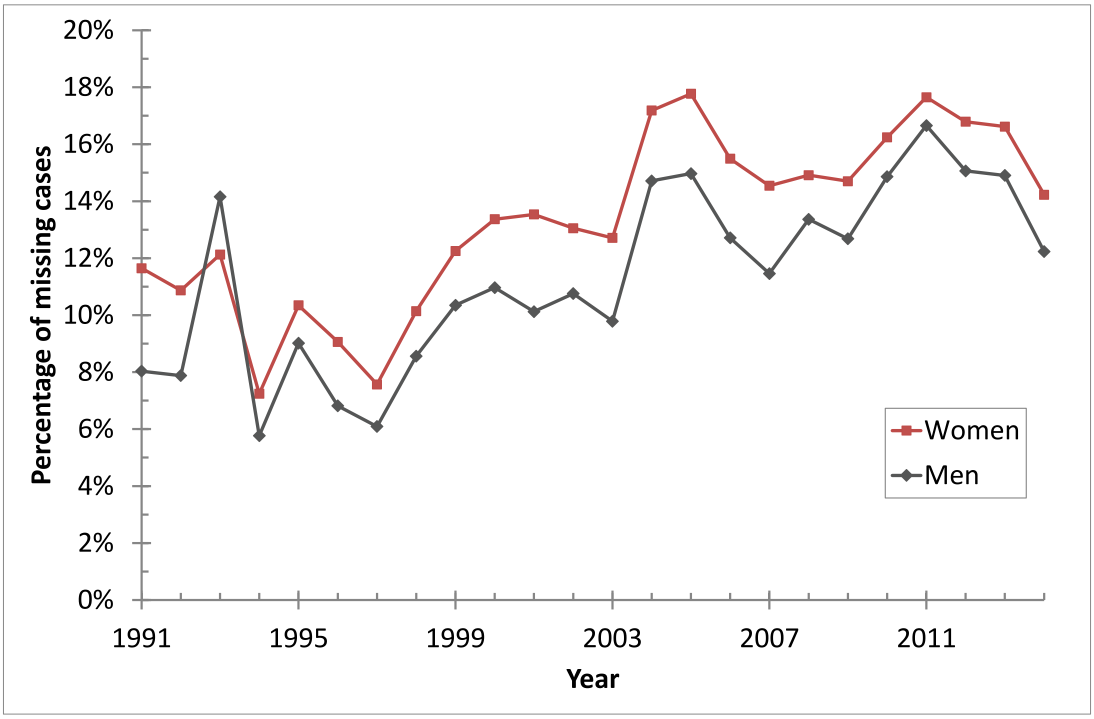


Figure 2.3 Percentage of missing values for each year.

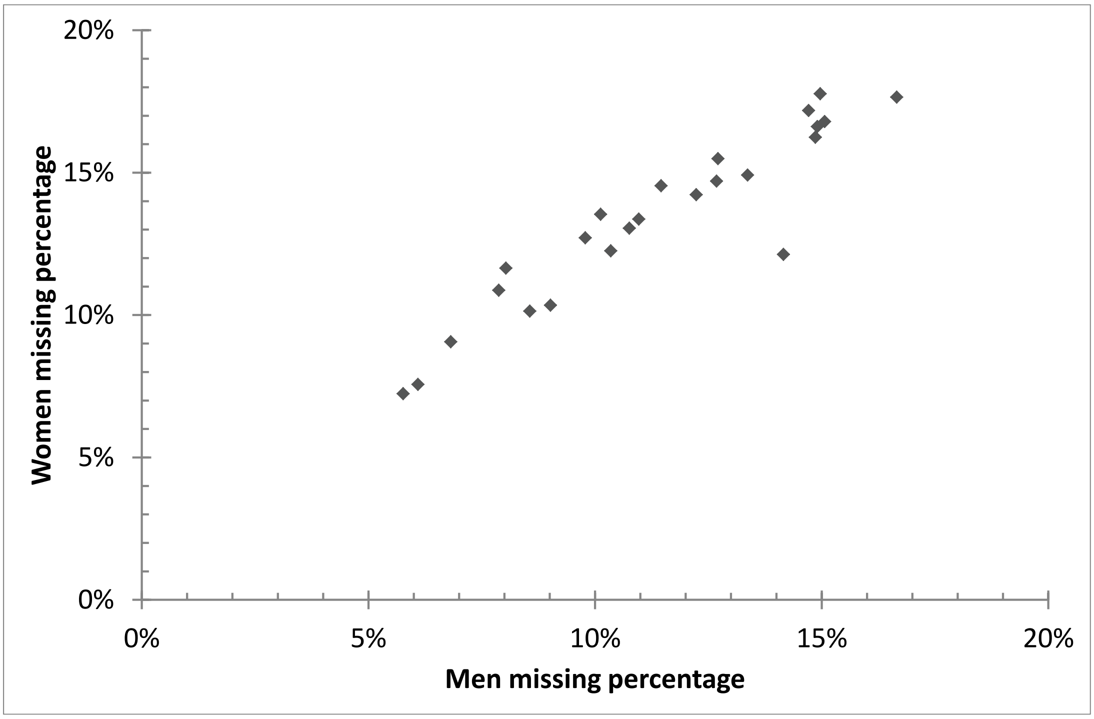


Figure 2.4 Relationship between the missing percentages for men and women for each year.

2.4.2 Missing values for each age

The missing value percentages were also compared by age. The overall percentage of missing values was calculated for each age in two ways. One way was to do the calculations using the total cases for each age across all years which treats each case equally. The other way was to calculate the average of the missing percentage values for each year which gives each year equal weighting. Unsurprisingly, both ways produce very similar results.

The chart using the total cases approach is shown in Figure 2.5. Note that age 90 is the total for ages 90 and over, since the age data in the HSE for 2014 is given in this way with ages of 90 and over grouped together in one category. The percentage of missing cases is roughly constant up to an age of about 65 and then increases in an exponential pattern above that. The increase above 65 is presumably due to not being able to measure some older people because they are unsteady on their feet. For ages up to 65 there is very little difference in the missing percentages for each age, and so across this age range there does not appear to be any big difference in the willingness of people to be measured or in the difficulty of obtaining measurements.

The age profile of missing cases for each year is compared in the charts in Appendix 2.III, where the lines are 15 point weighted moving average values to smooth the data. The general age profile is similar for each year. Some years have a relatively low or high percentage of missing values as shown in Figure 2.3 but generally this is reflected in the age profile curve for the year being higher or lower but with a similar shape. The year of 1991 is slightly different in having some of the highest missing percentages for older people but amongst the lowest percentages for younger people. This was the first year of the survey and has the smallest sample size. In particular, the number of older people in the 1991 survey is very small, and the numbers of men and women over 65 are just 275 and 375 respectively.

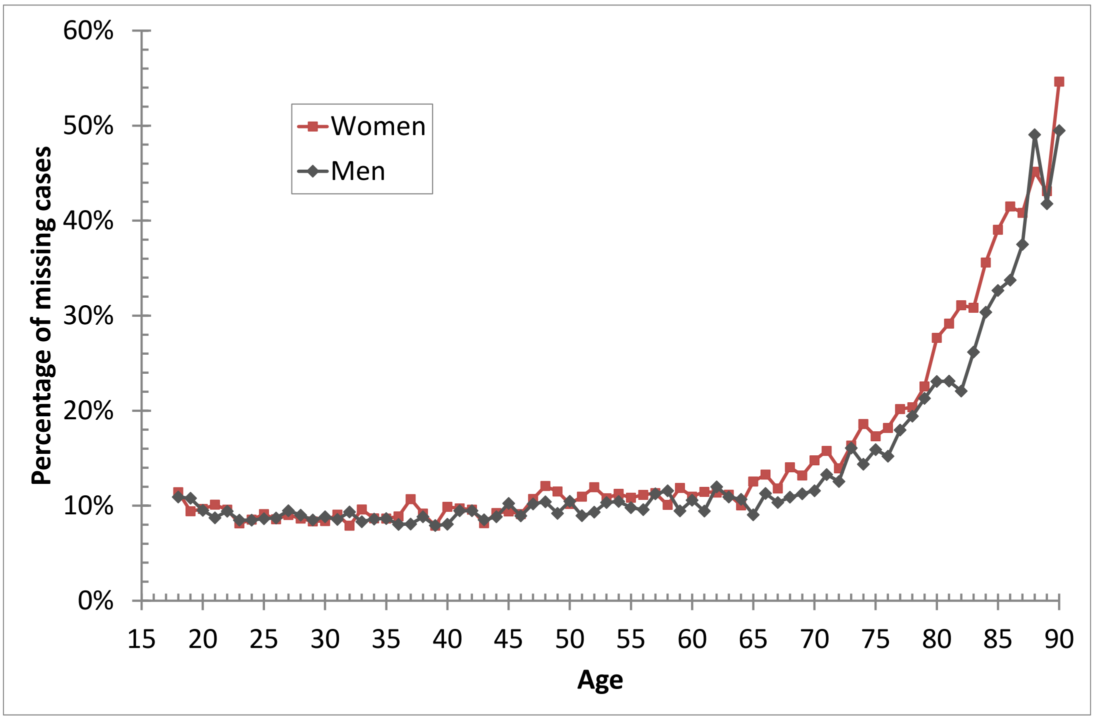


Figure 2.5 Overall missing percentages for each age for all the data.

2.4.3 Missing values discussion

The main aim of the missing value analysis was to consider whether the missing values might bias the results. The work in this report is focussed on the changes over time and so the main consideration is whether the data collection is similar over time. There is a slight increase in the percentage of missing cases over time but apart from that the pattern for each year seems fairly similar and so there are no major concerns as to the effects of the missing values on the data analysis.

Without having an estimate of the BMI of the missing values or knowing the reasons or the circumstances it is hard to assess specifically what the effect is of the missing values. If the proportion and age profile of missing cases is the same for each BMI value then the missing cases will not have any effect at all on the overall BMI distributions produced in the following chapters.

Speculatively, it could be that there is more chance of a missing value for cases with higher BMI values since people with a higher BMI might have more reluctance to be measured. This is also suggested by Wardle and Boniface (2008) in discussing missing cases in the HSE data that they used (final paragraph on page 532 of their paper) . As BMI is generally increasing over time this might explain the increase in the percentage of missing cases over time shown in Figure 2.3. If this is the case then the BMI distributions in this report may be slightly different to the true population distributions by underestimating the proportion of higher BMI values. BMI has increased over time and so if this is the case it will also have the effect of a small underestimation of the change in the BMI distributions over time.

There are more missing values for older people. Again, it is difficult to know the exact effect of this. If the missing values are spread across the BMI values then the result would just be to reduce the sample sizes for older people. This would not affect the BMI distributions as the analysis has adjustments to the case weights to match the age distribution in the population (Section 3.3). There could be more missing values for higher BMI cases which, as discussed above, would have some effect on the distribution.

2.5 Comparison with the Trend Table

As described in Section 1.2.5, an Excel file is produced each year as part of the HSE publications giving statistics from the HSE data. The latest year of HSE data used in the analysis described in this chapter is 2014. The statistics produced for the HSE for 2014 are called “Trend Tables” and are published on the NHS Digital website [19]. This website page also has an accompanying Trend Tables Commentary PDF document that is authored by NatCen Social Research and University College London (UCL).

Section 1.2.5 gives the website and link for the more recent HSE data for 2019 and describes the BMI statistics provided in the Excel file. In 2019 the term “Trend table” is not used, and instead the statistics files are called “Data tables”.

The Trend Table spreadsheet includes a worksheet for BMI statistics for each year of the HSE since 1993. For the 2014 spreadsheet this is Table 4. My analysis follows the same general approach in choosing and weighting the cases that is used for the Trend Table statistics. One of the checks done as part of the analysis here was to try and reproduce the Trend Table figures. This was to provide a check that the correct data and the correct case weightings were being used for the work here.

The statistics given in the spreadsheet are explained in Section 1.2.5. The specific statistics used for checking the data were those for all men, all women, and all adults and the statistical values compared were the mean BMI, the percentage of people in each BMI category, the number of actual cases, and the total interview weight of the cases.

The Trend Table approach is understood to be basically the same as set out in Section 2.2 in the way that the cases are chosen. The main difference is that the Trend Table uses ages 16 and over. Also, all valid cases were used for the Trend Table whereas five cases which have BMI < 10 were not included here (Section 2.2.1). Therefore, reproducing the Trend Table statistics involved calculating the statistics for ages 16 and over using the criteria for valid cases in Section 2.2 (except not excluding BMI values less than 10). In particular, the same basic spreadsheet formulae were used for the check as for the main analysis in the work here, just altering the parameter values for the criteria (minimum age 16 rather than 18, and minimum BMI 0 rather than 10).

Very good agreement was achieved with the Trend Table statistics. In most years the statistics match exactly, and if there were differences then these are extremely minor. The detailed results are given in Appendix 2.IV which explains what the differences are. There is a very small difference in the number of cases in some years but this amounts to a total of just six cases in the Trend Table data that are not in the data used here, and one case in the data here that is not in the Trend Table data. Therefore, this gives confidence in the data used including the interview weight values, and in the method used for identifying valid cases.

2.6 Weight limit of the scales

One issue with the HSE data is that the scales for weighing the person have a weight limit. This was originally 130 kg up until 2010, and then from 2011 was 200 kg. As explained in Section 2.2.1, the HSE data includes the variable BMIval from 1997 onwards and the variable BMIval2 from 2012 onwards. BMIval records BMI values based on estimated weights over 130 kg. BMIval2 records BMI values based on estimated weights over 200 kg.

As set out in Section 2.2.1, the valid cases include BMIval cases for 1997-2011 and BMIval2 cases since 2012. BMIval cases are not included after 2011 because the scales can then weigh over 130 kg. This approach follows that of NatCen and UCL with their Trend Table (Section 2.5).

An analysis was done of the percentage of cases with a weight over 130 kg in the data by simply adding the case weight of such cases each year and calculating this as a percentage of the total sample size of the data set. The results are shown in Figure 2.6. There are some differences in treatment over time. Before 1997 there was no BMIval variable and the cases in 1991-1994 that have a weight value

[19] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2014-trend-tables Excel file link: Health Survey for England, 2014: Trend Tables – Adult tables

over 130 kg are simply recorded as actual weights and valid BMI values. It is not clear exactly what these cases are or how they were measured. Perhaps the scales can show values above 130 kg but they are not considered accurate. In 1995 and 1996 there are no cases presumably because of specific instructions not to include weight measurements over 130 kg. The BMIval variable started in 1997 and all values over 130 kg from then until 2010 are from the BMIval variable. From 2011 onwards the scales were able to measure up to 200 kg and so most cases up to this weight are just those with a valid BMI variable value (Section 2.2.1). BMIval cases are still included for 2011, in line with the Trend Table treatment but there are only two cases (both in the data for women). For 2012-2014 the BMIval variable was not used for valid cases since the BMIval2 variable was then available. For these years the BMIval2 values are included to add in cases with a weight over 200 kg, although across the three years there are only two such cases (both in the data for men).

Figure 2.6 shows a fairly consistent trend of the percentage of cases increasing gradually over time. The percentages are quite high in 1991 but the sample sizes are very small in this year (as given in Table 3.1 in Chapter 3) and the number of cases was just four men and six women. Having no cases in 1995 and 1996 does mean that the right tail of the BMI distribution will be slightly reduced in those years. The main encouraging feature of Figure 2.6 is that the values for 2011-2014 when the new scales were in use are consistent with the trend and with the immediately preceding years – i.e., there is not a visual step change in the data. It also provides good support for using the BMIval variable in earlier years in that there is a fairly consistent trend across all the years for both men and women. Overall, this gives confidence that there does not appear to be a major effect with the change in the scales.

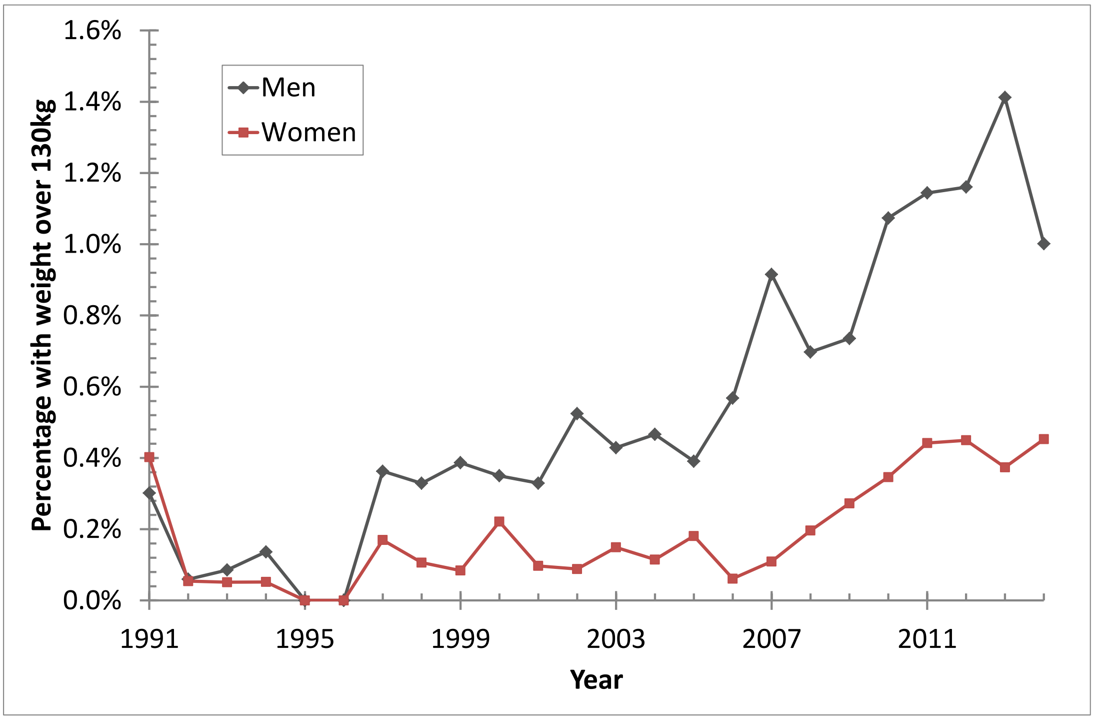


Figure 2.6 Percentage of cases with weight over 130 kg.

2.7 High BMI values

One noticeable feature of the changes in the data over time has been the considerable increase in the prevalence of high BMI values. Some further analysis was therefore done focussing on these high BMI values.

To look at the extremes, BMI values of 40 and over for men and 45 and over for women were analysed for each year by simply calculating the percentages in the HSE data (using the case weights). These values were chosen to give reasonably similar percentages across the years. The results are plotted on the chart in Figure 2.7. Obviously, a lot of these values are those with a weight over 130 kg, as discussed in Section 2.6, and so reflect the methodology for those cases. However, the groups here are larger in number particularly for women and so include many cases measured and recorded in the standard way in the HSE.

The percentages in Figure 2.7 go up to about 1.8% but the actual numbers are small and so it is not surprising that there are fluctuations from year to year. Nevertheless, both series show a clear trend of an increasing percentage that looks approximately linear, with the trend increase being larger for men than for women.

More detailed analysis of high BMI values is done in Chapter 4 in Section 4.5, although using slightly different BMI cut-off values to give larger sample sizes.

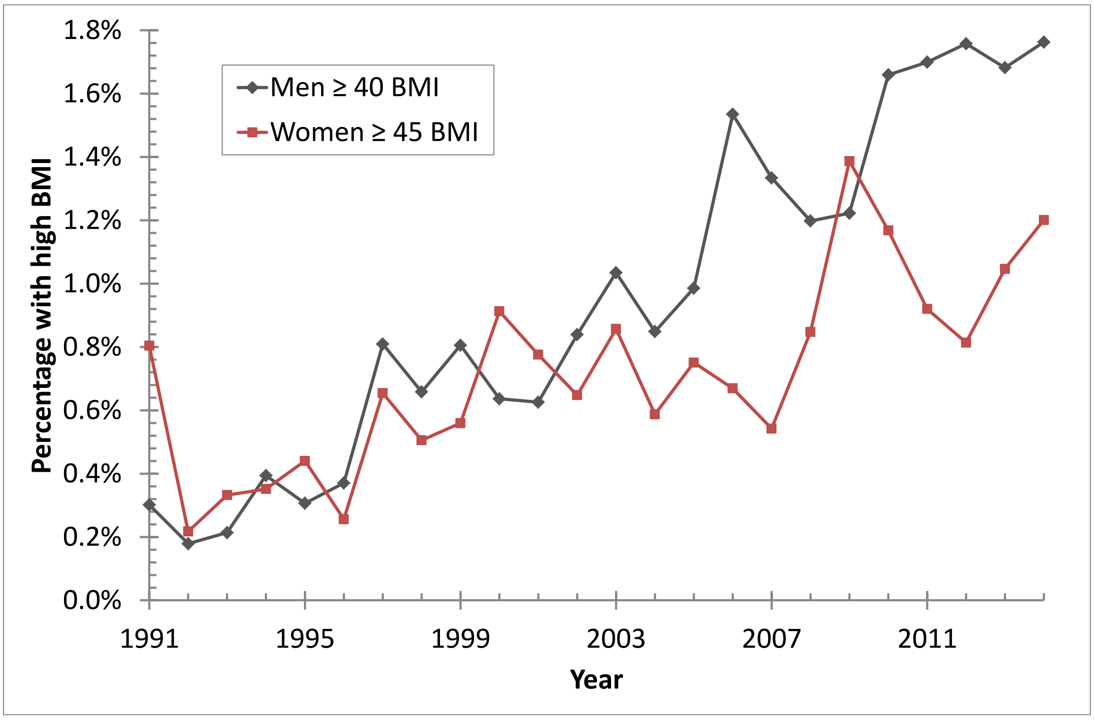


Figure 2.7 Percentage of high BMI values over time.

Chapter 3: Obtaining the BMI distributions

Key points from this chapter:

- The HSE data is combined in groups of four years, and is adjusted so that each group has the same age profile.
- The final estimated population distributions are produced by grouping the data in BMI intervals of width 0.5 and smoothing with the Savitzky-Golay method. This method is very good at smoothing the data as it maintains the shape and ensures that the mean and standard deviation stay the same.
- This gives population BMI distributions for 1993, 1997, 2001, 2005, 2009, and 2013. The distributions are analysed in Chapter 4.

3.1 Approach for estimating population BMI distributions

Chapter 2 has described the process for obtaining the BMI data for each year of the Health Survey for England (HSE). The next stage in the analysis was to use this data to estimate the population BMI distributions at different points in time.

A population BMI distribution gives the percentage frequency of the population who have BMI values within each given BMI interval. Intervals of width 0.5 are used here. One of the main aims of the work here is to produce BMI distributions using narrow intervals, to give detailed information of how BMI varies in the population.

The BMI distribution therefore consists of a set of intervals for BMI and a percentage frequency for each interval. To show this as a chart, this can be plotted as a histogram or a frequency polygon. A frequency polygon is used throughout this report, where the percentage frequency is plotted as a point at the mid-point of the interval. The points are connected with straight lines.

In the charts the y value for each point is a density. This is because the frequency is actually given by the area for the interval (the area of the bar for a histogram or, equivalently, the interval width multiplied by the density for the polygon). The density values on the y-axis are expressed per 0.5 BMI interval in the charts here. Therefore, the y values are the percentage frequency in the interval if the intervals are of width 0.5, which applies to most of the charts here. For the scaling transformations in Chapter 5, some of the interval widths are wider and so the y values are adjusted accordingly, as explained in Section 5.2.

This is very similar to, and an approximation of, a probability density function. A probability density function can be thought of as the limit of the percentage frequency distribution as the interval widths reduce to zero. It is a continuous function where the probability or percentage frequency in any given interval is the area under the curve across the interval. This is given by the integral of the function over the interval.

This chapter explains the approach used for producing the population BMI distributions. The process starts by combining four years of HSE data to give a larger sample size. Section 3.2 explains how the data is combined in terms of how each case is weighted, and also gives the sample sizes. Population adjustments are then made to use a consistent age profile and this is explained in Section 3.3. The age profile is used throughout the work in this report. The data is grouped into BMI intervals of width 0.5 to give a frequency distribution, and the BMI intervals used are stated in Section 3.4. Smoothing is then applied to give the final distribution. This is an important part of the process and the smoothing method used is described in Section 3.5. Then, Section 3.6 has a couple of charts comparing the data distribution and the smoothed distribution. The resulting distributions are analysed in Chapter 4.

3.2 Sample sizes and combining the years

The data used initially was from the start of the HSE in 1991 until 2014. Combining the data in groups of four years gives six grouped data sets: 1991-1994, 1995-1998, 1999-2002, 2003-2006, 2007-2010, 2011-2014. In this report the groups are denoted “1993”, “1997”, “2001”, “2005”, “2009”, “2013”, as the resulting distributions are estimates of the population BMI distributions for those years. By the time of the final version of the report, HSE data had become available for 2015-2018 which enabled distributions to be produced for “2017”, and this is covered in Chapter 14.

Table 3.1 shows the number of cases used for each group of four years and shows how they are split between each year for men and women. The group for 1997 has quite a lot more cases than the other groups, with the sample sizes for other five groups being fairly similar (between 12432 and 16625 for men, and between 15076 and 19557 for women).

There are two main ways in which the different years could be combined. The first method is to alter the case weights so that each year has the same overall weighting. The case weights produced from the HSE analysis described in Section 2.2.2 have a mean of 1 and so this requires multiplying each case weight by: (total sample size of the four year group) / (4 × sample size of the year). In other words, cases from a year with few cases get a higher case weighting and cases from a year with many cases get a lower case weighting.

The second method is to give the cases equal weighting (apart from the non-response interview weight). In this option the case weights already calculated would be used as they are.

The limitation of the first method is that cases from years with few cases get extra emphasis simply because of the low number of cases in the sample for that year. The limitation of the second method is that the cases are not weighted evenly across the time period. The choice is subjective but the second method was chosen to avoid cases from years with few cases (such as 1991 and 1992) having much greater emphasis. In other words, the case weights already calculated for each HSE year were used directly. As explained in Section 2.2.2, these are simply the HSE interview weights adjusted so that the mean case weight for the year is 1. A population age profile adjustment is then made as described in Section 3.3.

Consideration can be given as to exactly what point in time the groups of data represent in terms of the mid-point of the data. The HSE data is collected continuously throughout the year and so the mid-points of the four year time intervals are the start of years used to denote them. For example, the mid-point of 2011-2014 is the start of 2013. Therefore, the date of each distribution can best be thought of as corresponding to the start of the particular year. The number of cases does vary from year to year and this affects slightly how the cases are spread out across the time period.

A mid-point can be calculated as the average of the mid points of each of the years (i.e., 1991.5, 1992.5 etc.), weighted by the number of cases in the year. The only four year period where this gives a value more than 0.1 years away from the start of the third year (i.e., 1993.0, 1997.0, etc.) is the first time interval. There are many more cases in 1993 and 1994 than 1991 and 1992, giving mid-points of 1993.6 and 1993.7 for men and women respectively. In other words, the first four year period relates more to a mid-point date roughly in the middle of 1993 rather than at the start of 1993.

Table 3.1 Number of HSE cases used for 1991-2014.

Year	Men	Women	Total
1991	1326	1492	
1992	1678	1841	
1993	7021	7824	
1994	6600	7681	
"1993"	16625	18838	35463
1995	6525	7497	
1996	6751	7816	
1997	3583	4126	
1998	6382	7519	
"1997"	23241	26958	50199
1999	3104	3574	
2000	3143	3615	
2001	6073	7218	
2002	2861	3395	
"2001"	15181	17802	32983
2003	5787	6882	
2004	2365	3050	
2005	2819	3312	
2006	5359	6313	
"2005"	16330	19557	35887
2007	2633	3132	
2008	5660	6782	
2009	1790	2061	
2010	3046	3746	
"2009"	13129	15721	28850
2011	3099	3779	
2012	3032	3696	
2013	3245	3893	
2014	3056	3708	
"2013"	12432	15076	27508
All years	96938	113952	210890

3.3 Population age profile adjustment

The age profile of the population varies over time. One strong feature is a sharp peak in numbers for the post war baby boom for people born in 1946 and 1947. There is also a broader peak for those born in the second baby boom in the mid-1960s. These peaks obviously move one year later in the population age profile as each year passes. Since BMI tends to increase with age (examined in Chapters 7-10), part of any changes in the BMI distributions could be due to the change in age profile, although any effect is likely to be small. To eliminate this factor an adjustment was made to the data to give each data set the same age profile.

Any age profile could be used but the one chosen was the average profile for the last time interval of 2011-2014. The mid-year population values were obtained from spreadsheets provided by the Office of National Statistics in the website given in the footnote [20], and the data used can be obtained from the Excel file link as follows: drop down arrow for “Mid-2001 to mid-2019 detailed time series”, then below “Supporting files” select “UK population estimates, 1838 to 2019”. The data is in Table 11 of the file (“Table 11: Population estimates for England, by sex and single year of age, mid-1971 to mid-2019”).

The population values for each year for England for 2011 to 2014 were extracted and the percentage for each age amongst the adult population was calculated for men and for women for each of the four years. The percentages for each year from 2011 to 2014 were then averaged. The total population values for each of these years are very similar and an alternative method of using the sum of the population values for the four years gives results with negligible difference.

The age profiles used are shown in Appendix 3.I. Both the population data and the HSE 2014 data give ages by year up to age 89, with “90” being ages 90 and over. Therefore, the analysis used ages 90 and over as a single category. The HSE before 2014 does give actual ages but ages of 90 and over were grouped into one category in the analysis.

The adjustment for the age profile for the data sets was done by altering the case weights. For a given four year grouped data set the percentage of each age in the data set was calculated. This is the total case weights for the age divided by the total case weights (= total sample size). The case weights were then each multiplied by a factor of: target age percentage / actual age percentage. For example, for men the 2011-2014 population age profile has 1.65% of adults being age 18. The 2011-2014 HSE data has 1.75% of age 18, and so the case weight of every case with age 18 was multiplied by 1.65 / 1.75 (although more decimal places were used in the spreadsheet). This rescales the case weights of each age so that their percentage equals that of the 2011-2014 population. The sum of the case weights still equals the total sample size of the four year data set and so the mean case weight is still 1. There will be an effect on the relative weighting of each year within the data set depending on differences between the age profiles in the data for each year, although any effect is likely to be small.

The main patterns in the age adjustment factors are generally to increase the case weighting of cases of ages in the 20s and reduce it for ages in the 30s. In other words, the HSE data has fewer cases in age 20s and more cases in age 30s than the population profile for 2011-2014. Since BMI tends to increase with age this has a small effect on the distributions and the statistics. For example, it slightly reduces the mean in each of the four year sets of data. The values for the mean BMI of the population data after age adjustment compared to the data before age adjustment for the six population years for men and women are reductions in mean BMI of between 0.01 and 0.12 with the average reductions for both men and women being 0.05. This is a very small change.

[20]https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland

3.4 BMI intervals for the relative frequency distribution

A frequency distribution of the BMI data was produced using BMI intervals of width 0.5. To be consistent with the categories used for the usual BMI categories (Section 1.2.4), the lower end point was included in each interval giving, for example, intervals of: 18.5 ≤ BMI < 19.0, 19.0 ≤ BMI < 19.5. The total case weights after making the population adjustment in Section 3.3 were added up for each interval and then the percentage of the total was calculated. This gives a relative frequency distribution for the data consisting of the percentage frequency in each 0.5 BMI interval.

3.5 Smoothing the data for the final population distributions

3.5.1 Savitzky-Golay method

The relative frequency distribution for the data, after the stages described in Sections 3.2 - 3.4, consists of a percentage frequency value for each 0.5 BMI interval. The HSE data is a sample and so inevitably has considerable irregularities. BMI is a simple function of weight and height, which are physical quantities that differ from person to person and are on a continuous scale. Across a large population, such as adults in England, such a function would be expected to vary in a smooth manner. Hence, the final step in generating the estimated relative frequency distribution for the population was to smooth the data frequency distribution values.

The Savitzky-Golay method was used for all the smoothing for the work in this report. The method was presented in a paper by Abraham Savitzky and Marcel J. E. Golay in 1964 (Savitzky and Golay, 1964). This paper had some errors in the smoothing weight values that were corrected by Steinier et al. (1972).

The Savitzky-Golay method is often called the Savitzky-Golay filter. The approach taken in Savitzky and Golay (1964) is to use least squares to fit the best polynomial curve through a defined number of points. The middle value from the curve is then the smoothed value at the middle point. This is applied to each point in the data series to give the sequence of smoothed values. The method is applicable to data with constant intervals between the points on the x-axis, as is the case here where the points are all at a BMI distance of 0.5 from the next point. The method is extremely easy to apply. The mathematics of fitting the polynomial curve turns out to just require using a weighted moving average on the data, using the weights provided in the papers. There are also weights for obtaining values of the derivatives of the fitted smooth curve, but these are not used in the work in this report.

One important advantage of the Savitzky-Golay method compared to a simple moving average smoothing is that it maintains the patterns in the data better, particularly the peaks and the tails. A simple moving average will tend to flatten the data peaks and increase the heights of the tails and so will distort the shape to some extent.

A further advantage of the method is that the smoothed data has the same statistical values as the data that is being smoothed for the mean and for some of the higher moments of the data. Precisely which higher moment values are maintained depends on the Savitzky-Golay version. The quadratic / cubic Savitzky-Golay method was used in all the analysis here. Fitting either a quadratic or cubic polynomial gives the same values for the moving average weights for smoothing and these are the ones used, as given in Section 3.5.2. This version maintains both the standard deviation and the skewness value. Hence, each smoothed population distribution in this report will have exactly the same mean, standard deviation, and skewness value as the data distribution from which it is derived.

Where the frequency of the data distribution is equal to or close to zero at very low and very high BMI values, the smoothing method can sometimes produce a negative value slightly below 0 for the smoothed frequency because it is fitting a polynomial curve. The smallest value for all the population distributions for 1993 to 2017 is just -0.012%. Having a few negative frequency values of very small magnitude is not considered a problem. The values were left as they are so that the total percentage sums to 100% and the statistics for the mean, standard deviation, and skewness, including using the negative frequency values in the statistics calculations, are exactly the same as for the data distribution.

3.5.2 Choice of smoothing weights in applying the Savitzky-Golay method

As explained in Section 3.5.1, the quadratic / cubic version of the Savitzky-Golay method was used for all the smoothing in the work presented in this report. The choice that then has to be made is the number of points to use for the smoothing. The weights are given in Table I of Savitzky and Golay (1964) for between 5 and 25 points. The number of points needs to be odd so that the middle point corresponds to the point being smoothed and so the values are given for 5, 7, 9, … , 25 points. Corrections are given in Steinier et al. (1972) to Table I in the original paper for the values for 23 points and 25 points.

The spreadsheets used in the work in this report for producing the population distributions were set up to be able to easily apply any of the smoothing weights from 7 to 19 points. The values for these are given in Table 3.2. No corrections to these were applied by Steinier et al. (1972) and so all the values in Table 3.2 are from Savitzky and Golay (1964). The point index refers to the position in the series being smoothed. Index 0 is the position of the smoothed value. Next to it, -1 is the position before index 0 and 1 is the position after. Then, -2 is two positions before and 2 is two positions after, and so on for the other indices.

Table 3.2 Weights used for the smoothing (from Table I in Savitzky and Golay, 1964).

Point index	7 points	9 points	11 points	13 points	15 points	17 points	19 points
-9	0	0	0	0	0	0	-136
-8	0	0	0	0	0	-21	-51
-7	0	0	0	0	-78	-6	24
-6	0	0	0	-11	-13	7	89
-5	0	0	-36	0	42	18	144
-4	0	-21	9	9	87	27	189
-3	-2	14	44	16	122	34	224
-2	3	39	69	21	147	39	249
-1	6	54	84	24	162	42	264
0	7	59	89	25	167	43	269
1	6	54	84	24	162	42	264
2	3	39	69	21	147	39	249
3	-2	14	44	16	122	34	224
4	0	-21	9	9	87	27	189
5	0	0	-36	0	42	18	144
6	0	0	0	-11	-13	7	89
7	0	0	0	0	-78	-6	24
8	0	0	0	0	0	-21	-51
9	0	0	0	0	0	0	-136
Total	21	231	429	143	1105	323	2261

The smoothing is applied as a weighted average. This is done by taking the sum of y values multiplied by the weights, and then dividing by the total of the weights. Taking a simple example with the 7 point smoothing, suppose the data series includes these seven successive (x,y) values: (18.75, 0.54%), (19.25, 0.75%), (19.75, 1.02%), (20.25, 1.37%), (20.75, 1.54%), (21.25, 1.96%), (21.75, 2.31%). The x values are the mid-points of the BMI intervals (18.5 ≤ BMI < 19.0, 19.0 ≤ BMI < 19.5, etc.) and the y values are the frequencies for each interval for the data distribution. For the 7 point smoothing, using the weights from Table 3.2, the smoothed value for the middle point of x = 20.25 is given by the formula:

$$(-2\times0.54\% + 3\times0.75\% + 6\times1.02\% + 7\times1.37\% + 6\times1.54\% + 3\times1.96\% - 2\times2.31\%) / 21 = 1.30\%.$$

So, the smoothed distribution has the point (20.25, 1.30%). The other smoothed points would be calculated in the same way by applying the weights to the 7 points around each x value from three before to three after.

The same approach applies for smoothing with a different number of points. For example, the 11 point method uses the 11 points around each x value from 5 before to 5 after.

Fewer points mean that the smoothed values follow the data more closely. More points mean that the data is smoothed more. The choice is subjective but, from initial experimentation, the best choices were considered to be 11 points and 13 points. For example, it was felt that 7 points follows the irregularities in the data too closely whereas 19 points applies too much smoothing and alters the shape slightly by flattening the peak and increasing the tails.

Another way to get more smoothing is to apply the Savitzky-Golay method more than once. Any combination of weights could be used including applying the same choice two or more times. The option tried in the spreadsheets was double smoothing consisting of 11 points followed by 13 points. In other words, the 11 point smoothing was applied first. Then the 13 point smoothing was applied to the smoothed distribution from the 11 point smoothing.

For each population distribution described in Chapter 4, the smoothed distributions were produced for both 11 and 13 points, and for the double smoothing. All were considered suitable and to work well. It was therefore felt that any of these choices would be fine, and they are actually each used in different places in the work in this report.

For the distributions in Chapter 4, the 11 point smoothing version was used as this was considered to follow the general pattern of the data well and to give sufficient smoothing. As shown in Table 3.2, this has weights of -36, 9, 44, 69, 84, 89, 84, 69, 44, 9, -36, divided by the total 429. So, the smoothed value, s_i, for the *i*th point in the series is obtained from data values y_i as follows:

$$s_i = (-36y_{i-5} + 9y_{i-4} + 44y_{i-3} + 69y_{i-2} + 84y_{i-1} + 89y_i + 84y_{i+1} + 69y_{i+2} + 44y_{i+3} + 9y_{i+4} - 36y_{i+5})/429$$

The other options were sometimes chosen for the various distributions in this report. The method used is stated in each chapter. For example, the age group distributions in Chapters 8, 9, and 10 use the 13 point smoothing as these have less data and it was felt that a bit more smoothing was beneficial.

The double smoothing was used for the 2017 population distributions in Chapter 14 and also for the deprivation and height category distributions in Chapter 15. This method provides additional smoothing and can be a good choice. It tends to lower the peak very slightly compared to the 11 or 13 point methods but any differences in the distributions here are extremely small. For the 2017 distributions the 11 point smoothing was also analysed and the differences in the results between the double smoothing and the 11 point smoothing were very minor (as explained in Section 14.3 in Chapter 14). The additional smoothing provided by the double smoothing method was felt to be useful for the particular distributions in Chapters 14 and 15 and could be a good choice for future analysis.

3.6 Smoothing examples

Figures 3.1 and 3.2 show the BMI data distribution and the smoothed distribution for men and women for 2013, as examples, using the 11 point Savitzky-Golay smoothing described in Section 3.5. Charts for all six years are shown in Appendix 3.II. The charts show how the smoothing does well in both smoothing out the irregularities in the data and maintaining the general shape.

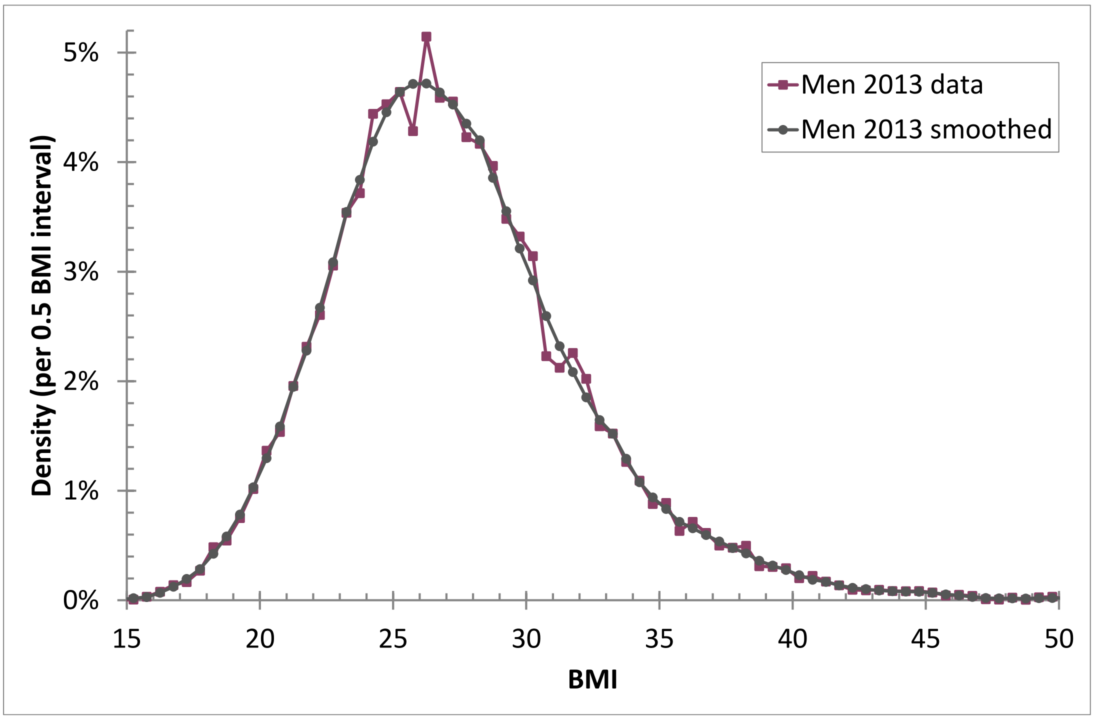


Figure 3.1 Data distribution and smoothed distribution for men for 2013.

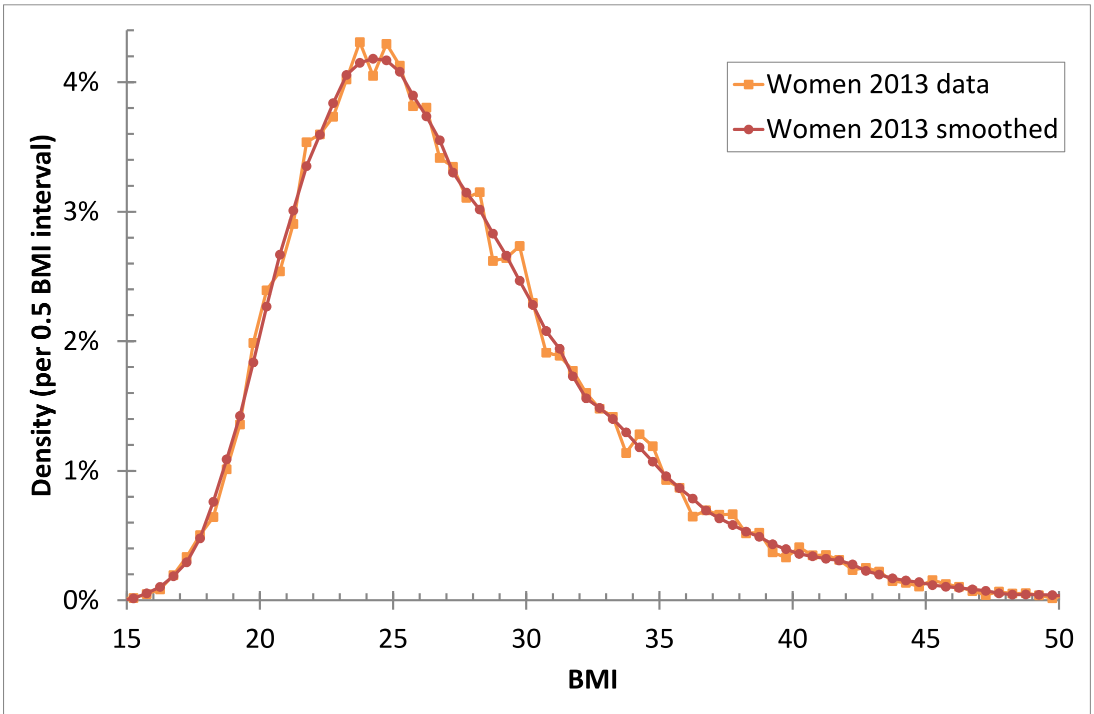


Figure 3.2 Data distribution and smoothed distribution for women for 2013.

Chapter 4: Analysis and comparison of the BMI distributions

Key points from this chapter:

- The BMI distributions produced from the HSE data by the methodology in Chapters 2 and 3 are analysed and compared. These are the estimated population BMI distributions for men and for women in England for 1993, 1997, 2001, 2005, 2009, and 2013. Charts of these distributions are in Figure 4.1 for men and Figure 4.5 for women, and show the shapes of the distributions and how they have changed over time. For both men and women the distributions are essentially a smooth curve with a single peak. They are positively skewed with a long right tail. However, the details of the distribution shapes are quite different for men and women, as discussed in the last two bullet points below.
- For both men and women, the pattern of the changes in the BMI distribution over time is for the left end of the distribution to stay basically the same and for the rest of the distribution to get more and more stretched out to the right. This means a worsening situation for BMI and obesity, but in a particular way where the pattern of the changes looks like a linear scaling transformation. The modelling of the changes using a linear scaling is explored further in Chapter 5. Changes in the BMI distributions over time are considered to represent how BMI has changed in equivalent people in the population through living in different times, and this interpretation is discussed further in Chapter 6.
- A variety of statistics are presented including basic descriptive statistics, BMI category percentages, percentiles, and the prevalence of high BMI values. Population BMI gets worse over time from one distribution to the next with, for example, the mean and median increasing and the healthy category percentage decreasing.
- Charts of the increases in the percentile values from one distribution to the next give approximately linear relationships. So, higher percentiles that are at higher BMI values in the initial population increase in BMI by more than lower percentiles at lower BMI values. Again, this indicates that a linear scaling is likely to be a good model of the changes over time, as considered further in Chapter 5.
- The increases in BMI are much greater between the early distributions than between the later ones. For example, the increases in the mean for the eight years from 1993 to 2001 for men and women of 0.85 and 0.89 are much greater than for the twelve years from 2001 to 2013 of 0.51 and 0.47. Hence, BMI increased at a faster rate in the early years of the HSE data in the 1990s than in the more recent years.
- The overall increases in BMI from 1993 to 2013 are substantial. For men, the mean BMI increases by 1.37 from 26.05 to 27.42, and the healthy category percentage reduces by 9.1% from 40.4% to 31.3%. For women, the mean BMI increases by 1.37 from 25.76 to 27.13 and the healthy category percentage reduces by 9.4% from 49.0% to 39.6%. The changes in these statistics are very similar for men and women. The percentile increases from 1993 to 2013 are also very similar for men and women, and have an approximately linear pattern with higher BMI increases for higher 1993 BMI values. For example, the 70th percentile BMI increases are 1.57 and 1.77, respectively. The 90th percentile increases are 2.65 and 2.72, respectively. BMI values and increases can be converted into weight values using average height and, for example, these 90th percentile BMI increases correspond to substantial weight increases of 8.10 kg (17.9 lbs) and 7.14 kg (15.7 lbs), respectively.
- The shapes of the distributions for men and women are quite different. For a given year, compared with the distribution for men, the distribution for women has a shorter peak at a lower BMI value. The distributions for women have a higher standard deviation and skewness than those for men. Nevertheless, the changes over time are very similar for the distributions for men and women.

- The different shapes mean that the distributions for women have higher percentages for low and high BMI values and lower percentages for middle values compared to the corresponding distributions for men. The percentage values are higher in the distribution for women for BMI values in most of its left tail and at the end of its right tail. Specifically for 2013, compared to the distribution for men, the distribution for women has 9.4% more with BMI < 24.0, 12.5% less with 24.0 ≤ BMI < 33.5, and 3.1% more with BMI ≥ 33.5. For the BMI categories, the total percentage being overweight or obese for each year is about 9% higher for men than for women and in this sense the general issue of obesity is more widespread for men. For the highest categories of obese II and severely obese, the percentages are about 3% higher for women and so the more specific issue of very high BMI values is more prevalent for women. Whilst there are these differences in prevalence, there still considerable numbers in all the overweight and obese categories for both men and women.

4.1 Overview of the analysis

The method described in Chapters 2 and 3 was used to produce BMI distributions for men and for women for 1993, 1997, 2001, 2005, 2009, and 2013. The distributions consist of 0.5 BMI intervals and a percentage frequency for each interval.

These distributions are the estimates from the HSE data of the population BMI distributions for adults aged 18 and over in England. In this and the next chapter they will simply be labelled and referred to as men 1993, women 1993, men 1997, women 1997, etc. This chapter compares these population distributions through charts and various statistics.

As explained in Section 3.1, the charts of the BMI distributions are plotted as frequency polygons (e.g., Figures 4.1 and 4.5). The percentage frequency for each interval is plotted at the mid-point of the interval, and the points are connected by straight lines. For a frequency polygon (as for a histogram), the frequency is actually given by the area for the interval which is the interval width multiplied by the density. The density values on the y-axis are expressed per 0.5 BMI interval in the charts throughout this report. Therefore, the y values are the percentage frequency in each interval if the intervals are of width 0.5, which applies to all the charts in this chapter. Charts of the distributions are shown and discussed in Section 4.2 for men and Section 4.3 for women.

Other types of charts are also used in this chapter. Cumulative distribution charts have the cumulative frequencies plotted at the end points of the intervals (e.g., Figures 4.2, 4.4, 4.6, 4.8), which is the usual format for this type of chart. As discussed in Section 1.6, the charts showing data for the different years use blue for 1993 and 1997, orange for 2001 and 2005, and purple for 2009 and 2013. These colours are used for the data for both men and women.

Descriptive statistics were worked out for the distributions in the usual way for such data. These are presented for men and for women in Sections 4.4 and 4.5 respectively. The mean, standard deviation and skewness were calculated using the mid-points of the intervals and the percentage frequency in each interval. To get a value for the median, linear interpolation was used within the interval that contains the median to estimate the 50th percentile value. The category percentages were obtained easily as the sum of the frequencies for the appropriate 0.5 BMI intervals since the intervals used tie in with the standard BMI categories.

Percentile values were also calculated for the distributions and are analysed in Section 4.6. These were also obtained by linear interpolation, with some examples of the calculations given in Section 4.6.1. Section 4.7 looks at high BMI values in the distributions. The distributions for 1993 and 2013 are compared and contrasted in Section 4.8 to look at the difference across the whole time period being considered. A comparison of the distributions for men and women is in Section 4.9. Some of the BMI values and BMI increases from 1993 to 2013 are converted into weight values in Section 4.10.

A relatively minor point is the precise date that the distributions correspond to. As discussed in Section 3.2, they can best be thought of as the start of the particular year. The slight exception is the 1993 distribution which is mostly derived from data in 1993 and 1994 (Table 3.1). The data used for this time interval has a mid-point near the middle of 1993. Therefore, when considering the changes between the distributions, strictly speaking the time from the 1993 to the 1997 distribution is slightly shorter than between the other distributions, being more like 3.5 years than 4 years.

The patterns of the changes in the distributions over time look like a linear scaling transformation. This is explored further in Chapter 5 where the changes are modelled using such a transformation. It is considered that changes in the BMI distributions over time represent the effect on equivalent people of living in different times, and this interpretation is discussed in Chapter 6 along with considering the various factors that might affect obesity. The results here are extended to distributions for 2017 in Chapter 14.

4.2 BMI frequency distributions for men

4.2.1 Shapes of the distributions for men

Figure 4.1 shows the BMI population distributions for men for the six years of 1993, 1997, 2001, 2005, 2009, and 2013 produced by the method set out in Chapters 2 and 3. The shape of each distribution in Figure 4.1 is a fairly smooth curve with a single peak at a BMI value of about 25 or 26. The distributions are reasonably symmetrical, although the right tail is a bit longer and more stretched out than the left tail and so the distributions are slightly positively skewed. Most BMI values are between 20 and 35.

The general pattern of the changes over time in the distributions is for the peak to become lower and the right tail higher. The peak also tends to shift slightly to the right over time. The start of the left tail is very similar for each curve whereas the right tail extends more to the right in each time interval. This reflects an increase in BMI where the frequency of the lowest values does not change much. Values around the peak of the distribution become less prevalent and high BMI values become more prevalent. This pattern of changes looks like a linear scaling transformation and this will be investigated further in Chapter 5.

The difference from one distribution to the next looks much greater from 1993 to 2001 than from 2001 to 2013. In other words, BMI increased a lot more in the HSE time period before 2001 than in more recent years. The biggest change visually from the charts is from 1997 to 2001. This is considered further in the next section.

The cumulative frequency distributions are shown in Figure 4.2. This chart helps to highlight the greater differences from 1993 to 2001, as well as the similarity of the left end of the distributions. The distributions for 2005, 2009 and 2013 are very close to each other for the lower part of the data for BMI values up to about 24.5 (about 30% of the data), with only fairly small changes for BMI values above that.

In Figure 4.1 the x-axis just goes up to a BMI value of 45 to help to see the detail of the main part of the curve. A chart going up to a BMI value of 50 showing more of the right tail is in Appendix 4.I. That appendix also has two separate charts of the distributions for 1993 to 2005 and 2005 to 2013 to see some of the details of the curves that are obscured where the lines overlap in Figure 4.1.

A table of values for all the distributions in Figure 4.1 (along with the distribution for 2017 in Section 14.4.1 in Chapter 14) is given in Appendix 14.III in Table 14.III.3.

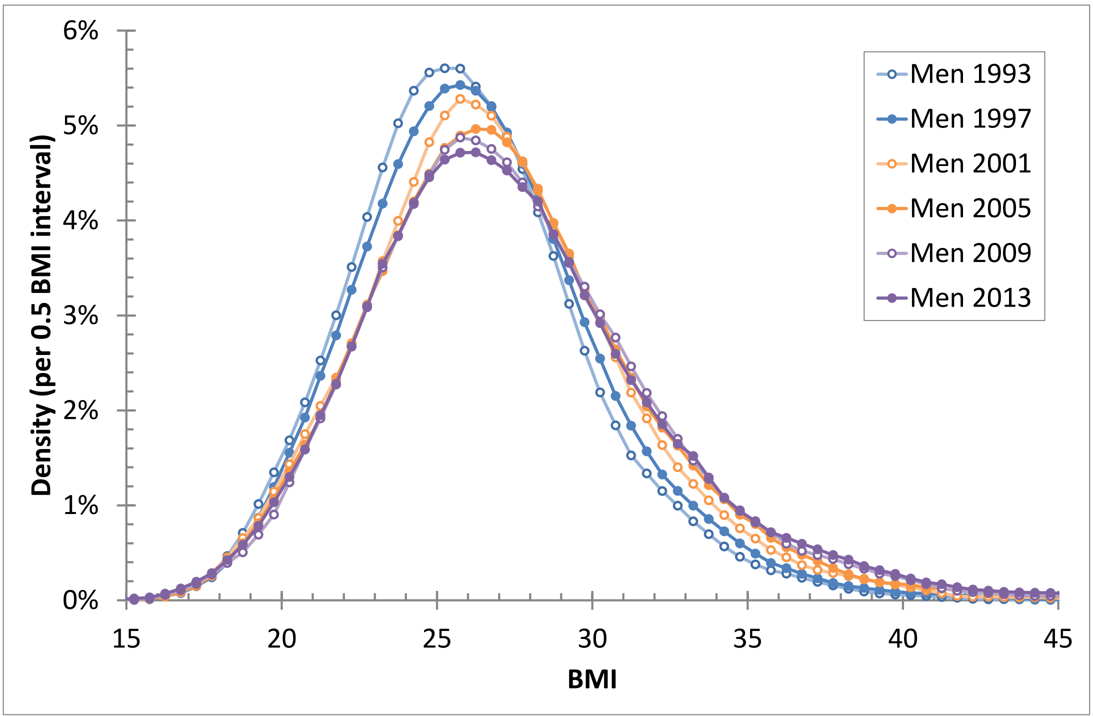


Figure 4.1 BMI population distributions for men for 1993 to 2013.

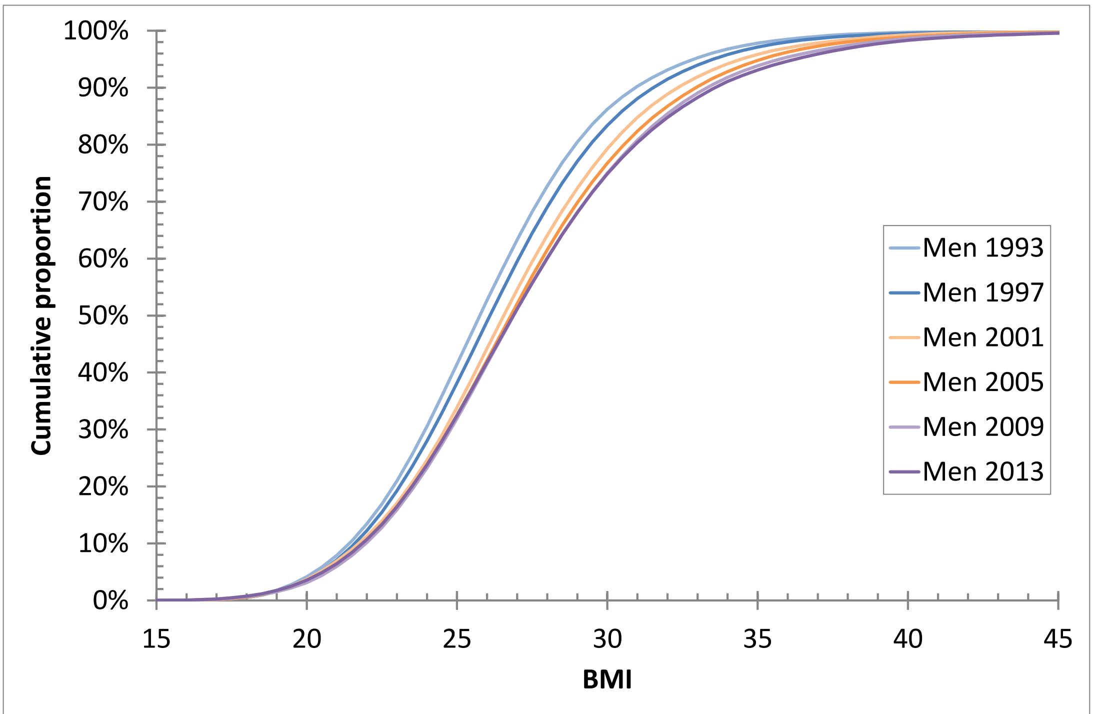


Figure 4.2 BMI cumulative frequency distributions for men for 1993 to 2013.

4.2.2 Differences between each year for men

The changes between successive population distributions were analysed by calculating the differences between the distributions. These help to show the detailed pattern and the relative magnitude of the changes.

The differences between the distributions are plotted in Figure 4.3. The series are labelled by the later distribution. So "men 1997" shows the percentage frequency of 1997 minus the percentage frequency of 1993 for each 0.5 BMI interval. These are just the differences in the y values between the distributions in Figure 4.1. Positive values are where the percentage frequency increased in the later distribution. Negative values are where the percentage frequency decreased in the later distribution.

In the case of 1997, the values in the chart are all negative up to a BMI value of 27. This means that the 1997 distribution curve is lower than the 1993 distribution for BMI values between 15 and 27, as can also be seen in Figure 4.1. So, the percentage of people decreased in each BMI interval between 15 and 27. Above 27 all the values in Figure 4.3 are positive and so the percentage in each of these intervals increased in 1997. In other words, from 1993 to 1997 the change was that fewer people had a BMI of less than 27 and more people had a BMI of 27 or more.

Figure 4.3 is useful in showing the magnitude of the differences and how they compare over time. So, for 1997 the lowest values are -0.43% for the BMI intervals 23.5 ≤ BMI < 24.0 and 24.0 ≤ BMI < 24.5 (plotted at 23.75 and 24.25). The changes are greater for 2001 with a lowest value of -0.62% for the BMI interval 22.5 ≤ BMI < 23.0. The changes are smaller for later years.

The differences in the cumulative distributions were also calculated and these are shown in Figure 4.4. The values in Figure 4.4 are plotted at the ends of the interval being the total difference up to the end of the interval. Figure 4.4 shows some characteristics of the patterns more clearly. There is some irregularity but the curves essentially have a single trough. This means that before the lowest trough point the percentage frequency values for the later year are lower, and after that point the later year values are higher. Hence, this is the main point at which the curves cross in Figure 4.1.

The values for the lowest points in Figure 4.4 are: 1997 -3.8% at BMI 27.0; 2001 -5.0% at BMI 27.5; 2005 -2.6% at BMI 28.5; 2009 -1.8% at BMI 29.5; 2013 -0.8% at BMI 33.0. This means, for example, that 1997 compared to 1993 has 3.8% fewer men with BMI < 27.0, and 3.8% more men with BMI ≥ 27.0. Based on the depths of the troughs, the ranking of the magnitude of changes from largest to smallest is 2001, 1997, 2005, 2009, 2013. Hence, after 2001 the changes get smaller in each time interval.

The 2013 changes have a slightly different pattern to the other years. Looking at both Figure 4.3 and Figure 4.4, there are slightly more values in 2013 compared to 2009 up to a BMI of 25 mainly due to more BMI values between 18.5 and 20.5. Hence, the percentage in the healthy category increases slightly (as also analysed in Section 4.4). The trough value in Figure 4.4 is not as deep as for the other years and so the changes are relatively small. The lowest trough point is at a BMI of 33.0, and so there is only an increase in prevalence of high BMI values above 33.0, with not much difference between 33.0 and 36.0. Hence the increase in the frequency of high BMI values happens over a smaller range of the right tail than in other years.

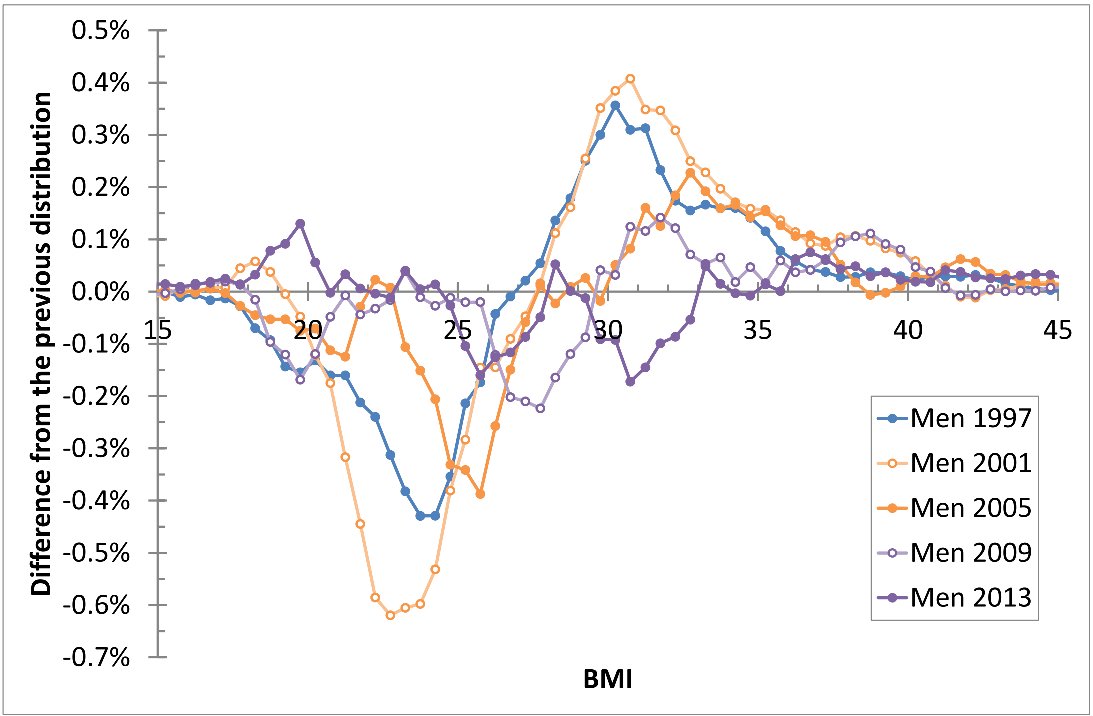


Figure 4.3 Difference from the previous for the distributions for men.

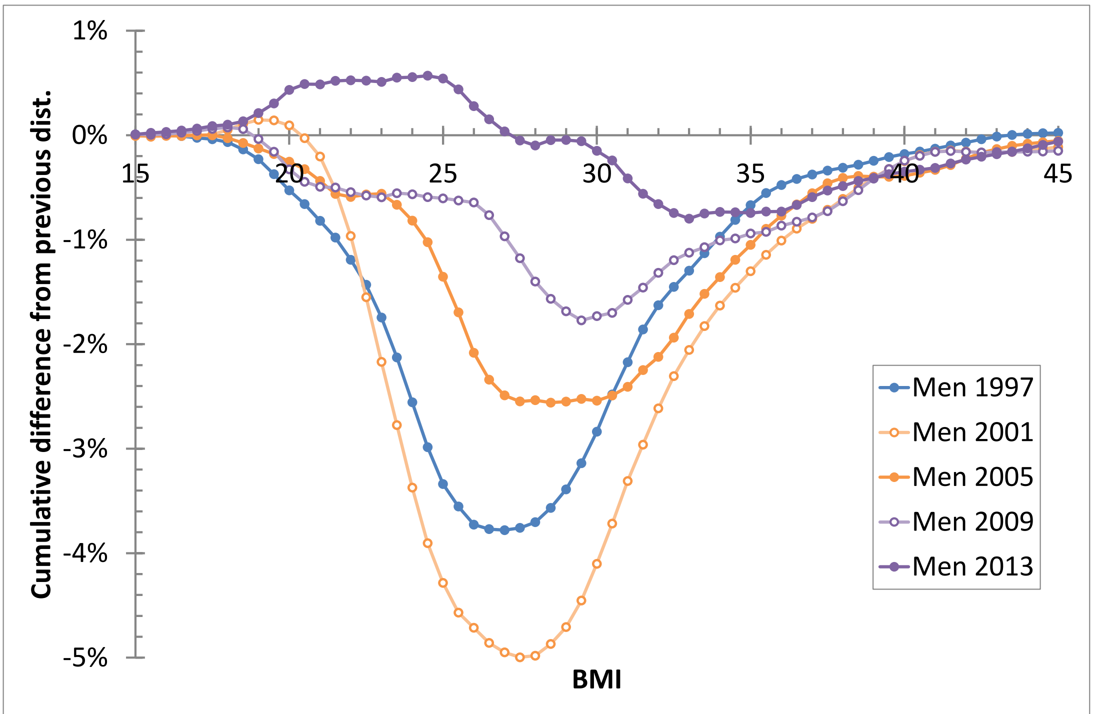


Figure 4.4 Cumulative difference from the previous for the distributions for men.

4.3 BMI frequency distributions for women

4.3.1 Shapes of the distributions for women

Figure 4.5 shows the BMI population distributions for women for the six years of 1993, 1997, 2001, 2005, 2009, and 2013 produced by the method set out in Chapters 2 and 3. The cumulative frequency distributions are shown in Figure 4.6.

Each of the distributions in Figure 4.5 is quite a smooth curve with a peak at BMI values of around 23 or 24. The peak is fairly narrow, particularly for 1993. The shape has a positive skew with a longer right tail than the left tail. Over time the peak becomes lower and tends to move a little to the right. The end of the left tail is much the same for all the distributions but the rest of the curve becomes more stretched out over time with the right tail becoming higher and longer. These changes look like a linear scaling transformation, and this is considered further and modelled in Chapter 5.

The changes are examined in the next section but visually from Figures 4.5 and 4.6 the changes from 1993 to 1997 and 1997 to 2001 are much greater and more pronounced than those since 2001.

Figures 4.5 goes up to a BMI of 45 to help see the detail of the main part of the distributions. Appendix 4.I has a chart with BMI up to 50 to see more of the right tail as well as the curves for 1993 to 2005 and 2005 to 2013 in separate charts.

A table of values for all the distributions in Figure 4.5 (along with the distribution for 2017 in Section 14.4.1 in Chapter 14) is given in Appendix 14.III in Table 14.III.4.

4.3.2 Differences between each year for women

The differences between successive distributions in Figure 4.5 are shown in Figure 4.7 with the cumulative differences in Figure 4.8. The general interpretation of these charts was covered earlier in this chapter in Section 4.2.2.

The patterns of the changes in the distributions as time passes are for fewer low BMI values and more high BMI values. So, the differences in Figure 4.7 are generally negative for lower BMI and positive for higher BMI.

The changes are large between 1993 and 1997, and between 1997 and 2001 with these both showing as a clear deep trough in the cumulative differences in Figure 4.8. The changes are much smaller for the years after that and the series for 2005, 2009, and 2013 show a roughly flat region at the lowest points in Figure 4.8 indicating that the middle of the right tails of the distributions are all quite similar in those ranges. For example, the Figure 4.8 curve for 2005 is quite flat between a BMI of 26.5 and 31, and this naturally corresponds to the values in Figure 4.7 in this range being close to 0 and the 2001 and 2005 curves in Figure 4.5 being very close together. One difference with the 2013 distribution compared to 2009 is that it has a slightly higher frequency of BMI values at just below 20. This was also the case for the distributions for these two years for men (Section 4.2.2).

The values for the minimum points for each year in Figure 4.8 are: 1997 -4.1% at BMI 25.5; 2001 -3.7% at BMI 27.5; 2005 -1.6% at BMI 30.5; 2009 -1.1% at BMI 29.0; 2013 -0.9% at BMI 32.5. As discussed in Section 4.2.2 these represent points at which the curves cross. The stated percentage gives the difference in prevalence for the interval before and after the given BMI value. For example, 1997 has 4.1% fewer BMI values less than 25.5 and 4.1% more BMI values over 25.5 compared to 1993.

The magnitude of the changes can be compared using the Figure 4.8 minimum point values and also visual inspection of the charts. Based on the minimum point values in the previous paragraph, the ranking of the magnitude of changes from largest to smallest is 1997, 2001, 2005, 2009, 2013. So, the changes get smaller each year with 1997 and 2001 being much greater than the later years.

There is a slight difference in where the changes happen for 1997 and 2001 in Figures 4.7 and 4.8. The shapes of the patterns are similar but with the changes occurring at lower BMI values in 1997. For example, the lowest trough point is at a smaller BMI value in 1997 (25.5) than in 2001 (27.5). To some extent, this is just a consequence of the trend of BMI values being lower in earlier years and hence the starting point for the changes is lower in earlier years.

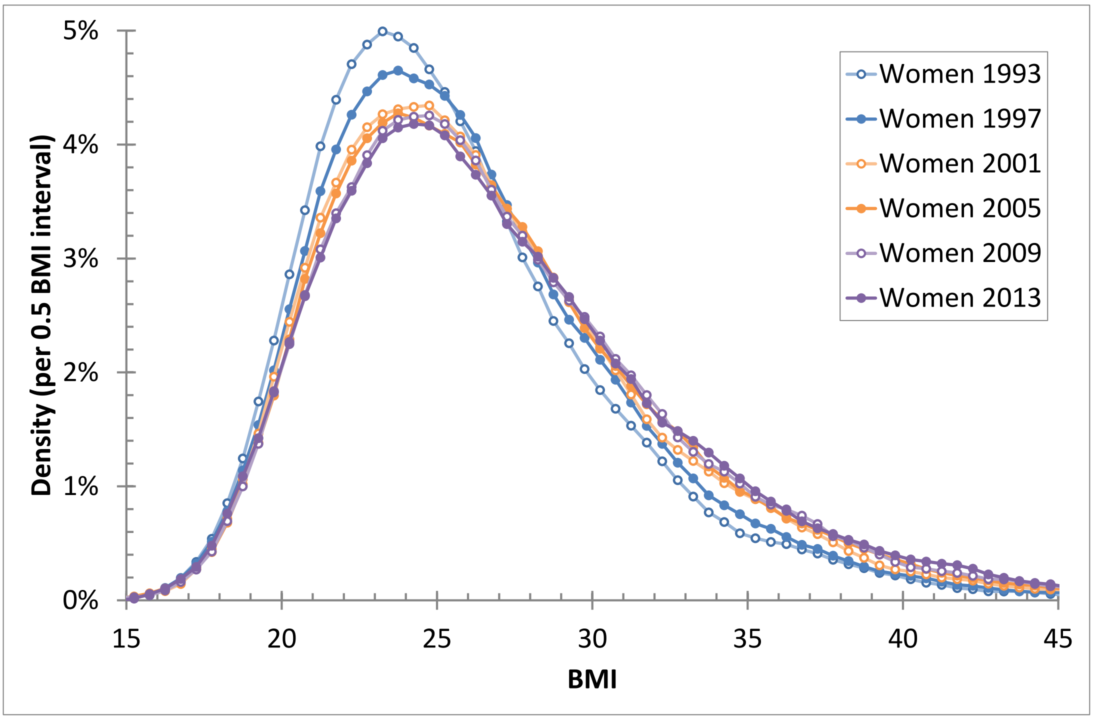


Figure 4.5 BMI population distributions for women for 1993 to 2013.

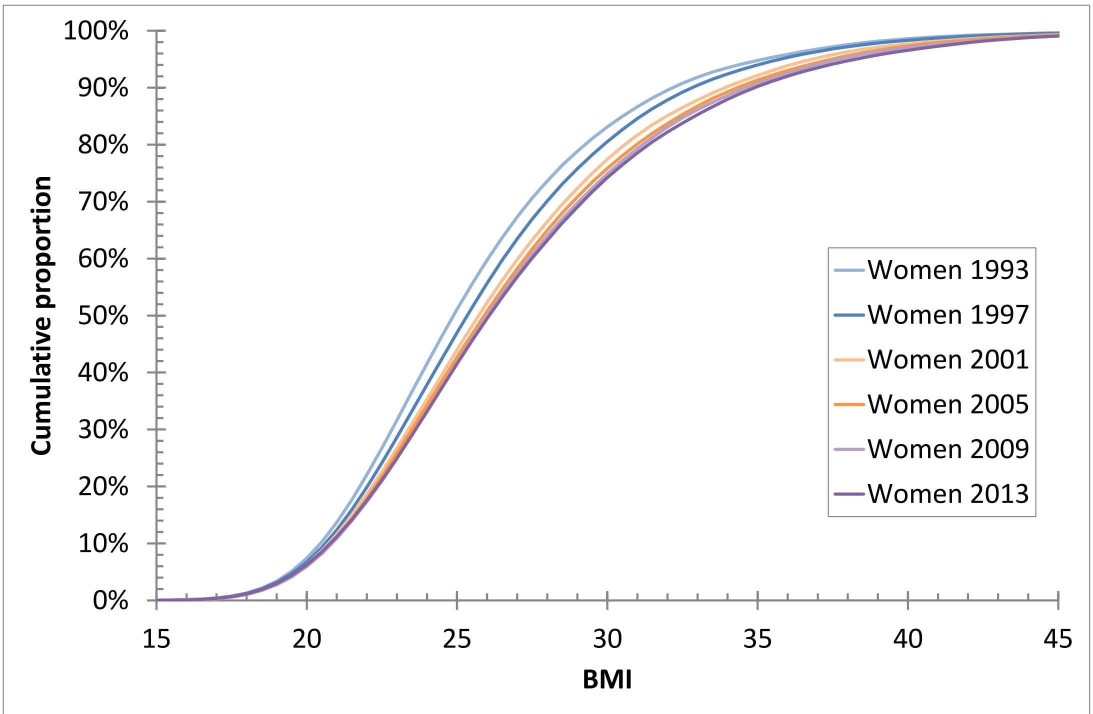


Figure 4.6 BMI cumulative frequency distributions for women for 1993 to 2013.

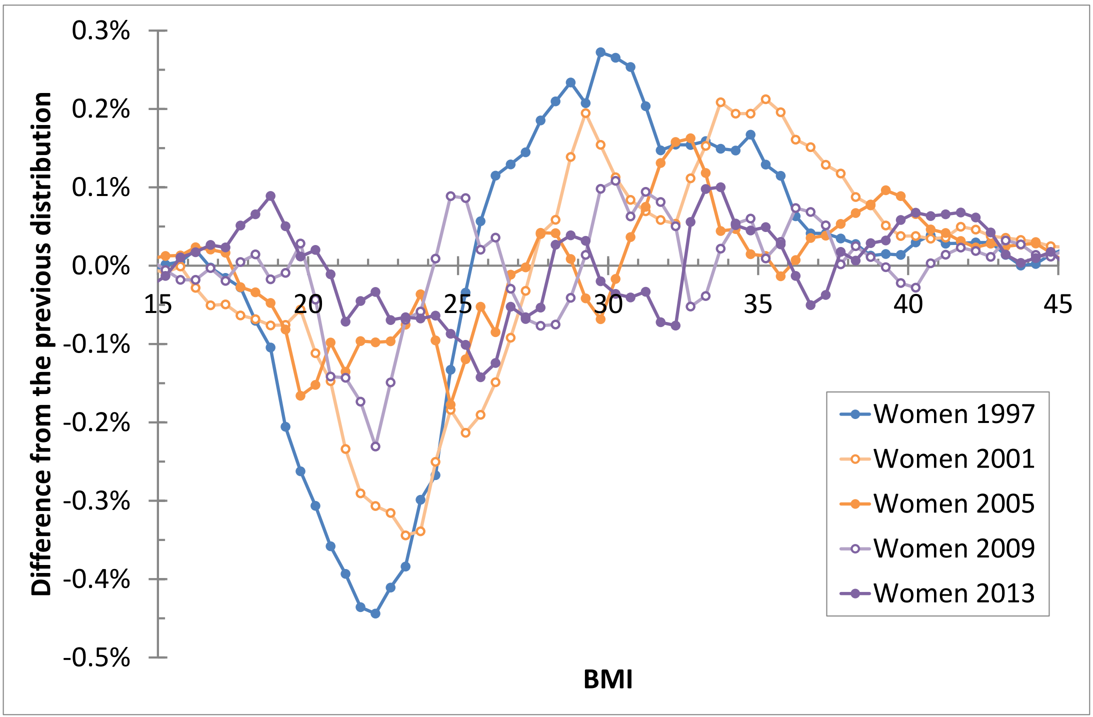


Figure 4.7 Difference from the previous for the distributions for women.

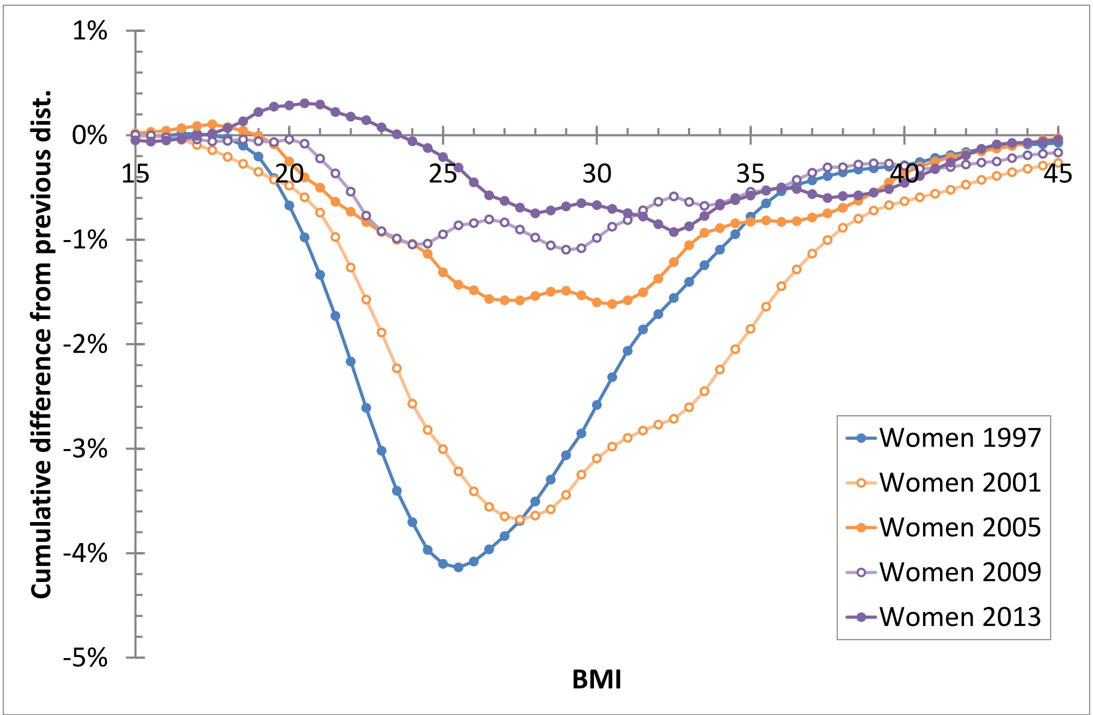


Figure 4.8 Cumulative difference from the previous for the distributions for women.

4.4 Descriptive statistics for men

The descriptive statistics for the population distributions for men are in Table 4.1. These were calculated in the usual way for this type of data, as described in Section 4.1. Some of the statistics are plotted as time series in Figures 4.9-4.13.

Table 4.1 Descriptive statistics for the population distributions for men.

	1993	1997	2001	2005	2009	2013
Mean	26.05	26.41	26.90	27.19	27.39	27.42
Median	25.76	26.10	26.55	26.79	26.88	26.87
Modal interval mid-point	25.25	25.75	25.75	26.25	25.75	26.25
Standard deviation	3.87	4.00	4.29	4.48	4.67	4.84
Skewness	0.75	0.63	0.71	0.72	0.87	0.89
Underweight (BMI < 18.5)	1.1%	0.9%	1.0%	1.0%	1.0%	1.2%
Healthy (18.5 ≤ BMI < 25.0)	40.4%	37.2%	32.8%	31.6%	30.9%	31.3%
Overweight (25.0 ≤ BMI < 30.0)	44.7%	45.2%	45.4%	44.2%	43.1%	42.4%
Obese I (30.0 ≤ BMI < 35.0)	11.6%	13.8%	16.6%	18.1%	18.8%	18.2%
Obese II (35.0 ≤ BMI < 40.0)	1.9%	2.4%	3.4%	4.1%	4.8%	5.2%
Severely obese (BMI ≥ 40.0)	0.3%	0.5%	0.7%	1.1%	1.3%	1.7%

Figure 4.9 shows the mean and median values, and Figure 4.10 highlights the changes in the mean and median by plotting the differences compared to the previous distribution. The biggest increase in both the mean and the median is from 1997 to 2001. There is also a large increase in both values from 1993 to 1997. The increases get progressively smaller after 2001. There is very little change from 2009 to 2013 with the median actually being 0.01 lower in 2013. From 2005 onwards the left side of the BMI distributions are similar (Figures 4.1 and 4.2), and so the median, in measuring the 50th percentile value, only has small changes from 2005 to 2009 and from 2009 to 2013. Ranking the increases in the mean and median from largest to smallest is consistent with the discussion in Section 4.2.2, with the order: 2001, 1997, 2005, 2009, 2013. The increases in the mean and median over the eight years from 1993 to 2001 are 0.85 and 0.79, compared to just 0.51 and 0.32 in the twelve years from 2001 to 2013. Strictly, the 1993 data is more towards the end of the four years and so the time period to 2001 can be considered as slightly shorter than 8 years, being more like 7.5 years (Section 4.1).

From Table 4.1, the skewness does not change much between the different years. However, the standard deviation increases over time as the distribution gets more spread out, as can be seen in the distributions in Figure 4.1. The mid-point of the modal interval is also shown in Table 4.1 but this is quite sensitive to the exact frequencies of the intervals at the peak of the distribution, and the peaks are quite flat for some of the distributions (Figure 4.1).

Table 4.1 shows the BMI category values and these are plotted in Figure 4.11. The main change over time is fewer people in the healthy category and more in the obese categories. From 1993 to 2013 the underweight and overweight category percentages do not change much. In the latter case this is presumably because the numbers moving from healthy to overweight are roughly balanced by those moving from overweight to obese I.

Figure 4.12 shows the changes in each BMI category from one distribution to next. Figure 4.12 is actually basically the same as Figure 4.7 but gives a smoother although less detailed picture as it combines the values across the BMI categories rather than using 0.5 BMI intervals. So, for example, the value for the healthy category is the totals of the values in Figure 4.7 between 18.5 and 25. Hence the patterns in the two charts are much the same, and it again highlights the largest changes being in 2001 and 1997. The nature of the changes for 2013 is a bit different to the other years with an increase in the healthy percentage and a reduction in obese I, although the changes are small.

Figure 4.13 plots the values for just the healthy category percentage to show the values more clearly than in Figure 4.11. This is an important category as it represents what is generally considered to be the ideal BMI values. This category percentage falls considerably from 40.4% in 1993 to just 31.3% in 2013. The biggest fall is from 1997 to 2001, with a large reduction also from 1993 to 1997. The changes are small in recent years and, as mentioned above, with even a small increase in 2013. The total percentage of the three obese categories is shown in Figure 4.14 and has a similar pattern in the way it increases with the largest increases from 1993 to 1997 and 1997 to 2001. Charts for each BMI category are in Appendix 4.II, giving the values for men and for women on the same chart.

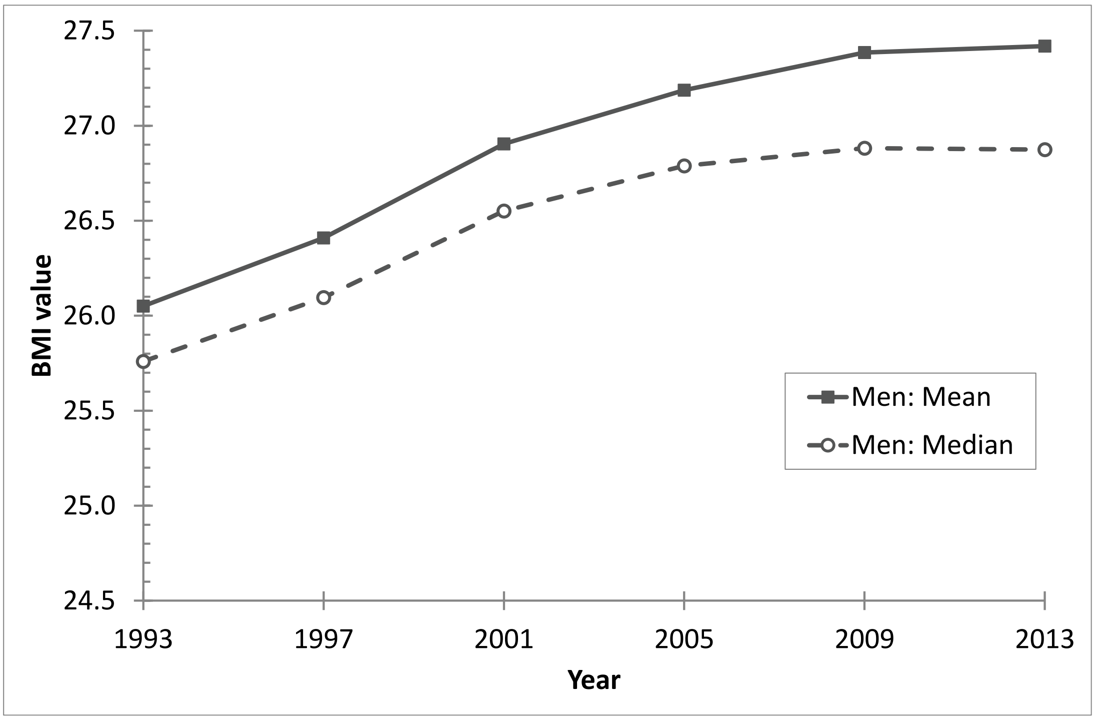


Figure 4.9 Mean and median values for the distributions for men.

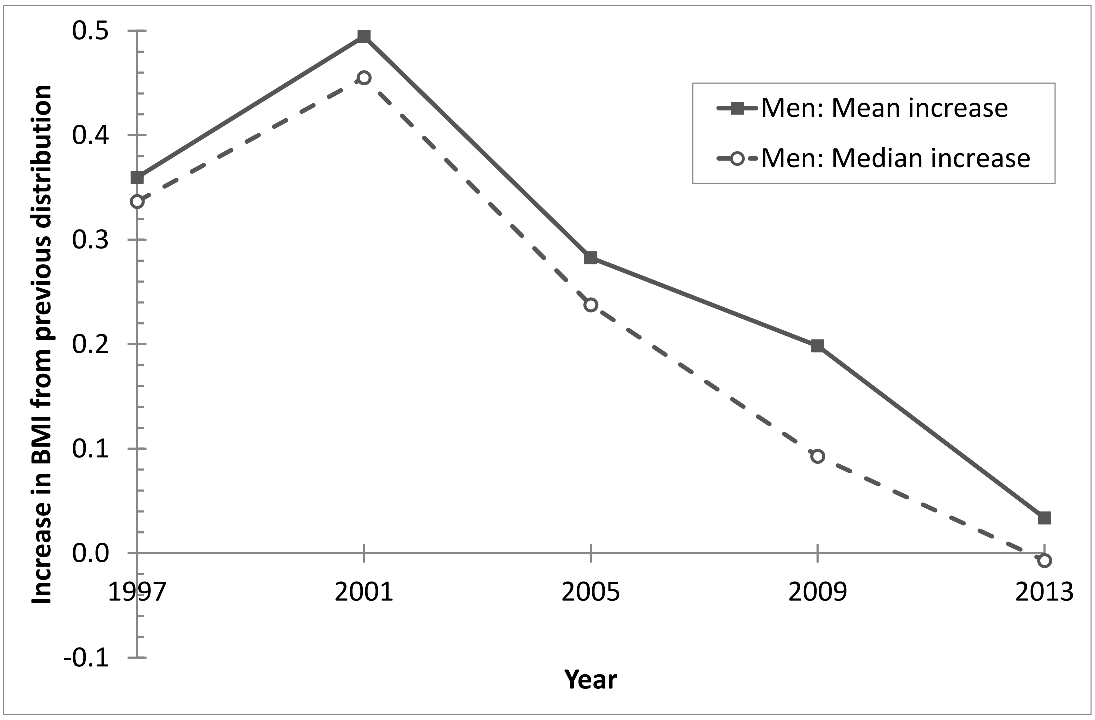


Figure 4.10 Change in mean and median values from the previous distribution for men.

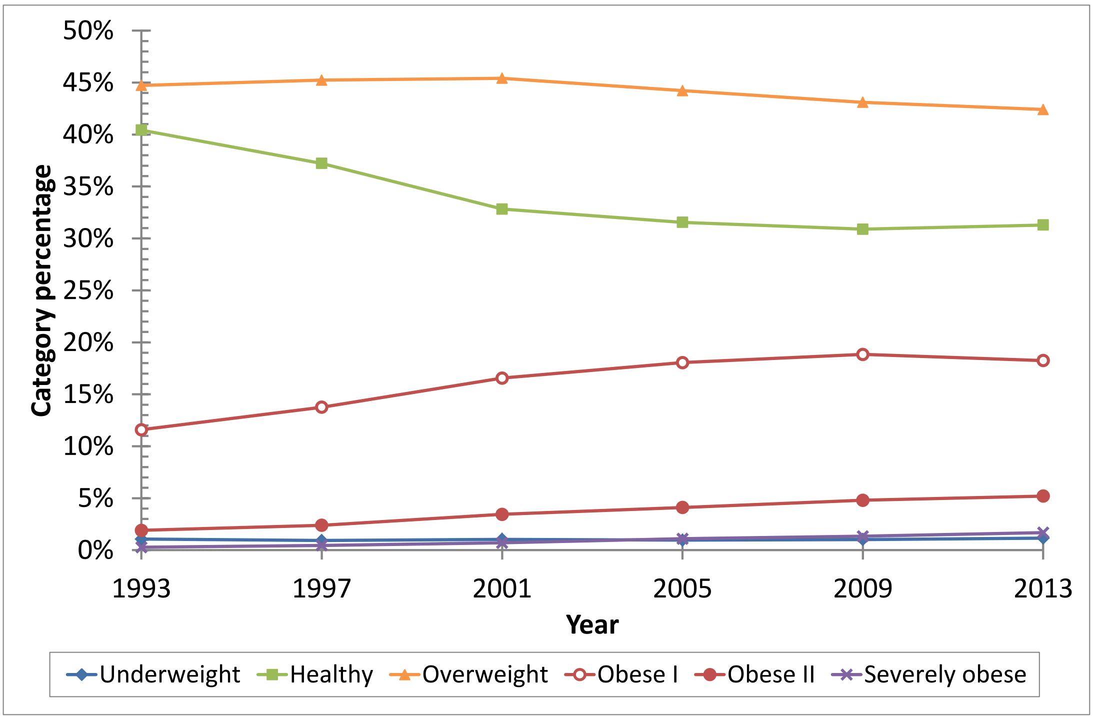


Figure 4.11 BMI category percentages for the distributions for men.

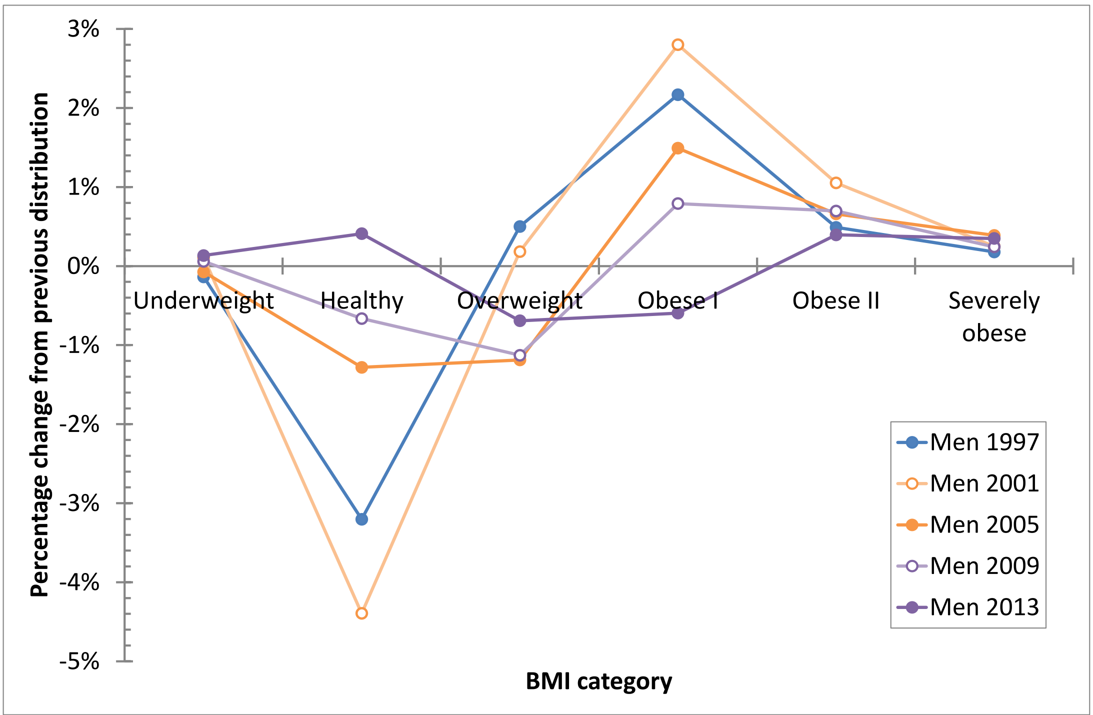


Figure 4.12 Change in BMI category percentages from the previous distribution for men.

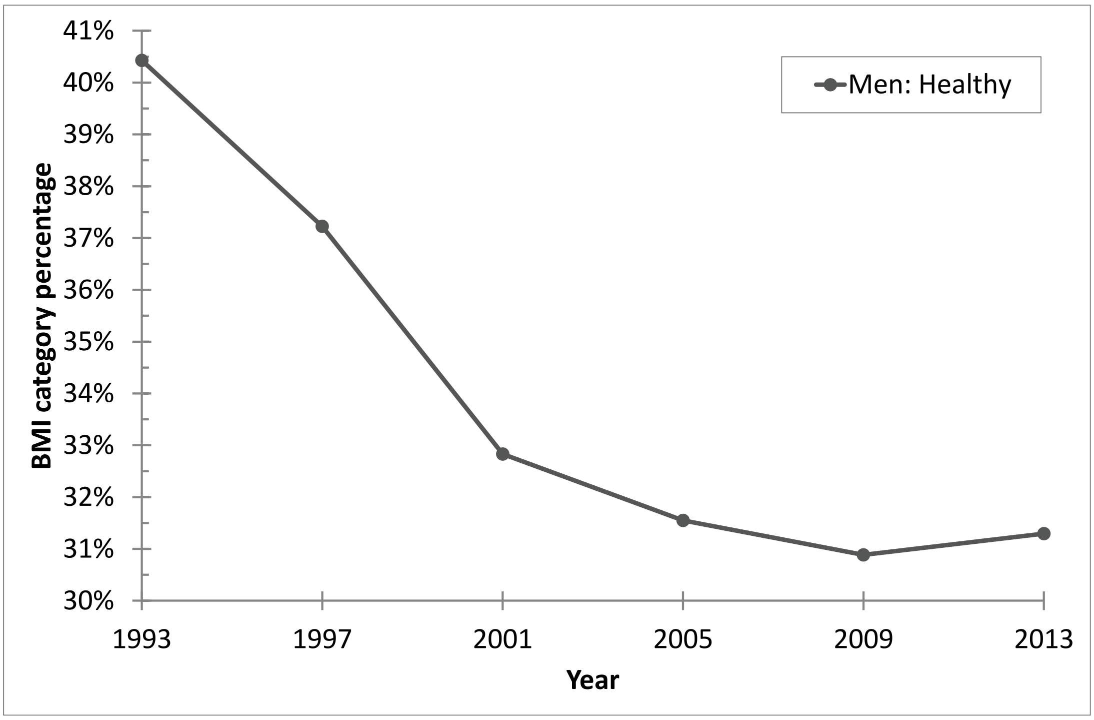


Figure 4.13 Healthy category percentages for the distributions for men.

Figure 4.14 Total of obese category percentages for the distributions for men.

Statistics were also calculated for the data before smoothing. For reference the values are in Appendix 4.III. The values are very similar to those in Table 4.1. In particular, the smoothing method (Section 3.5) does not change the mean, standard deviation or the skewness value and so these are exactly the same.

4.5 Descriptive statistics for women

The descriptive statistics for the population distributions for women are in Table 4.2. The mean and median are plotted in Figure 4.15, with the changes from the previous distribution in Figure 4.16. The category percentages are in Figure 4.17 with the changes from the previous distribution in Figure 4.18. The changes in the category percentages for healthy and for all obese are in Figures 4.19 and 4.20, respectively. Charts of the percentages for each category are in Appendix 4.II.

Table 4.2 Descriptive statistics for the population distributions for women

	1993	1997	2001	2005	2009	2013
Mean	25.76	26.18	26.65	26.88	27.03	27.13
Median	24.88	25.34	25.72	25.90	26.01	26.07
Modal interval mid-point	23.25	23.75	24.75	23.75	24.75	24.25
Standard deviation	4.95	5.04	5.38	5.54	5.63	5.77
Skewness	1.32	1.05	1.06	1.05	1.05	1.06
Underweight (BMI < 18.5)	2.1%	2.0%	1.7%	1.8%	1.7%	1.9%
Healthy (18.5 ≤ BMI < 25.0)	49.0%	45.0%	42.2%	40.9%	40.0%	39.6%
Overweight (25.0 ≤ BMI < 30.0)	32.0%	33.6%	33.5%	33.2%	33.1%	32.7%
Obese I (30.0 ≤ BMI < 35.0)	11.7%	13.5%	14.7%	15.5%	15.9%	16.0%
Obese II (35.0 ≤ BMI < 40.0)	3.8%	4.3%	5.5%	6.0%	6.2%	6.4%
Severely obese (BMI ≥ 40.0)	1.4%	1.7%	2.3%	2.7%	3.0%	3.4%

From Table 4.2 and Figures 4.15 and 4.16, the mean and median both increase in every time interval. The two largest increases in the mean and median are in 1997 (from 1993 to 1997) and 2001 (from 1997 to 2001). The mean has its largest increase in 2001 and its 2nd largest increase in 1997, whereas the median is the other way round with the largest increase in 1997 and the 2nd largest in 2001. The increases in both mean and median are much smaller in 2005 and reduce further in 2009 and 2013. The increases in the mean and median over the eight years from 1993 to 2001 are 0.89 and 0.84, compared to just 0.47 and 0.35 in the twelve years from 2001 to 2013. Strictly, the 1993 data is more towards the end of the four years and so the time period to 2001 can be considered as slightly shorter than 8 years, being more like 7.5 years (Section 4.1).

The skewness value is high in 1993 at 1.32 compared to values of 1.05 or 1.06 in the other years. The standard deviation increases each year as the peak gets lower and the distribution extends in the right tail. This can also be seen visually with the distributions in Figure 4.5 getting more spread out over time.

For the BMI categories (Table 4.2 and Figures 4.17-4.20), the main change over time is fewer people in the healthy category and more in the obese categories, with not much change in the overweight category. The healthy category percentage falls from 49.0% in 1993 to 39.6% in 2013, although the reductions from one distribution to the next get progressively smaller in magnitude over time. The pattern of increases in the total of all obese (Figure 4.20) is similar although the increase from 1997 to 2001 is slightly larger than from 1993 to 1997. The overall change is from 16.9% to 25.8%.

As discussed in Section 4.4, the chart of the changes in the category percentages (Figure 4.18) is basically the same as the chart of the changes in the distributions (Figure 4.7). Figure 4.18 combines all the intervals for each BMI category and so shows a less detailed but perhaps slightly clearer picture. For example, comparing the 1997 and 2001 changes, Figure 4.18 shows that 1997 has greater reductions in the healthy category and greater increases in overweight and obese I, but not as big an increase in obese II or severely obese.

The statistics for the data before smoothing are in Appendix 4.III, and are similar values to Table 4.2. As noted in Section 4.5, the smoothing method does not change the mean, standard deviation or the skewness value and so these are exactly the same.

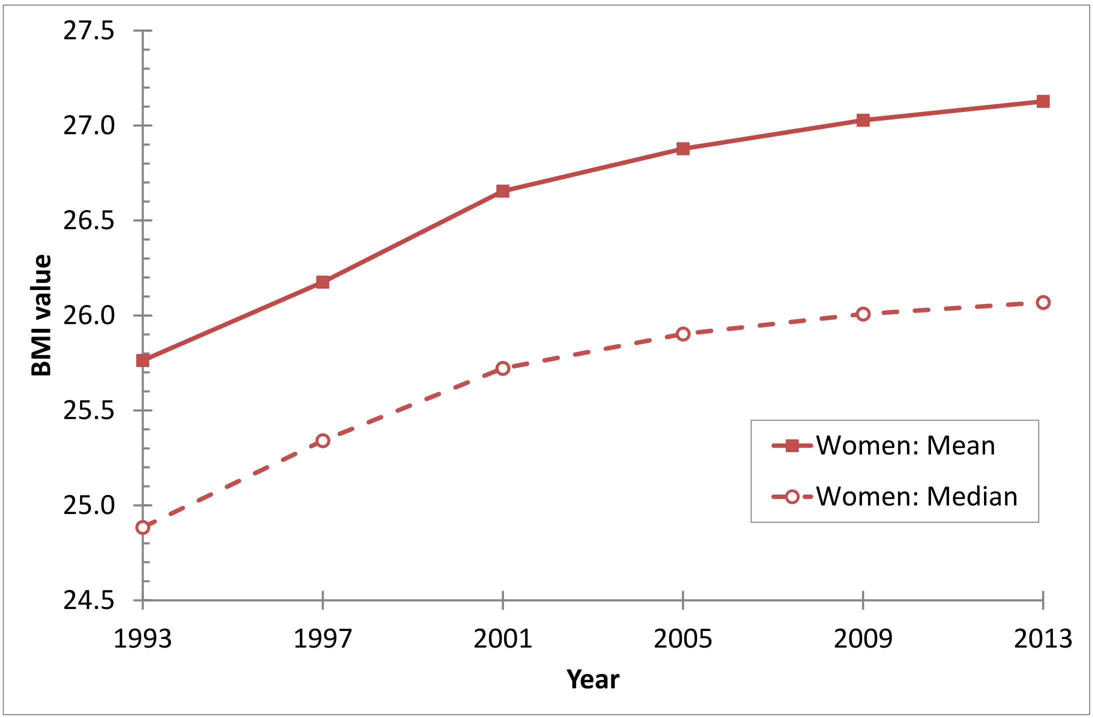


Figure 4.15 Mean and median values for the distributions for women.

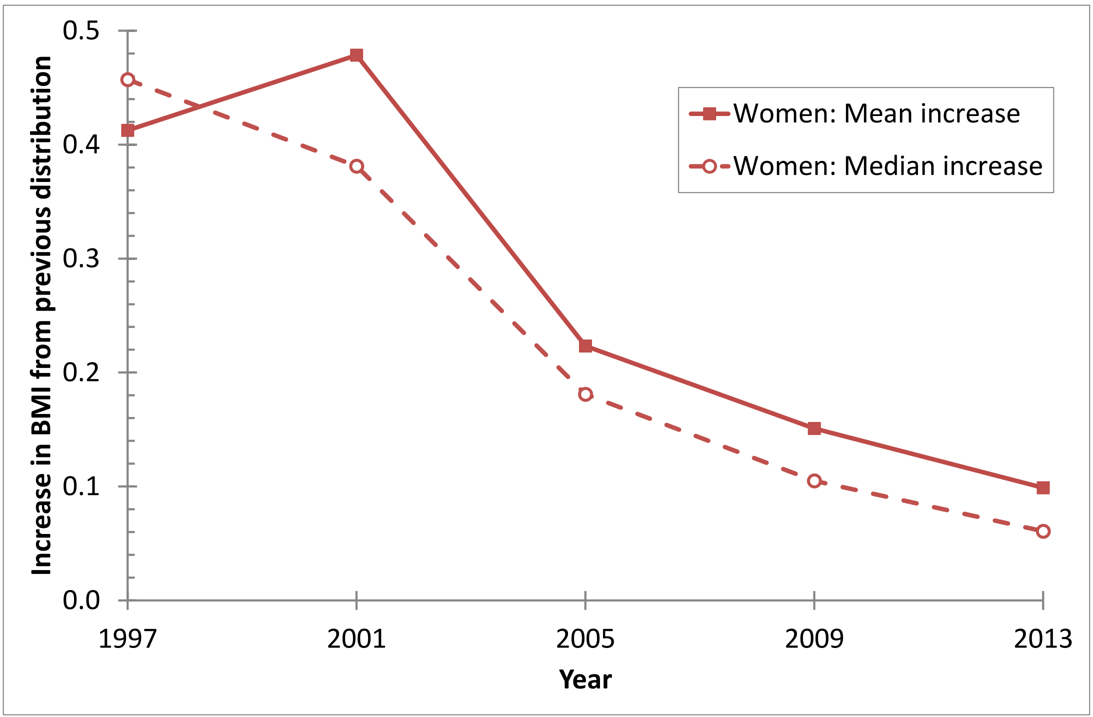


Figure 4.16 Change in mean and median values from the previous distribution for women.

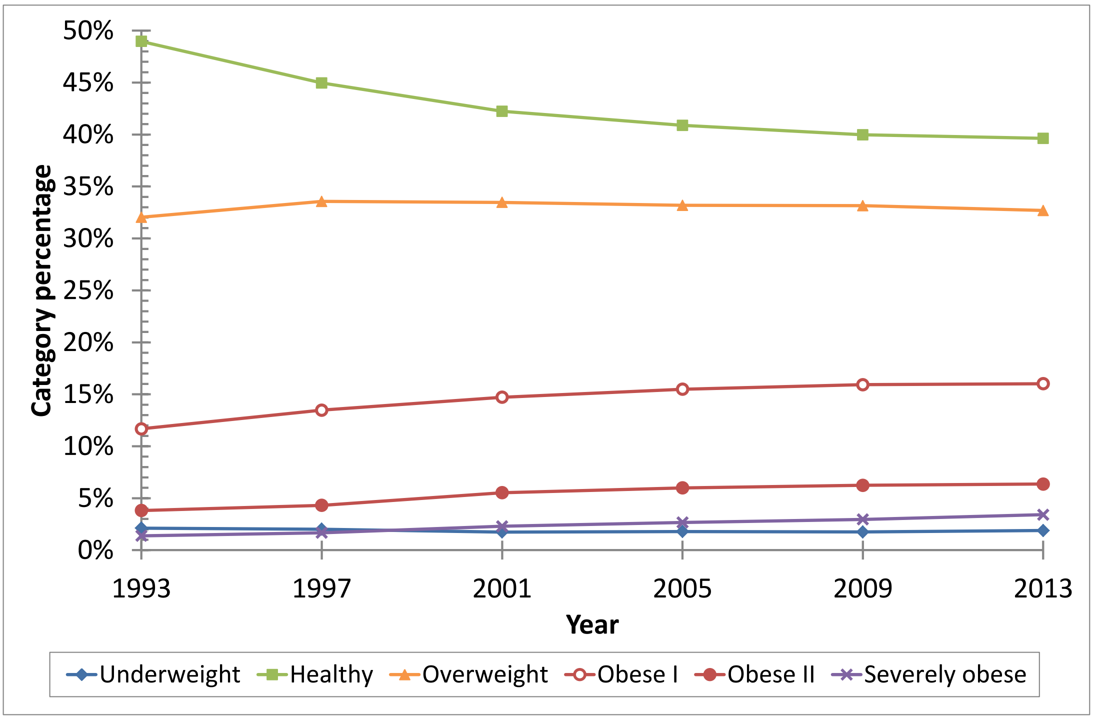


Figure 4.17 BMI category percentages for the distributions for women.

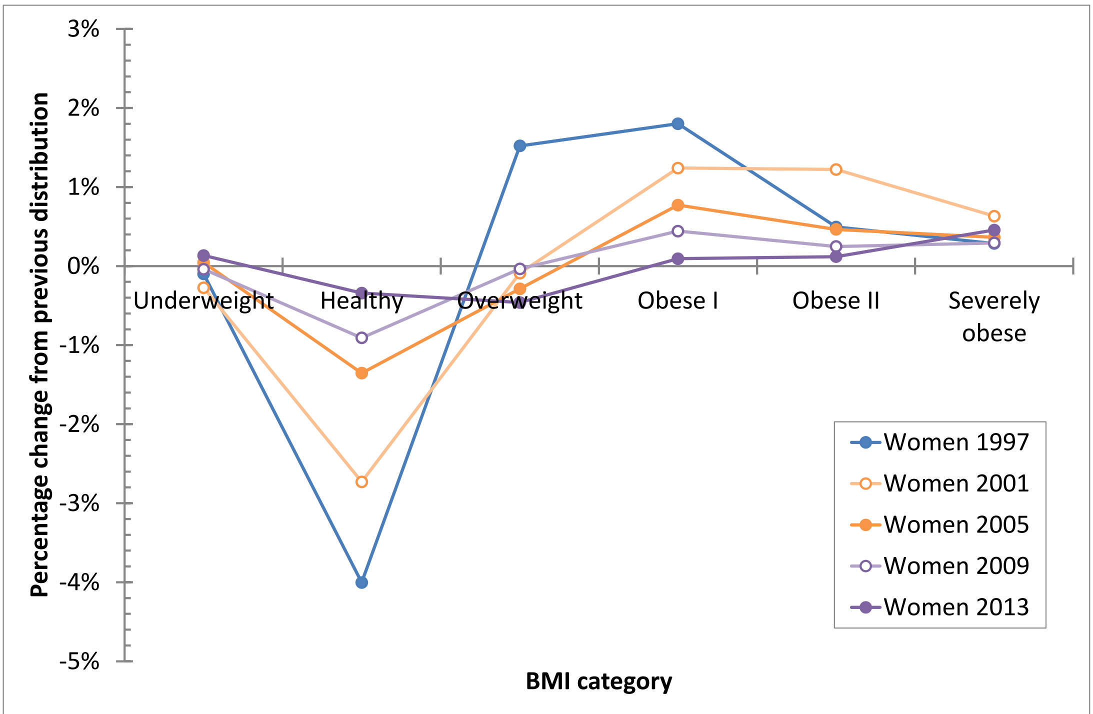


Figure 4.18 Change in BMI category percentage from the previous distribution for women.

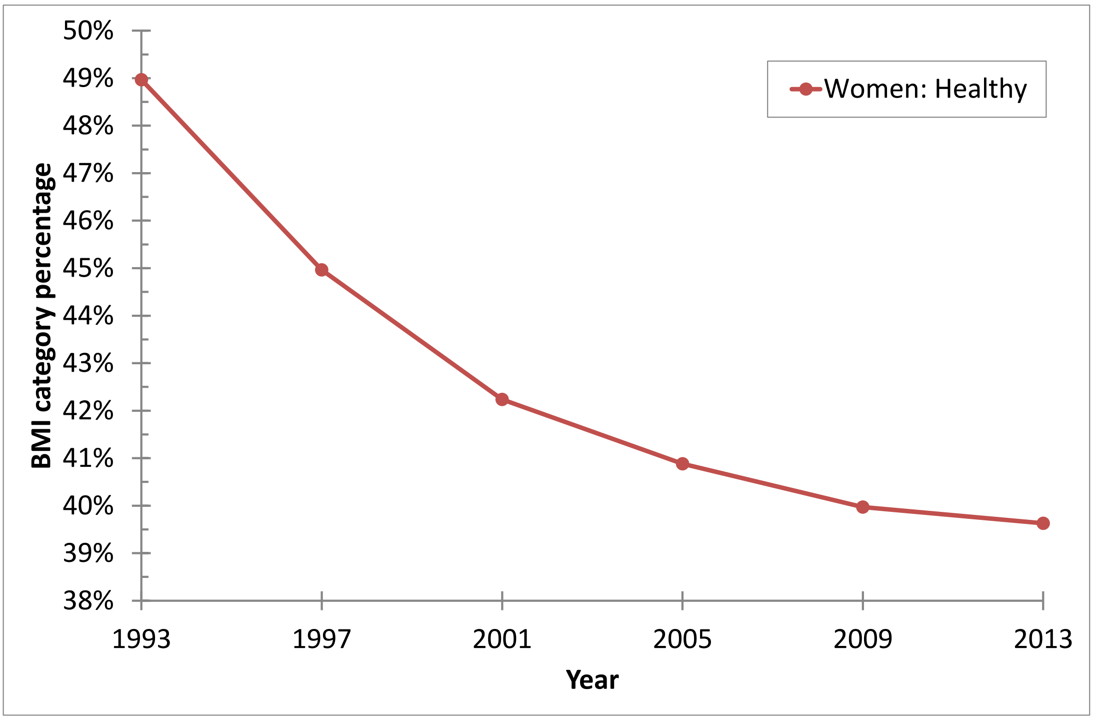


Figure 4.19 Healthy category percentages for the distributions for women.

Figure 4.20 Total of obese category percentages for the distributions for women.

4.6 Percentiles

A further analysis calculated and compared the percentile values for the distributions for the different years. The percentiles can be very helpful in seeing the patterns of how the distributions change over time. They are used extensively in the work presented in this report.

4.6.1 Calculation of the percentiles

The percentile values give the BMI value that is the given percentage of the way through the distribution. The values here, and used generally in the analysis in the report, are the percentiles for 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, and 99%. As for the median (Section 4.1), the percentile values were calculated from the smoothed BMI population distribution using linear interpolation within each 0.5 BMI interval. The 50th percentile value is the median.

As an example of the interpolation calculation, the 30th percentile value for the distribution for women for 2013 was obtained as follows: The value falls in the BMI interval from 23.5 to 24.0 because the smoothed distribution has 29.01% of values with BMI < 23.5 and 33.16% of values with BMI < 24.0. The interval of 23.5 ≤ BMI < 24.0 has a frequency percentage of 4.15%. The 30th percentile value was therefore calculated as: 23.5 + 0.5 × ((30% - 29.01%) / 4.15%) = 23.62. So, the BMI value for the 30th percentile is 23.62. This estimates the BMI value that is exactly 30% of the way through the distribution.

As another example, the 80th percentile for the distribution for women for 2013 is in the BMI interval from 31.0 to 31.5. This is because there are 78.56% of values with BMI < 31.0 and 80.50% with BMI < 31.5, and 1.94% in the interval 31.0 ≤ BMI < 31.5. So, the 80th percentile value is given by the calculation: 31.0 + 0.5 × ((80% - 78.56%%) / 1.94%) = 31.37. Tables of the actual percentile values are in Appendix 4.IV.

4.6.2 Percentile results

Figures 4.21 and 4.22 plot the total increase from 1993 of the BMI value of each percentile for men and for women. This shows the pattern that the higher the percentile the greater the increase. For example, the 1% values at the left end of the distribution hardly change at all whereas the 99% values at the right end have large BMI increases that get greater over time reaching values at 2013 of 4.9 for men and 3.6 for women. The pattern corresponds to the observations already made about the distributions that the start of the left tail stays about the same but the rest of the distribution stretches out more and more to the right over time. These charts also show how the increases get less over time with the curves flattening out, but that this happens more at the lower percentiles. The higher percentiles keep increasing which corresponds to the right tails still extending in recent years.

A particularly useful chart is to plot for each percentile the increase in BMI from one distribution to the next against the BMI value in the first distribution. These charts are Figures 4.23 and 4.24. To give an example, the 50th percentile values (i.e., the median, as given in Table 4.1) for men are 25.76 in 1993 and 26.10 in 1997, an increase of 0.34. Therefore the 1997 curve in Figure 4.23 includes the value of 0.34 plotted against 25.76, as the 7th point in the series: i.e., (25.76, 0.34). As another example, the 80th percentile value for women for 2013 is 31.37, as explained above in Section 4.6.1. The 80th percentile value for women for 2009 is 31.18 and so the increase from 2009 to 2013 is 0.19. So, the chart in Figure 4.24 has the value of 0.19 plotted against 31.18 as the 10th point in the 2013 series: i.e., (31.18, 0.19).

All the percentiles are plotted in the same way giving 13 points for each series. The points give the increase in BMI between the two populations for each starting BMI value, based on the equivalent percentile ranking in the population. It should be noted that the end points, particularly the 99% value, will be sensitive to the specific extreme values in the HSE data samples and so should be treated as much more uncertain than the other points.

The series in the charts all have a similar general pattern of the BMI increases tending to be larger for higher percentiles. The relationship between the BMI increase and the starting BMI value is approximately linear with a positive gradient across most values for each series, with the increases starting at about zero for the lowest percentiles. For a few of the series, particularly 2009 and 2013 for men, several of the lower percentiles are close to zero and then there is an upward approximately linear pattern for the BMI increases for the higher percentiles. For example, for 2013 for men the values up to the 70th percentile are all close to zero mostly being small negative values and then the values are a fairly steep upward gradient for the last four points for the 80th, 90th, 95th, and 99th percentiles.

With these charts, a linearly increasing pattern starting from 0 at a certain BMI value corresponds to the distribution changing with a linear scaling transformation above that BMI value. This pattern applies approximately for all the curves for both men and women indicating that a linear scaling transformation should be a good model of the changes in the distributions. Such a model is applied in Chapter 5.

Consistent with previous discussions in this chapter, the series in Figures 4.23 and 4.24 show smaller changes in recent years than in the early years of the HSE and these mainly just apply at higher BMI values at the right of the distributions. Based on whether the series in Figures 4.23 and 4.24 are generally higher or lower than each other, the biggest increase in the BMI distribution for men is from 1997 to 2001, followed by 1993 to 1997. The changes then get smaller over time after that. For women the first two time periods have the largest increases and are quite similar to each other. After that, 2005 has smaller increases and the last two time intervals of 2009 and 2013 are similar to each other and have the smallest changes.

The overall changes in the percentiles from 1993 to 2013 are plotted and discussed further in Section 4.8.3.

4.6.3 Percentile implications

The interpretation and implications of the results are discussed more fully in Chapter 6. Briefly, the interpretation of the changes in the distributions over time is that it is due to changes in the various obesity factors so that people have different lifestyles and experiences because of living in different times.

The nature of the changes in the BMI distributions over time is that the start of the left tail stays about the same and the rest of the distribution gets more and more stretched out to the right. In particular, the pattern from the percentiles is that this corresponds to higher BMI values increasing over time more than lower BMI values with approximately a linear relationship between the BMI value and the BMI increase. This also applies to the overall change from 1993 to 2013 in Section 4.8.3.

The implications of this are that the changes in the obesity factors over time have had a much greater effect on those with a higher BMI than on those with a lower BMI.

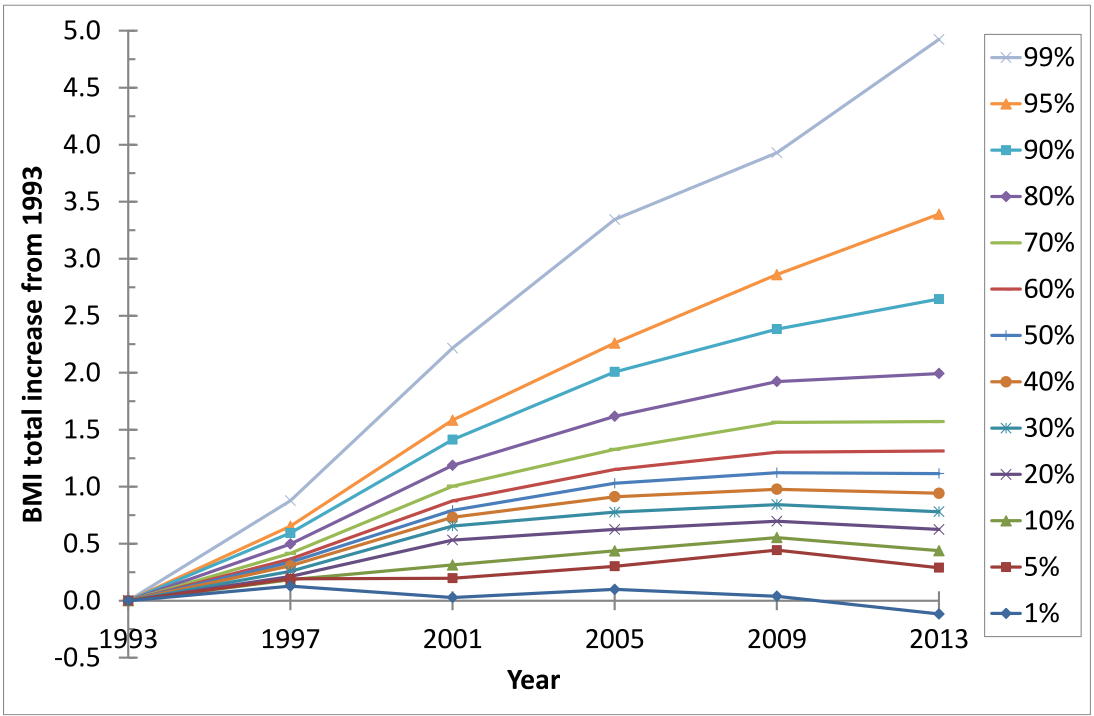


Figure 4.21 Total increases in BMI since 1993 for the percentile values for men.

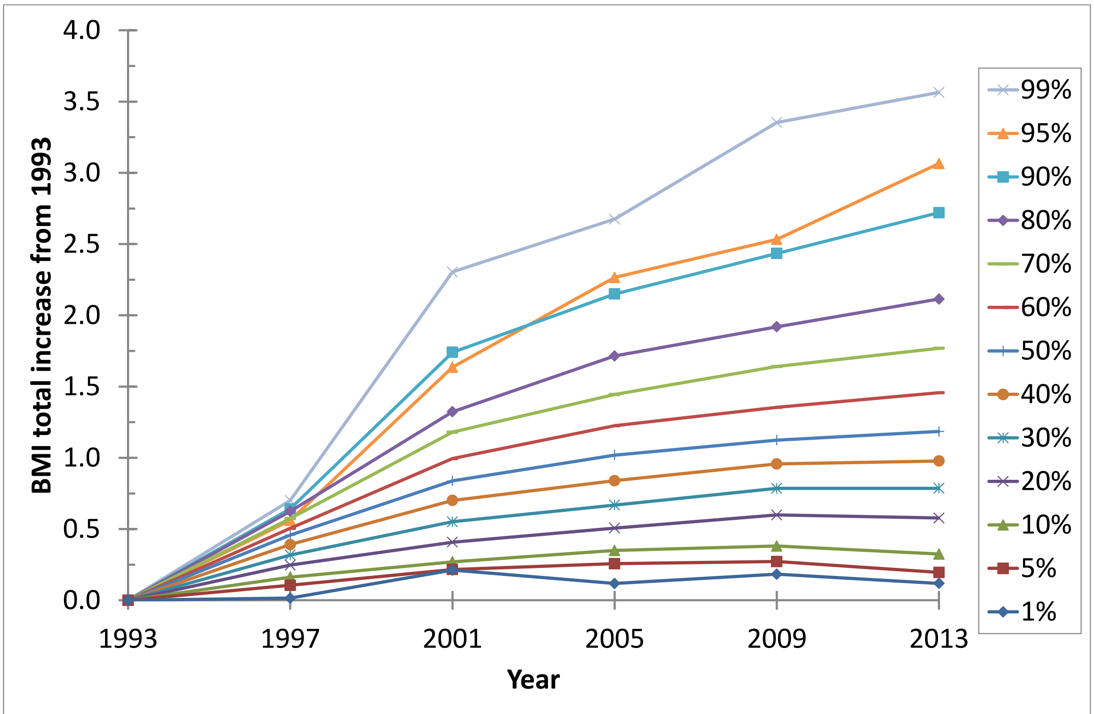


Figure 4.22 Total increases in BMI since 1993 for the percentile values for women.

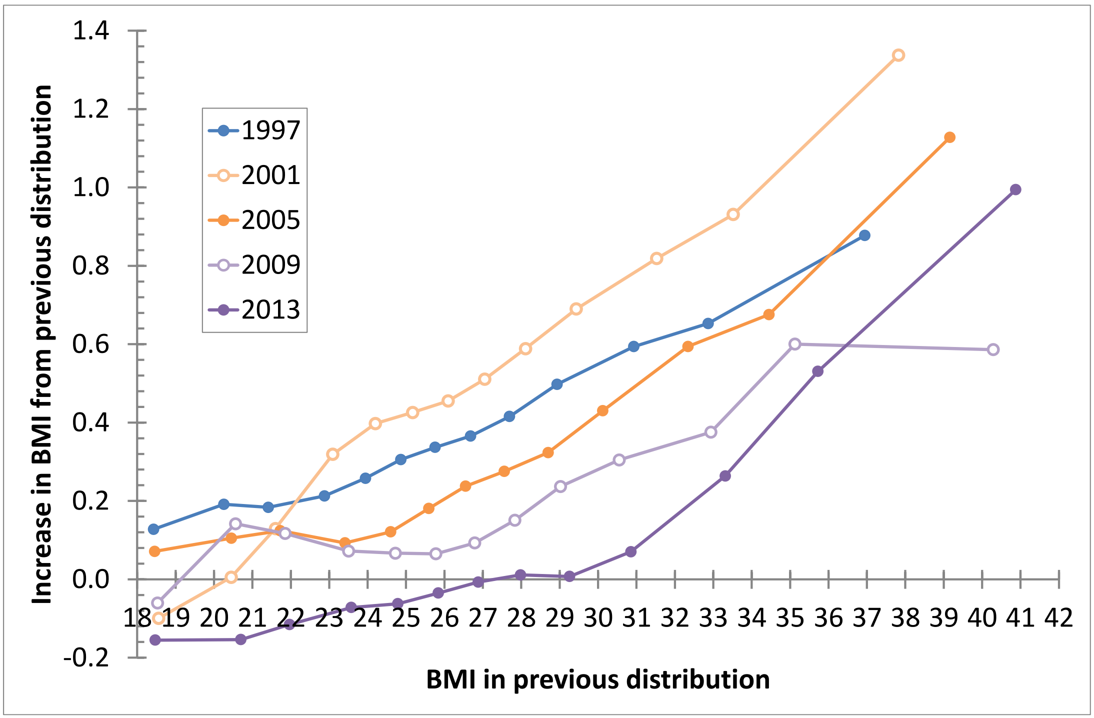


Figure 4.23 Increase in BMI against the previous BMI value for the percentiles for men.

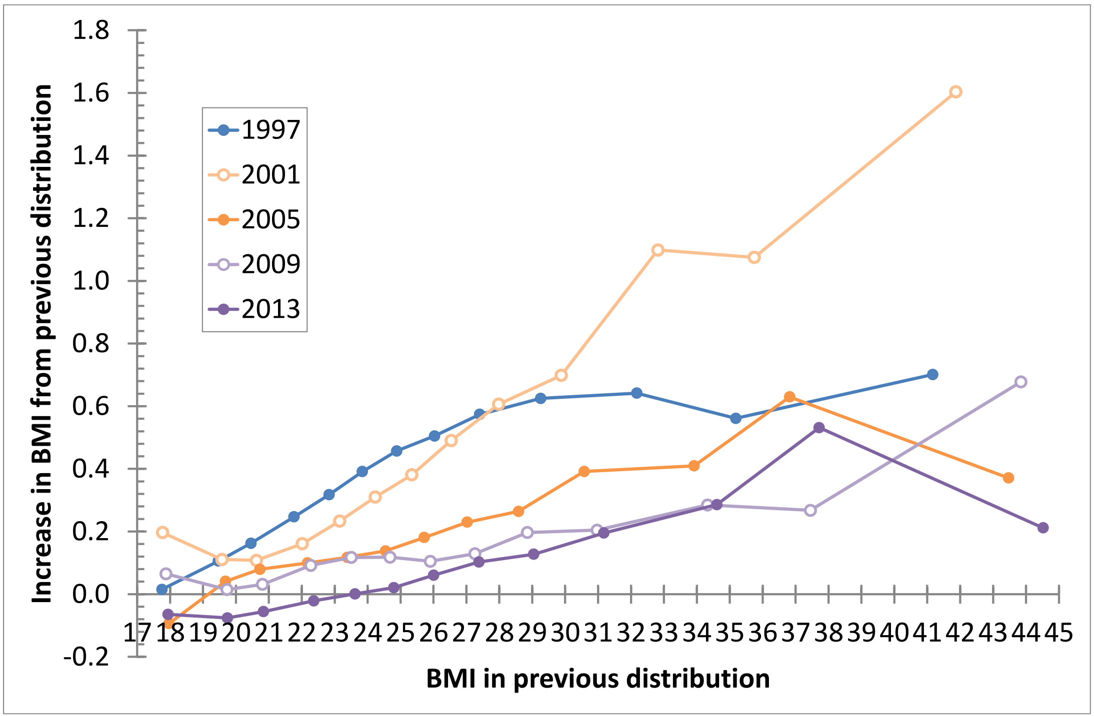


Figure 4.24 Increase in BMI against the previous BMI value for the percentiles for women.

4.7 High BMI values

The analysis in earlier sections of this chapter has shown that in terms of statistics such as the mean, median, and the total obese percentage, BMI increased at a faster rate from 1993 to 2001 than since 2001. However, the pattern may be a bit different for very high BMI values. The highest BMI values are important as they are the cases that are likely to have the greatest risk of obesity related health conditions, with probably also the highest risk of developing multiple obesity related health conditions. Therefore, some specific analysis was done on high BMI values and this is described in this section.

The percentiles analysis in Section 4.6 shows that the BMI values of the higher percentiles keep increasing at a high rate in recent years, whereas the lower percentiles values tend to increase less quickly than in earlier years or even stop increasing. Also, the higher BMI values tend to have the highest increase in BMI from one distribution to the next. In the charts of the distributions for men and women in Figures 4.1 and 4.5 the right tail has continued to extend over time.

Some indication of the prevalence changes can be seen in the BMI category percentage values given in Tables 4.1 and 4.2. The total of the obese category percentages tends to increase at a lower rate over time as shown in Figures 4.14 and 4.20. However, there are differences in the individual categories and charts of these are given in Appendix 4.II in Figures 4.II.1 – 4.II.8. From Tables 4.1 and 4.2 and the charts in Appendix 4.II, the obese I percentages tend to increase at a lower rate over time (Figure 4.II.4), and this is also the case for obese II for the data for women (Figure 4.II.5). However, the severely obese category percentages show a fairly similar and consistent increase over time for both men and women (Figure 4.II.6). This is also the case for obese II for men (Figure 4.II.5).

High BMI values for the annual HSE data are analysed earlier in this report in Section 2.7 of Chapter 2. Further analysis was done on the four year data used for the BMI distributions and is set out here. Values for "high BMI" were chosen for men and women so that the overall prevalence in 2013 is about 2.5%. This is to give a balance between looking at the more extreme high values (with 2.5% being, of course, 1 in 40 people) and having a reasonable sample size. This also means that the data is not just those cases where the weight of the person exceeds the limit for the weighing scales (Section 2.6) and so also includes a good proportion of measured cases from the HSE data. The BMI cut-off values used to give about 2.5% for 2013 are 38.6 for men and 41.3 for women.

The data used here was the actual data rather than the smoothed distributions. Using the smoothed distribution data would have required interpolation and it was considered better to calculate the actual percentages from the data for this particular analysis. Also, additional analysis was done on the prevalence by age. Looking at characteristics such as age for the cases requires using the data before smoothing. This is because the smoothing averages out the relative frequencies for the 0.5 BMI intervals and so the link with data for the individual cases is lost. The final data before smoothing was used. This includes the adjustments so that the overall data follows a standard population age profile (Section 3.3). The data with case weights before this adjustment was also analysed and the differences were very small. Simply counting the data (i.e., weighting each case equally) also gave similar results.

The total percentages for men and for women were calculated for the groups of four years of data used for each BMI distribution. Figure 4.25 shows the results. The labels on the chart refer to the distribution year. For example, 2013 refers to the data used for the 2013 distribution which is HSE years 2011-2014. There is a strong and fairly consistent increase across the whole time with a slightly greater increase for men than for women (between 1993 and 2013: 0.5% to 2.5% for men, 0.9% to 2.5% for women).

In order to look further into the increase in the high BMI values an analysis was done of the prevalence by age to see whether this has changed over time. For each distribution the prevalence of the high BMI values in different age groups was calculated. Figures 4.26 and 4.27 show the results for men and women respectively for all the distributions. The pattern is generally similar for men and women across all the years with the highest prevalence usually in the age categories of 40s, 50s, and 60s and low prevalence for age 80 and over. There is no obvious major difference in the pattern over the years, with similar increases in each of the age categories.

To look at this further, the pattern of the increases over time were examined by age category. To make the different age categories more comparable, each value was divided by the mean prevalence for its age category across the six year values. For example, each prevalence for age group 18-29 was divided by the mean of the six distribution year percentages for age 18-29. This rescales the prevalence values so that each age category is on a similar scale.

Charts of the values for the age groups up to 60s are in Figures 4.28 and 4.29. This shows a very similar pattern in the increase for each age group. They all look approximately linear over this time period with a very similar relative rate of increase. This implies that the causes of the increase in high BMI values are probably not particularly age related but apply across the population, such as general lifestyle changes. Factors affecting obesity are discussed in Chapter 6.

The older age groups of 70s and 80s have more irregularity as the number of cases is small and so these are not included in Figures 4.28 and 4.29 to make the charts clearer. These age groups still generally fit the same pattern.

Overall, it is interesting that these high BMI values have continued to increase at a similar rate across the whole time period. This contrasts with the pattern of BMI as a whole where the rate of increase has reduced in recent years.

Further work could be done using the HSE data to examine more details of these cases to try and identify any particular characteristics of the cases, such as demographics. This might be useful to give some indication of the causes of the changes, or in finding the best ways of targeting interventions and help.

More detailed analysis of how BMI varies with age, including taking into account birth year, is in Chapters 7-10.

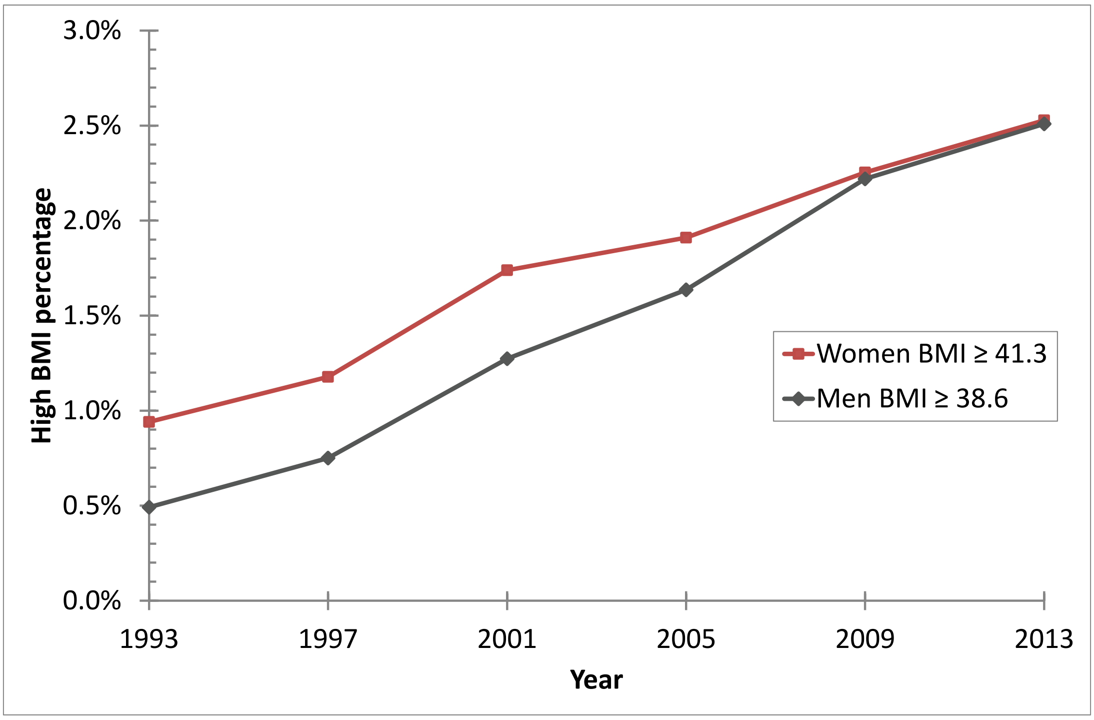


Figure 4.25 Percentage of high BMI values in the groups of four year data.

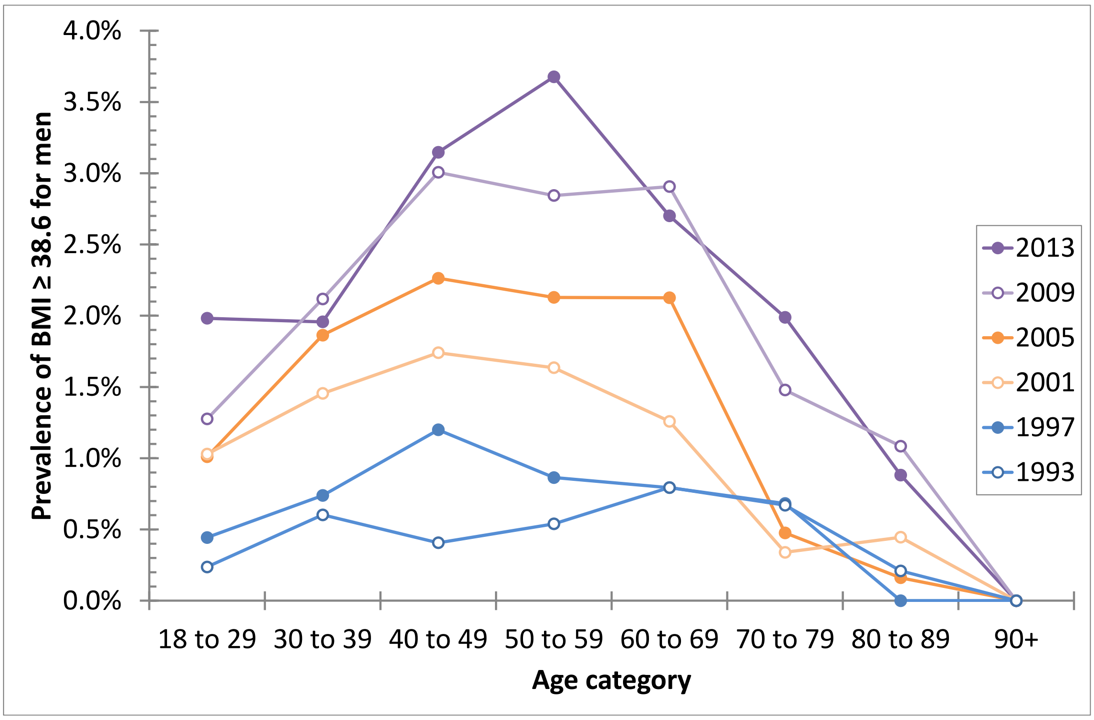


Figure 4.26 Percentage of high BMI values in different age groups for men.

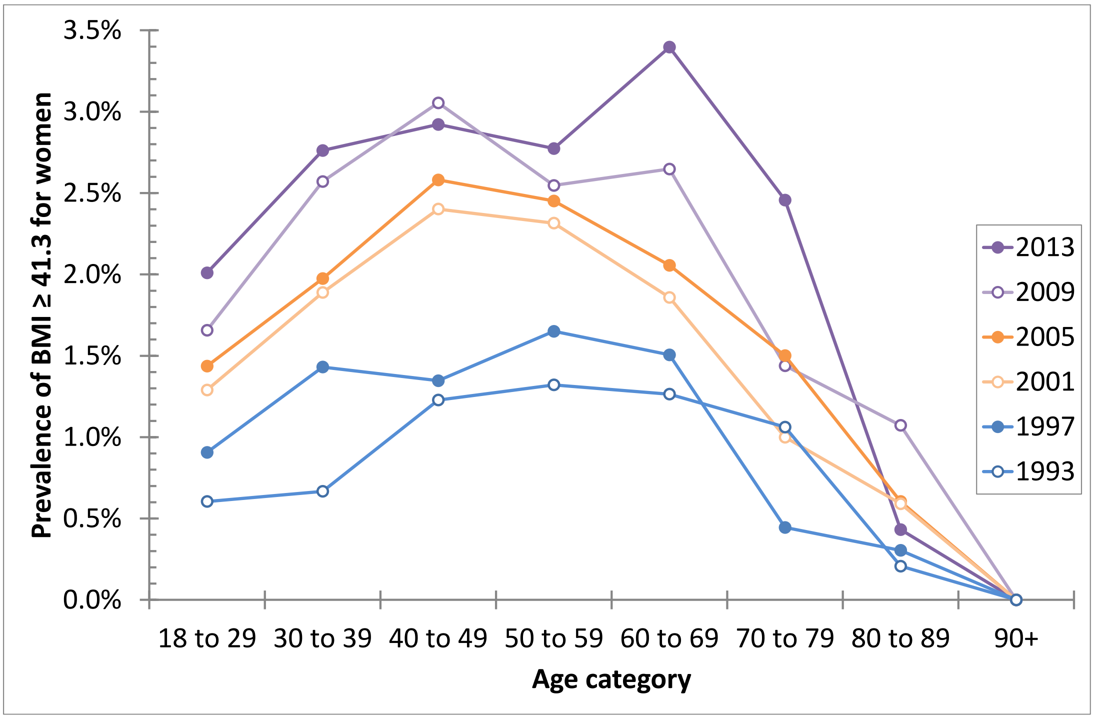


Figure 4.27 Percentage of high BMI values in different age groups for women.

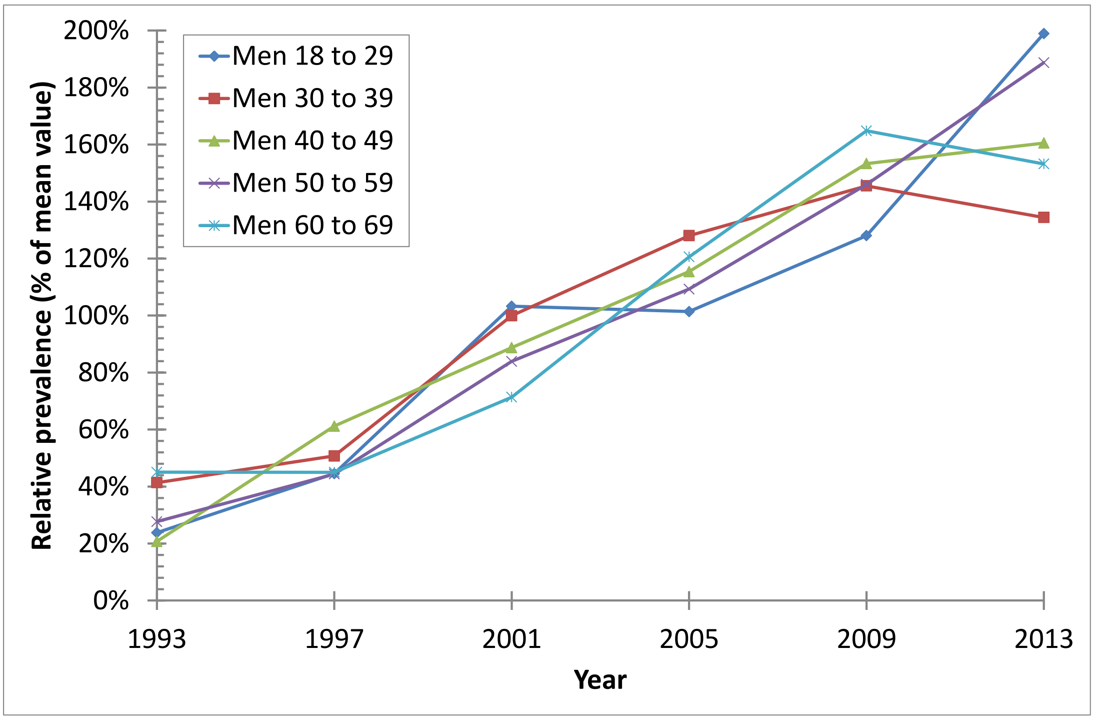


Figure 4.28 Relative prevalence of high BMI values in different age groups for men.

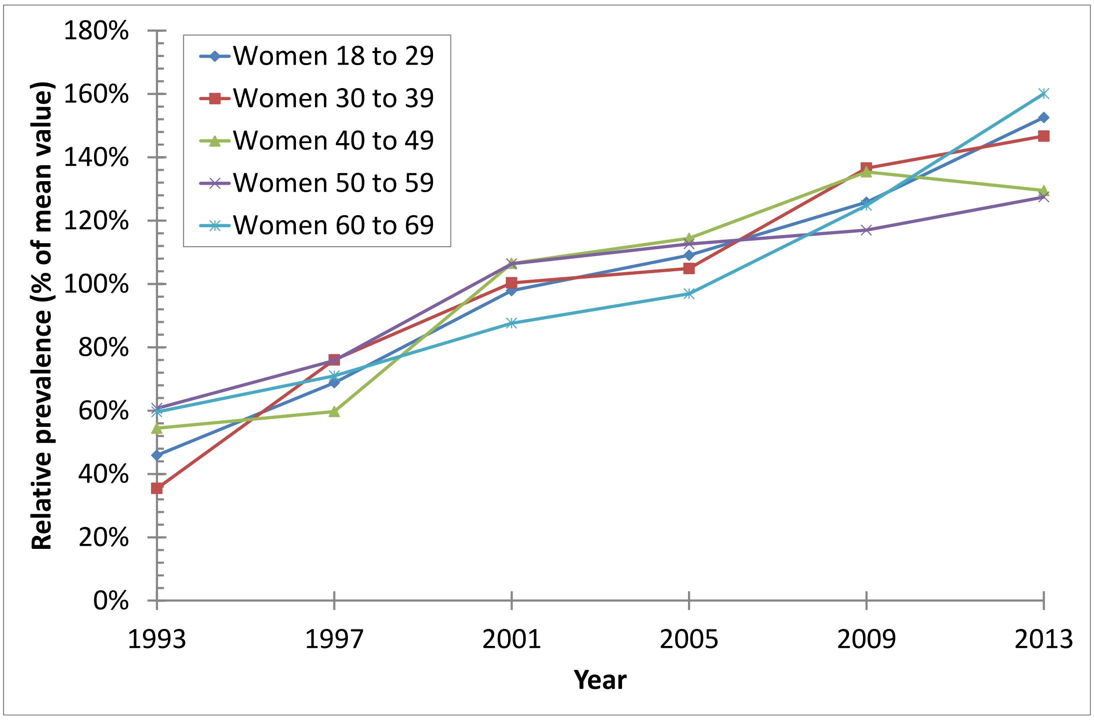


Figure 4.29 Relative prevalence of high BMI values in different age groups for women.

4.8 Overall change from 1993 to 2013 for men and for women

This section looks at the overall change across the time period of the HSE data used in this chapter by comparing the 1993 and 2013 distributions. To some extent this has been covered already in the previous sections of this chapter. For example, Figures 4.1 and 4.5 include the distributions for 1993 and 2013 for men and for women, whilst Tables 4.1 and 4.2 include the statistics for both years. However, the discussions here focus more specifically on the total changes from 1993 to 2013.

4.8.1 Changes in the distributions from 1993 to 2013

Plotting just the 1993 and 2013 distributions on their own helps to see more clearly the extent of the differences over time. This is done in Figure 4.30 for men and Figure 4.31 for women. The distributions change very considerably in the 20 years' time interval. For the distributions for both men and women, from 1993 to 2013 the end of the left tail stays much the same. The peak of the distribution becomes much lower and moves to the right. The right tail is higher and extends much more to the right. Naturally, this reflects the discussions earlier in Section 4.2 and Section 4.3 on the changes between successive distributions over the HSE time interval. A comparison for men and women of the distributions for 1993 and 2013 is in Section 4.9.

The differences between the 2013 and 1993 distributions can be plotted in the same type of chart as Figures 4.3 and 4.7, and the resulting chart is Figure 4.32. The differences over time are very similar for men and women even though the distributions for a particular year are quite different (discussed further in the next Section 4.9).

For men, the percentages are lower up until a BMI value of 28.0, where the curves for 1993 and 2013 cross, with the total difference being 12.7%. In other words, in the 2013 distribution compared to 1993, 12.7% fewer men have a BMI < 28.0, and 12.7% more men have BMI ≥ 28.0.

For women, the curves cross at a BMI of 27.5 with the total difference being 10.5%. Hence, 10.5% fewer women in 2013 compared to 1993 have a BMI < 27.5, and 10.5% more women have BMI ≥ 27.5.

Again, these values are very similar for men and women and the size of the percentage differences reflects large changes in the distributions.

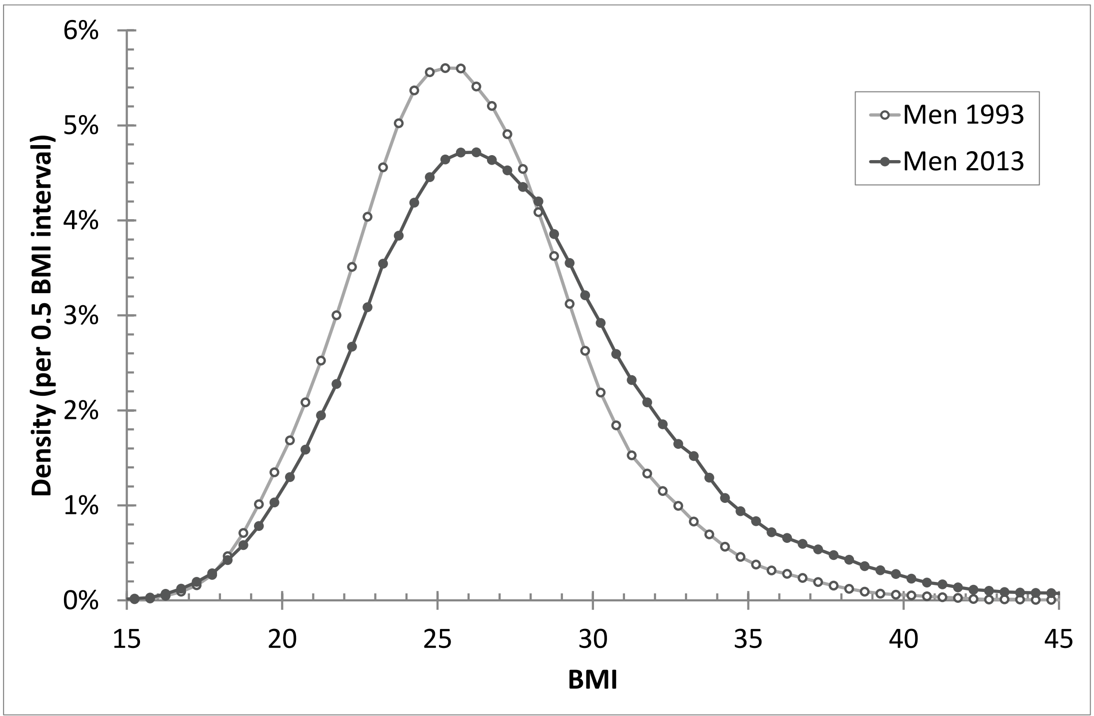


Figure 4.30 BMI distributions for 1993 and 2013 for men.

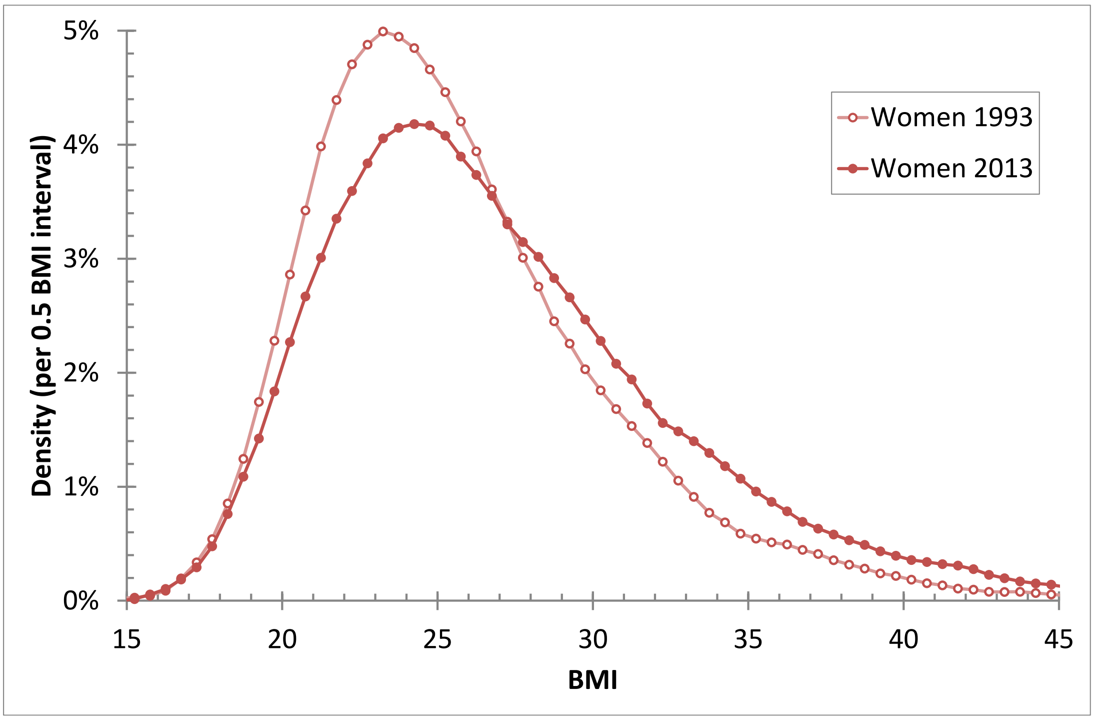


Figure 4.31 BMI distributions for 1993 and 2013 for women.

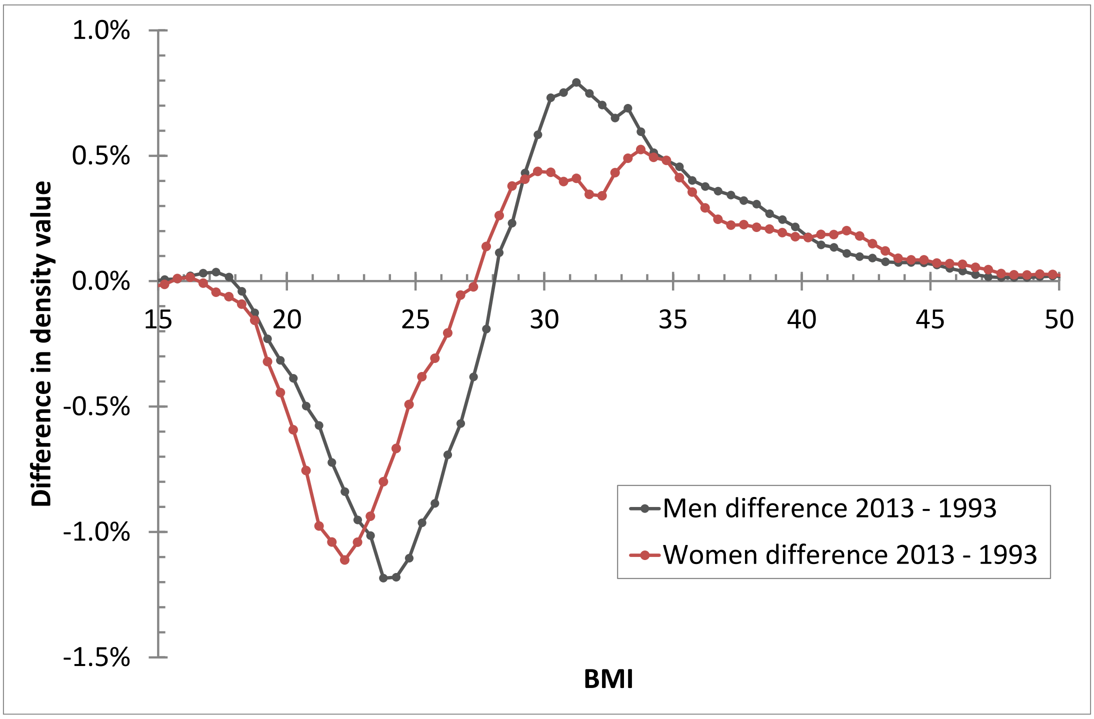


Figure 4.32 Differences between the BMI distributions for 1993 and 2013.

4.8.2 Changes in the statistics from 1993 to 2013

The descriptive statistics and BMI category percentages for each distribution from 1993 to 2013 are in Tables 4.1 and 4.2. The differences from 1993 to 2013 were calculated from those values and are given here in Table 4.3. The changes are substantial and are very similar for men and women.

The mean BMI increases for both men and women by about 1.4. The standard deviation values also increase by about the same amount for men and women (0.97 and 0.82) as the distributions get more stretched out to the right in 2013 compared to 1993.

For the BMI categories, the healthy category percentage reduces for both men and women by about 9%. The actual changes are from 40.4% to 31.3% for men (Table 4.1) and from 49.0% to 39.6% for women (Table 4.2). Each of the obese category percentages increase, with the total for all the obese categories being increases of 11.4% for men and 8.9% for women.

As discussed in Section 4.7, the prevalence of high BMI values has increased by large factors. The percentage increases for severely obese in Table 4.3 are quite small values but they represent large changes in relative terms as the values in 1993 are very small. Specifically, the increases are from 0.3% to 1.7% for men and from 1.4% to 3.4% for women.

Bar charts comparing the category percentages are included in Appendix 4.II in Figures 4.II.9 and 4.II.10. The bar chart shapes are naturally similar to the distributions themselves in that each bar is the combined percentage of a particular interval of BMI values.

Table 4.3 Change in the descriptive statistics for the population distributions from 1993 to 2013

	Men	Women
Mean	1.37	1.36
Median	1.12	1.18
Modal interval mid-point	1.00	1.00
Standard deviation	0.97	0.82
Skewness	0.14	-0.26
Underweight (BMI < 18.5)	0.1%	-0.2%
Healthy (18.5 ≤ BMI < 25.0)	-9.1%	-9.3%
Overweight (25.0 ≤ BMI < 30.0)	-2.3%	0.6%
Obese I (30.0 ≤ BMI < 35.0)	6.7%	4.3%
Obese II (35.0 ≤ BMI < 40.0)	3.3%	2.5%
Severely obese (BMI ≥ 40.0)	1.4%	2.0%

4.8.3 Percentile increases from 1993 to 2013

The percentile increases from 1993 to 2013 were also calculated as the difference between the percentile values for 1993 and 2013 (each calculated as explained in Section 4.6.1). The results are shown in Figure 4.33, which plots the increases in the percentile values from 1993 to 2013 against the starting value in 1993. The values are for all the 10% values along with 1%, 5%, 95%, and 99%. One striking aspect, as noted before in the previous sections, is how similar the changes are for men and women.

For both men and women, the increase for the left most value for 1% is close to 0 with the increase values being -0.12 and 0.12 respectively from 1993 BMI values of about 18 (18.43 and 17.75 respectively). Then each subsequent point in the series has a larger increase in approximately a linear pattern.

The gradient of the percentile increases is quite steep with large increases in BMI for the higher BMI values. For example, the 50th percentiles have BMI increases of 1.12 and 1.18 for men and women respectively from 1993 values of 25.76 and 24.88. The 70th percentiles have increases of 1.57 and 1.77 from values of 27.70 and 27.41. The 90th percentiles have increases of 2.65 and 2.72 from values of 30.93 and 32.18. Finally, the 99th percentiles have increases of 4.92 and 3.56 from values of 36.95 and 41.17. These are considerable increases, and the weight values that these correspond to are discussed in Section 4.10.

Hence, the pattern of the way that BMI has increased over time from 1993 to 2013 is that the increases have been much greater on average for higher BMI values. As mentioned in Section 4.6.3, the interpretation is discussed in Chapter 6, but the implication is that changes in various obesity factors over time have affected those with a higher BMI much more than those with a lower BMI.

The percentile increases having this approximately linear pattern starting from the x-axis indicates that the changes in the BMI distributions are approximately a linear scaling transformation, and this is examined further and modelled in Chapter 5.

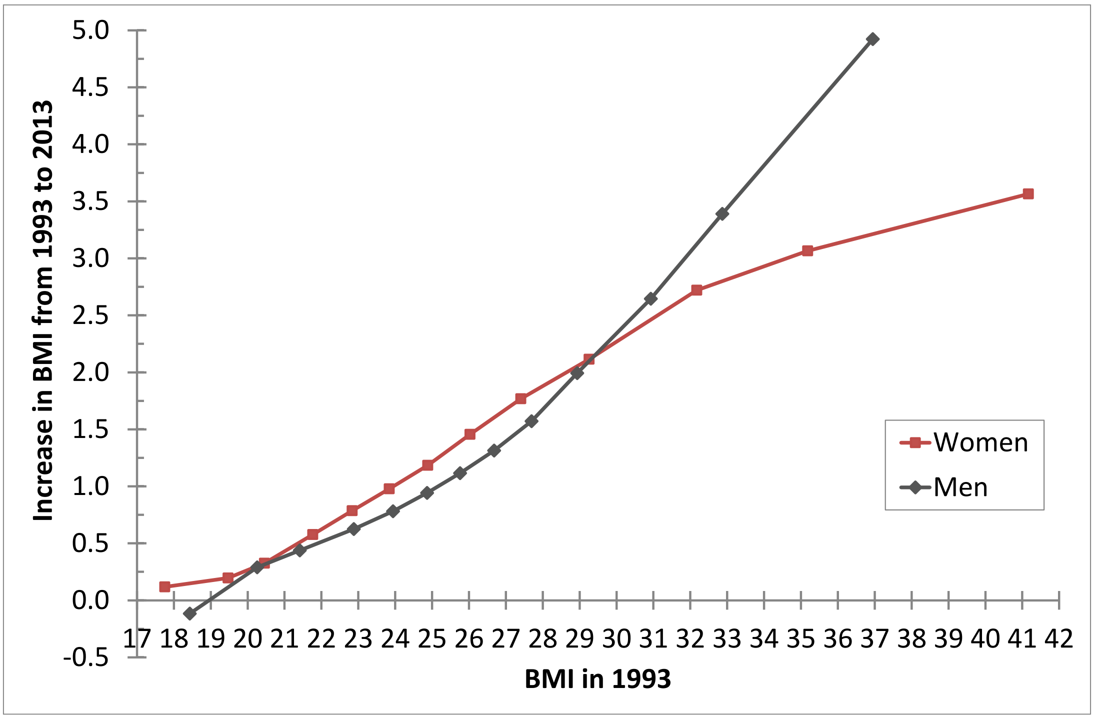


Figure 4.33 Percentile increases from 1993 to 2013 for men and women.

4.8.4 Comparison with the percentile results of Wardle and Boniface (2008)

The approach taken in this report in calculating the changes over time in the percentile values is similar to that used in earlier work by Wardle and Boniface (2008). They analysed changes in percentile values for BMI and for waist circumference using the HSE data. They compared HSE data for 2002-2003 with HSE data for 1993-1994. They used cases in the age range 18-64. The percentiles they calculated were each 10% value along with 2.5%, 5%, 95%, and 97.5%. The results are presented in the paper as charts of the percentile increases on the y-axis and the mean of the starting and ending values on the x-axis. The format of the charts of percentile values in this report is very similar, although here the x-axis value is the starting percentile value. The pattern of the BMI results in the charts in the Wardle and Boniface paper is similar to those discussed in this chapter with higher percentiles having higher increases in approximately a linear relationship.

There are differences in the details of the approach in the work here compared to that used by Wardle and Boniface. The work here uses four years of HSE data rather than two years. It also smooths the data to produce estimated population BMI distributions and obtains the percentile values from the distributions using interpolation in the intervals, whereas Wardle and Boniface used the HSE data directly. The age range applied in both cases starts at age 18 but there is no upper limit here whereas the Wardle and Boniface analysis had an upper limit of age 64. The end point values here are 1% and 99%, compared to 2.5% and 97.5%, but the other percentiles used are the same. The charts here use the percentile value from the starting distribution as the x-axis value whereas Wardle and Boniface used the mean of the values from the start and end data.

There are therefore a lot of differences in the precise details of the method but none of them should make a big difference to the results. The Wardle and Boniface results for 1993-1994 to 2002-2003 were compared with the data here. The results should be similar to the differences in the percentile values from the 1993 distribution (using 1991-1994 data) to the 2001 distribution (using 1999-2002 data) and to the 2005 distribution (using 2003-2006 data). The data used by Wardle and Boniface is not exactly the same with different combinations of years and with age being limited to a maximum of 64. As mentioned above the method is also slightly different. However, the results should be reasonably close. The values from the Wardle and Boniface results were estimated by measuring the points on the charts in their paper against the axis values on their charts. The x-axis values were adjusted to the format used here of the percentile value in the starting data rather than the mean of the start and end data. This adjustment made was simply to deduct half of the percentile increase from the x-axis value in the paper. This is a small adjustment and does not make much difference to the values. Charts comparing the results are in Appendix 4.IV in Figures 4.IV.2 and 4.IV.3. The results are indeed very similar and the Wardle and Boniface values are generally closest to the increases in the population distributions here from 1993 to 2005. The general data source is the same and so the results were expected to be close to each other. However, as with any replication or comparative work (Section 1.6), the similarity gives added confidence to the calculations done for both sets of results.

The percentile results in this chapter therefore extend the previous results of Wardle and Boniface (2008). New HSE data has become available since the time of the paper and the results in this chapter use data up to 2014. In particular, the percentile results in Section 4.8.3 are the percentile increases over a longer time interval from 1993 to 2013. Chapter 14 takes this further in producing distributions for 2017 and analysing the percentile increases from 1993 to 2017 (Section 14.4.5). This chapter also includes the percentile increases between each of the distributions calculated and so enables a comparison of the increases over each four year time interval. The work here also has a wider scope in producing population BMI distributions and having other analyses of the changes. As discussed in this chapter, the pattern of the percentile increases is similar to that of a linear scaling transformation and scaling models are applied to model the changes in the distributions in Chapter 5.

The Wardle and Boniface paper has some analyses that are not done here. They calculated 95% confidence intervals for the percentiles using a bootstrapping approach. The percentile values in the paper come from the data directly whereas the percentiles here are calculated from a smoothed population distribution. Potentially, a bootstrapping approach could also be used for the results here to obtain confidence intervals. This would involve multiple bootstrap sampling with replacement from the

HSE data and then for each bootstrap sample applying the full process (from Chapter 3) of creating the smoothed distribution and from this the interpolated percentile values. This could be done as future work.

Wardle and Boniface also calculated percentiles for waist circumference. The results are similar to those for BMI with the percentile increases mostly being higher for higher percentiles. The pattern is not quite as close to a straight line as for BMI with the general gradient of the increases for the lower percentiles being greater than for the higher percentiles. Estimating the values by measuring the charts in the paper, and adjusting the x-axis values to be the starting value as explained above, the 2.5% value for men is an increase of 0.8 cm in waist circumference from a starting value of 74 cm. The values then increase steeply to the 40th percentile with an increase of 3.7 cm from a starting value of 90 cm. The values then only get slightly larger up to the highest value for 95% with an increase of 4.9 cm from a starting value of 112 cm. For women the lowest percentiles for 2.5% and 5% have increases of 1.7 cm and 1.8 cm from starting values of 61 cm and 66 cm. The gradient of the values is quite steep up to the 60th percentile with an increase of 5.1 cm from a starting value of 81 cm. The highest value for the 95th percentile has an increase of 6.6 cm from a starting value of 103 cm. The values for 2.5% and 5% for women are the smallest values but they are quite a bit greater than zero and so this indicates some increases in waist circumference across the whole distribution even at the lowest values. The lowest percentile for men for 2.5% has an increase of 0.8 cm and so this is closer to the x-axis.

Wardle and Boniface compared the percentile increases for two social categories for manual and non-manual occupation. The results are very similar for the two categories. They also compared the results for age categories of 18-44 and 45-64. The chart of the results for BMI for the age categories is not included in the paper but the comments in the paper are that the increases for higher percentiles tend to be larger for the younger age group. A chart of results for waist circumference is in the paper with the percentile increase values being larger for the younger age group for all the values (although with some overlapping confidence intervals). There is a comment in the paper that the difference between the age groups is greater for waist circumference than for BMI.

In the work in this report, the relationship between age and BMI is examined in Chapters 7-10. This includes some analysis of percentile increases over time for different age groups in Sections 8.3.2 and 9.3.2. These use narrower age ranges than those in the Wardle and Boniface paper but wider ranges of HSE data, although as mentioned above the Wardle and Boniface paper does not show the BMI percentile results for the age categories. Regarding social categories, Chapter 15 analyses BMI distributions and the changes over time for the five deprivation variable categories that are given in the HSE data since 2001. These are similar general topics to the analysis in the Wardle and Boniface paper mentioned in the previous paragraph, but the results are not comparable directly.

The discussion in Wardle and Boniface (2008) emphasises the importance of understanding precisely how population BMI has changed because different BMI levels will have different levels of risk of health conditions, and also because it can give insights into the causes and possible mechanisms of the changes in BMI. There are also health cost implications. The aim of the work in this report is the same. As mentioned above the percentile results here hopefully give useful extensions of the results in Wardle and Boniface (2008).

4.9 Comparison of the BMI distributions for men and women

4.9.1 Differences in the distributions for 1993 and 2013

The BMI distributions for men and women for 1993 and 2013 are shown in the previous section in Figures 4.30 and 4.31. The distributions are plotted together on a single chart in Figure 4.34. There are clear differences in the shapes of the BMI distributions for men and women.

The distributions for women for 1993 and 2013 have their peak at a smaller BMI value compared to the distribution for the same year for men. The peak is also lower. For 2013, the peak value for women is 4.18% at a BMI of 24.25, whereas for men it is 4.72% at a BMI of 26.25. For 1993, the peak for women is 4.99% at a BMI of 23.25 compared to 5.60% at 25.25 for men.

The distributions for women have higher frequencies of BMI values compared to men throughout most of the left tail up to about the peak value of the distribution for women. Then the frequencies are higher for men around the peak of the distribution for men. There is then higher prevalence for women than for men at the end of the right tail. This pattern means that the BMI distributions for women are more spread out than for men having more variability with a higher standard deviation and a higher skewness (values given in Tables 4.1 and 4.2).

The differences between the distributions are plotted in Figure 4.35, which shows the value for the distribution for women less the value for the distribution for men for each 0.5 BMI interval. These are just the differences between the curves in Figure 4.34. The pattern of the differences is very similar for the two years, although slightly smaller in magnitude for 2013. As discussed in the previous paragraph, the frequencies for women are higher in the left tail and the end of the right tail, but lower in the middle. These intervals and the total difference in frequencies (frequency for women – frequency for men) are shown in Table 4.4. Positive frequencies mean that the value for women is higher, negative that the value for women is lower. The region is labelled in brackets by whether the value for women or men is higher (W / M). Hence, for 2013 there are 9.4% more women than men with BMI < 24.0, 12.5% fewer with 24.0 ≤ BMI < 33.5, and 3.1% more with BMI ≥ 33.5.

Table 4.4 Intervals over which the frequency values for women are higher or lower than for men.

Region	1993 interval	1993 frequency	2013 interval	2013 frequency
Left tail (W)	BMI < 23.5	11.1%	BMI < 24.0	9.4%
Middle (M)	23.5 ≤ BMI < 31.0	-14.7%	24.0 ≤ BMI < 33.5	-12.5%
End of right tail (W)	BMI ≥ 31.0	3.6%	BMI ≥ 33.5	3.1%

In summary, compared to the equivalent distribution for men, the distribution for women has a smaller peak at a lower BMI value, with higher proportions in most of its left tail and at the end of its right tail.

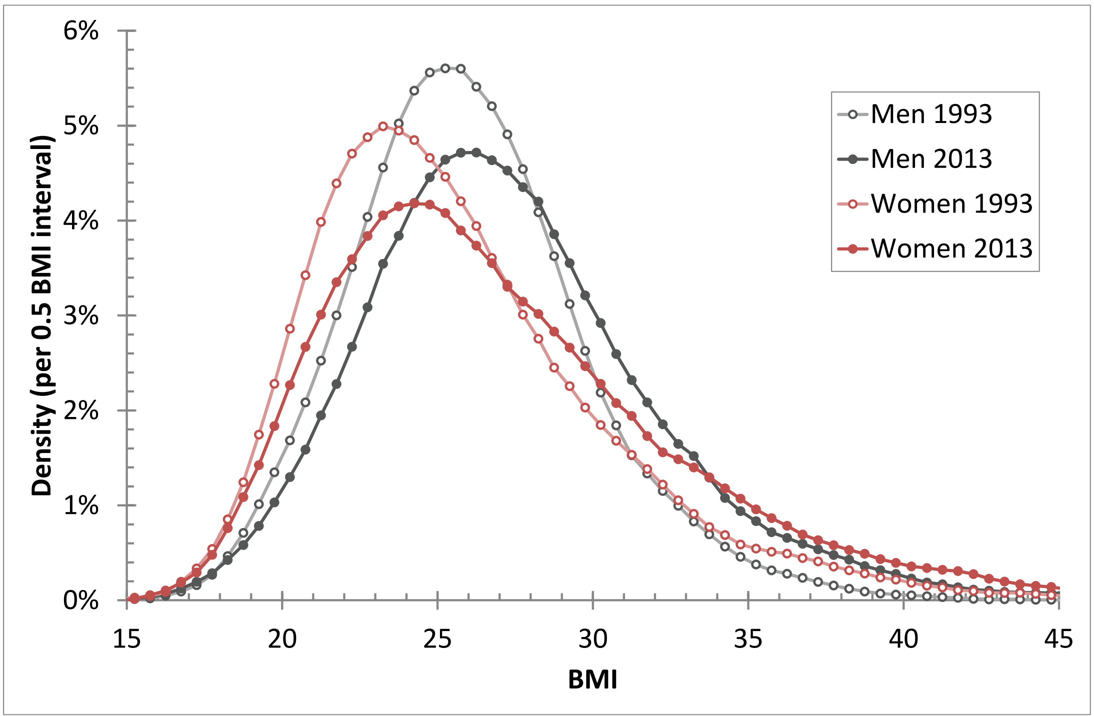


Figure 4.34 BMI distributions for men and women in 1993 and 2013.

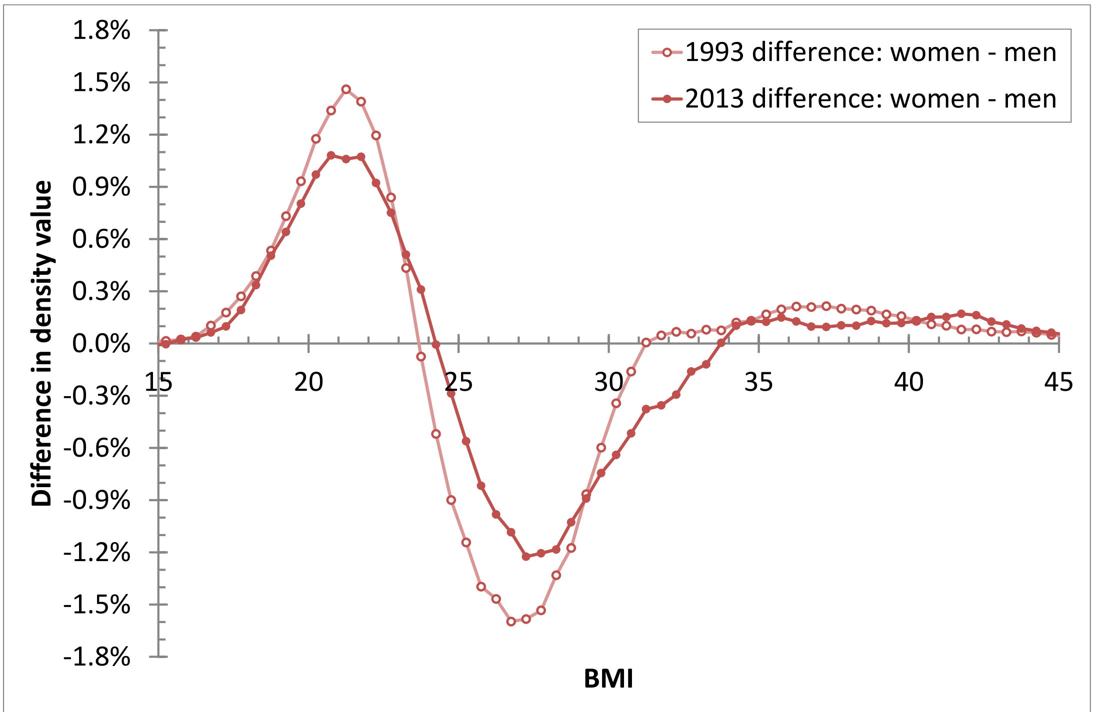


Figure 4.35 Differences between the distributions for men and women in 1993 and 2013.

4.9.2 Comparison of the category percentages and statistics for each year

The BMI category values are given in Tables 4.1 and 4.2, and time series are plotted together for men and women for each category in Appendix 4.II. The differences in the shapes of the distributions for men and women means that the category percentages differ. In particular, the regions where the curves are relatively higher or lower, as discussed above in Section 4.9.1, is naturally reflected in the category percentages. With one minor exception, for all years the category percentages are higher for men for overweight and obese I, and lower for all other categories. The one exception is that the obese I category percentage is 0.1% less for men in 1993 than for women. The differences between the distributions for men and women are very consistent across the six years, as shown by the charts in Appendix 4.II. Across the years, the mean differences for each category for women relative to men (women percentage – men percentage) are:

underweight 0.9%, healthy 8.7%, overweight -11.2%, obese I -1.6%, obese II 1.7%, severely obese 1.5%.

In particular, the healthy category percentage for women is consistently about 9% higher than the percentage for men. This is a big difference with far fewer men being in the healthy range. Instead, the percentage of men that are in the overweight category is about 11% higher than for women. The obese II and severely obese categories with the highest BMI values are more prevalent for women. Across all the overweight and obese categories the mean difference is 9.6% higher prevalence for men. For 2013 this value is 9.1%.

Despite these differences, there is considerable similarity in the changes over time for men and for women. This has been evident in the discussions throughout this chapter where very much the same patterns are described for the changes in the distributions for men and for women. For example, the changes in the shape of the BMI distribution over time for both men and women are that the left tail stays much the same while the distribution stretches out more and more to the right. The overall percentile changes from 1993 to 2013 for men and women in Figure 4.33 are very similar.

The descriptive statistics can also be compared. Figures 4.36 and 4.37 plot the mean and median values and their changes on the same charts for men and women (combining Figures 4.9 and 4.15, and 4.10 and 4.16). The mean and median values are higher for men, particularly the median, but the changes over time between the six distributions (Figure 4.37) are very similar.

The earlier discussions in this chapter have noted that the distributions for both men and women have considerable increases in BMI between 1993 and 2001 with much smaller increases after that. One aspect that is a little different is that the biggest change for men is from 1997 to 2001 followed by 1993 to 1997, whereas these two are about the same for women.

4.9.3 Implications of the comparison of the distributions

The differences in the distributions for men and women mean that there is a lower percentage of men in the healthy BMI category and a higher percentage overall in the overweight and obese categories. The difference in each year is about 9%. In that sense, the issue of overweight and obesity is more widespread for men.

On the other hand, the prevalence of the highest BMI values is greater for women with about a 3% higher frequency in the obese II and severely obese categories. Whilst there are these differences in prevalence, there still considerable numbers in all the overweight and obese categories for both men and women.

The relationship with health conditions is outside the scope of this report, although some of the literature was mentioned in Chapter 1. However, the difference in the distributions means that potentially obesity is a more common issue for men for conditions related to some degree of overweight or obesity, but a more common issue for women for conditions related mainly to very high BMI levels. This also depends on whether the same BMI value has the same relationship with disease risk for the particular condition for men and women.

Despite the differences in the BMI distributions for each year, it is striking how similar the changes over time have been in the distributions for men and women. The factors that might have affected BMI, such as changes over time in lifestyle and the general environment, are discussed in Chapter 6. The similarity in how the distributions have changed may well indicate that the changes in the factors have tended to have much the same effect on the BMI of both men and women.

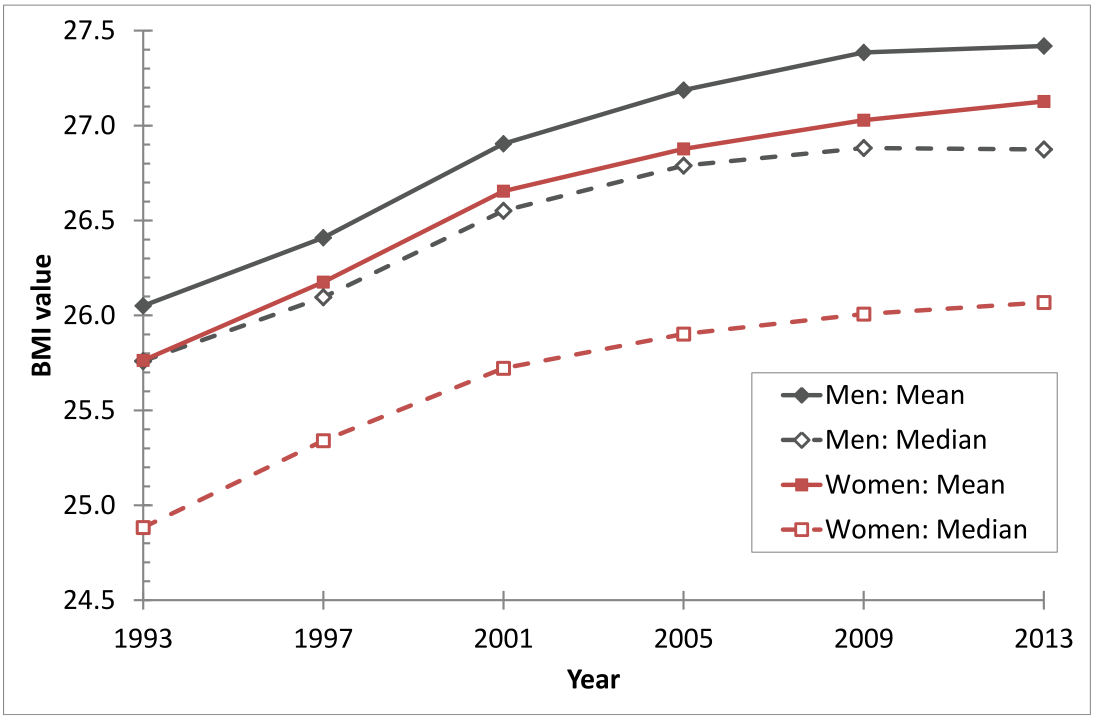


Figure 4.36 Mean and median values for the distributions for men and women.

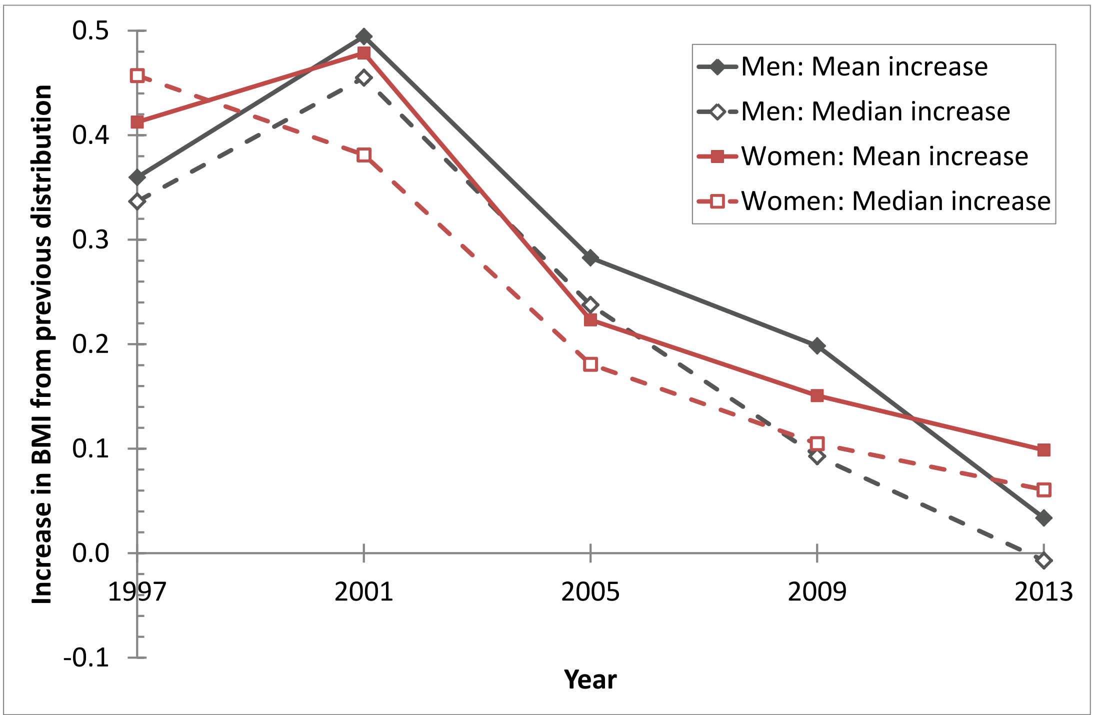


Figure 4.37 Changes in the mean and median values for the distributions for men and women.

4.10 Weight values for the BMI changes

The relationship between a person's BMI and their weight is a simple equation but one that depends on their height. As set out in Section 1.2.1, the equation for BMI is: $b = w / h^2$ where b is the BMI value, w is the weight in kg, and h is the height in metres. The equation for BMI can be re-arranged easily to show how weight is related to the BMI value (Section 1.2.3): $w = bh^2$. As discussed in Sections 1.2.3 and 1.2.4, for typical average height values one unit of BMI corresponds to quite a large difference in weight, and the standard BMI categories are a wide range of weights.

Weight is a much more familiar concept than BMI and so it can be useful to consider what the typical weight values are for some of the BMI changes over time to get a better understanding of the magnitude of the changes. Any height values could be used, but the distributions in this report are for the adult population of England and so it makes sense to use the mean population heights. Height tends to reduce slightly with age for the adult population as shown in Table 13.2 in Section 13.3 of this report that gives values from SACN (2012), and also shown in Section 15.3.2. From the values in Table 13.2, overall mean adult height values are 1.75 m (5 ft 8.9 in) for men and 1.62 m for women (5 ft 3.8 in).

Any of the BMI values or the BMI changes could be converted into weight values using a given height. Here, some of the overall changes from 1993 to 2013 in Section 4.8 are analysed. A weight value in kg is obtained from a BMI value by multiplying it by h^2 and so, using the mean height values, this is $1.75^2 = 3.0625$ for men and $1.62^2 = 2.6244$ for women. Hence, for these mean height values the increases in mean BMI from 1993 to 2013 of 1.369 for men and 1.364 for women correspond to increases in weight of $1.369 \times 3.0625 = 4.19$ kg (9.2 lbs) for men and $1.364 \times 2.6244 = 3.58$ kg (7.9 lbs) for women. Hence, this is another way of assessing the magnitude of the increase and again shows that these are substantial changes.

Using the same height values and therefore the same factors, the percentile increases from 1993 to 2013 in Figure 4.33 were also converted into weight values. Both the starting BMI in 1993 and the BMI increase values were altered to weight values. This was done for units of kg, with the resulting chart being Figure 4.38, and for units of stone (st) and pounds (lbs), with the resulting chart being Figure 4.39 (note: 1 stone = 14 lbs).

Some example values of the 50th, 70th, and 90th percentile values on the charts are given in Tables 4.5 and 4.6. Again, as for the mean values, these are substantial increases. For example, the 90% increases are both greater than one stone, at 17.9 lbs and 15.7 lbs respectively.

The changes could also be expressed as percentages, which are the same values whether the variable is BMI or weight. For example, the increases for the 50th, 70th, 90th percentiles are 4.3%, 5.7%, 8.6% for men and 4.8%, 6.5%, 8.5% for women. A chart of the percentage changes is given in Figure 4.IV.1 in Appendix 4.IV.

Table 4.5 Percentile increases for men from 1993 to 2013 as weights (using average height).

Percentile	1993 BMI	BMI increase	1993 weight	Weight increase
50%	25.76	1.12	78.89 kg (12 st 6 lbs)	3.42 kg (7.5 lbs)
70%	27.70	1.57	84.82 kg (13 st 5 lbs)	4.81 kg (10.6 lbs)
90%	30.93	2.65	94.72 kg (14 st 13 lbs)	8.10 kg (17.9 lbs)

Table 4.6 Percentile increases for women from 1993 to 2013 as weights (using average height).

Percentile	1993 BMI	BMI increase	1993 weight	Weight increase
50%	24.88	1.18	65.30 kg (10 st 4 lbs)	3.11 kg (6.8 lbs)
70%	27.41	1.77	71.92 kg (11 st 5 lbs)	4.64 kg (10.2 lbs)
90%	32.18	2.72	84.44 kg (13 st 4 lbs)	7.14 kg (15.7 lbs)

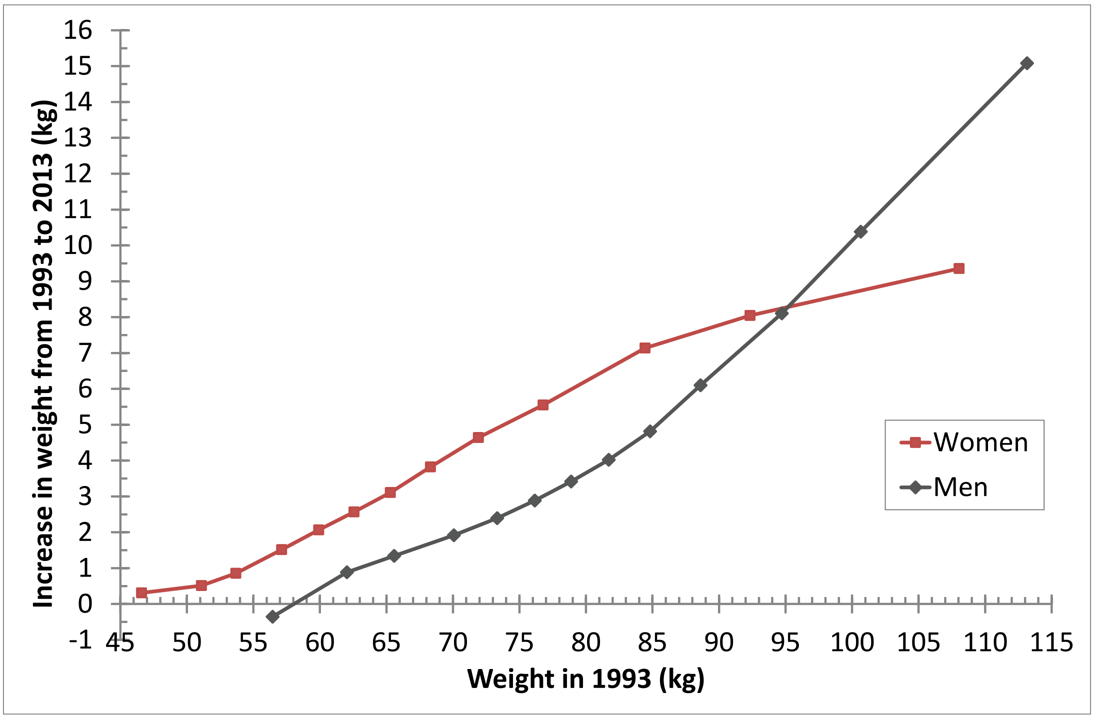


Figure 4.38 Percentile increases in kg (using average height), from 1993 to 2013.

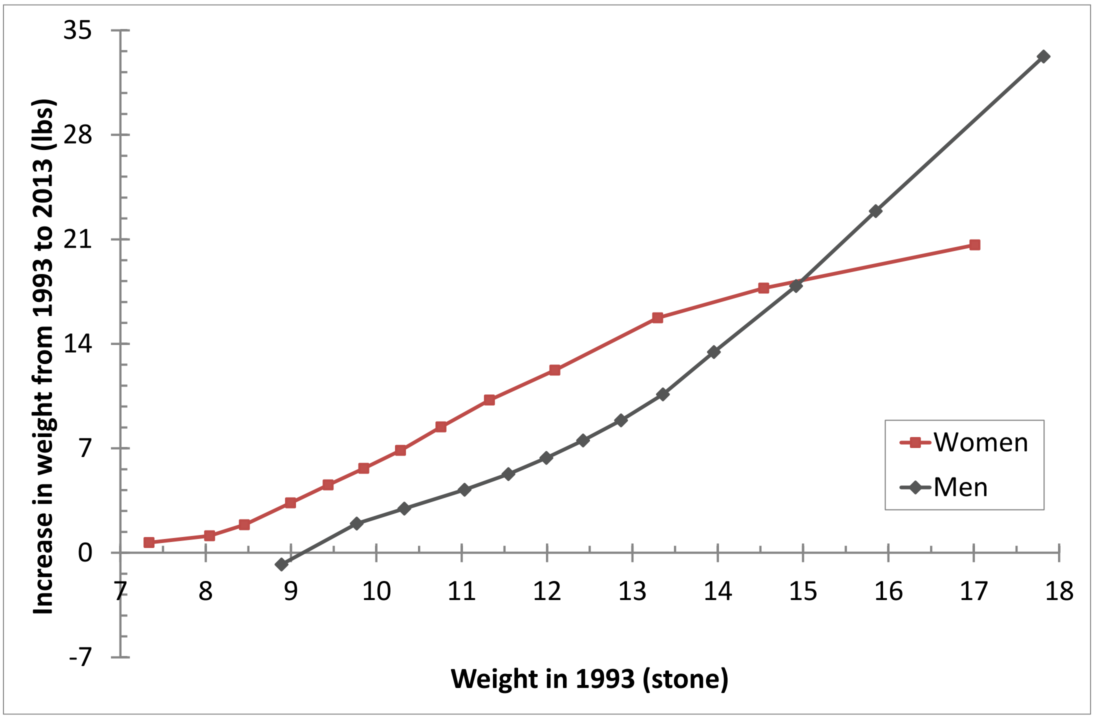


Figure 4.39 Percentile increases in stone and lbs (using average height), from 1993 to 2013.

Chapter 5: Modelling the BMI distribution changes using linear scaling transformations

Key points from this chapter:

- Linear scaling transformations were used to model the changes from one distribution to the next for the BMI distributions set out in Chapter 4. The basic approach was to apply the scaling to each distribution to model the next distribution four years later as closely as possible. Excel solver was used to find the scaling starting point and the scaling factor that gives the best match.
- The scaling transformations are able to model the changes well and give a close match to the target distributions. This shows that the way that BMI in the population has changed in each four year time interval is similar to a linear scaling pattern. This implies that changes in obesity factors over time have affected those with a higher BMI more than those with a lower BMI. The implications and interpretations are discussed further in Chapter 6.
- The magnitudes of the changes over time naturally follow the analysis of the distributions in Chapter 4 with the scaling models having much greater increases in BMI in the 1990s than in recent years. The changes for men and women are similar.
- Scaling transformations were also applied to model the change over the whole time period from 1993 to 2013. Again, this works well in modelling the changes and matching the 2013 distributions. Several alternative models were applied and these all work well and give similar results.

5.1 Linear scaling overview

As discussed in Chapter 4, the changes between the successive four year BMI distributions look like linear scaling patterns. This is because in the charts of the distributions the changes in the shape of the distributions over time are that they get more stretched out. The starts of distributions in the left tails are very similar but the rest of the distribution is stretched out to the right with a lower peak and longer right tail. This corresponds to a linear scaling transformation of the BMI values where the scaling starts at some point in the left tail of the distribution. Another indication of this pattern is that the percentile increases plotted against the starting BMI value are approximately a straight line, where the points along the line have a positive gradient and start approximately from the x-axis (i.e., with the lower percentiles have increases of about 0).

Therefore, linear scaling transformations were applied to see how well they can model the changes from one distribution to the next, and for the whole time interval from 1993 to 2013. This chapter explains what was done and the results.

The general approach was to start with the actual BMI distribution for the initial year. The BMI distribution consists of 0.5 BMI intervals and a percentage frequency for each interval. Charts of the distributions are Figures 4.1 and 4.5 in Chapter 4 and they plot the density value at the midpoints of the intervals (Section 3.1). The density value is the percentage frequency for the 0.5 BMI interval.

The effect of a linear scaling transformation, in general terms, is to spread out the distribution along the x-axis. Scaling was applied to the BMI distributions by altering the start and end points of the intervals using the scaling function. With a scaling transformation, the total percentage frequency for each interval stays the same but the scaling makes the intervals wider. The density values (the y values of the points on the curve) therefore have to be adjusted accordingly for the revised interval widths.

The scaled distribution was then plotted using the new interval mid-points and density values. Each scaling is specified by the starting value and the scaling factor. These two values were chosen by searching for the values that give the transformed distribution that matches the target distribution most closely. Excel solver was used for the search for the optimal values.

5.2 Equation for the linear scaling model

The population BMI distributions give an estimate of the percentage frequency of the population in each BMI interval. Hence, BMI is a random variable with the probability or percentage frequency being given by the distribution. The distribution is the proportion of the population in each interval, or the probability of a randomly chosen person having a BMI value in each interval.

In applying a linear scaling transformation to the BMI distributions, a scaling function is used to transform each BMI value into a new scaled value. This is used here to stretch out the distribution along the x-axis. There is a starting point for the scaling, and the distribution is unchanged to the left of this point and is stretched out to the right of this point.

Hence, there are two parameters for the linear scaling function, which are the starting BMI value for the scaling and the scaling factor.

In the equations here the initial BMI values are denoted by B, and the new transformed values by B_{new}. The scaling starting point is denoted by p, and the scaling factor by s. Then the scaling function is:

$$For\ B \leq p: B_{new} = B$$

$$For\ B > p: B_{new} = p + (B - p) \times s$$

Alternatively, this can be written in a single equation as:

$$B_{new} = B + max(B - p, 0) \times (s - 1) \qquad (5.1)$$

This can be explained further with a numerical example, as follows. This will use one of the models from the full results in the later Section 5.4 as the example. The scaling for women for 1993 to 1997 has a scaling starting point, p, of 17.50 and a scaling factor, s, of 1.063 (Table 5.1). The 1993 BMI distribution for women for the interval 22.5 ≤ BMI < 23.0 has a percentage frequency of 4.878%, which is plotted on the charts at the x-axis mid-point value of 22.75.

For the scaling model distribution, the interval end points are adjusted by the scaling equation. So, they become 22.5 + (22.5 – 17.5) × 0.063 = 22.8150, and 23.0 + (23.0 – 17.5) × 0.063 = 23.3465. The interval width is now 23.3465 – 22.8150 = 0.5315. Therefore, the scaling model has a percentage frequency of 4.878% in the BMI interval from 22.8150 to 23.3465.

For the chart (whether plotted as a histogram or here as a frequency polygon) the y-axis value for this interval needs to be expressed as a density per 0.5 BMI interval (Section 3.1). Hence, the scaling model y coordinate for the interval is 4.878% × 0.5 / 0.5315 = 4.589%. This is plotted at the interval mid-point of (23.3465 + 22.8150) / 2 = 23.081.

Hence, the line plotted for the scaled model distribution for 1997 includes the point (23.081, 4.589%).

5.3 Finding the best scaling

The scaling transformations were assessed by a fitness score to find the optimal values. This was calculated as the weighted sum of the squared differences between the scaled model distribution and the target distribution density values. Therefore, lower fitness values are better.

The intervals used were those of the scaled model distribution. Therefore, the differences were measured at the mid-point of each scaled distribution interval and the squared differences weighted by the width of the interval. The value for the scaled distribution is simply the distribution density value for the interval, that is plotted at the mid-point on the distribution charts. Linear interpolation was used to get the value for the target distribution at that mid-point using the nearest neighbouring points for the distribution.

In other words, the fitness function is the sum of ((scaled model value – actual interpolated value)^2 × interval width. Values for all the intervals up to a BMI value of 50 were used. If the last of these intervals goes beyond 50 then the weighting used for that final interval was just 50 – interval start value.

Excel solver was used to search for the minimum fitness value. It was run several times with different initial values to give more chance of finding the best solution. In some cases all the starting points converged to essentially the same solution. In others the solver stopped at different solutions indicating that there are probably local minima in the fitness space.

The final model was obtained by taking the model with the best fitness value and rounding the values. The scaling starting point BMI value was rounded to the nearest 0.25 and then the scaling factor was adjusted up and down slightly, working to 3 decimal places, to find the value that gives the best fit.

The one exception was the changes for the distribution for men from 2009 to 2013, where a couple of alternative models were considered. The one with a slightly higher (worse) fitness value was preferred, and the reasons for this are explained in Section 5.5.

Other fitness values could have been used such as, perhaps, using the absolute difference or the absolute percentage difference. Even if a scaling model applies very well, there are bound to be small differences due to small irregularities in the shape of the distributions. However, the aim was to try and avoid large differences. The squared difference was used to particularly penalise large differences. The charts of the results in this chapter indicate that the method works well.

An alternative fitness measure could have been to measure the differences at the actual target distribution mid-points rather than the scaled distribution mid-points. This would have required interpolating the scaled distribution values. A slight advantage of this would have been that the differences would be measured at the same BMI values for all alternative model distributions. A disadvantage would be that the scaled points are more spaced out than the actual distribution points and so the interpolation would be greater introducing a little more approximation and uncertainty. The overall approach would be the same and any change in the results would be likely to be negligible.

5.4 Scaling results between successive distributions

The scaling values for the models chosen as the best are shown in Table 5.1 along with the fitness values. The best fitness value for men is 1993 to 1997 even though there is a big change between these two distributions. For women the fitness values are all fairly similar. The precise fitness numbers are quite hard to interpret but the charts and other statistics in this chapter give a visual impression and further information to evaluate the scaling models.

The charts for the resulting distributions are in Figures 5.1 to 5.5. In each case, the aim with the scaling model is to scale the first distribution (the earlier year with the white filled markers) to get as close as possible to the second distribution (the later year with solid grey or solid red markers). The turquoise line shows the model for men and the purple line shows the model for women. In all cases, the scaling model provides a pretty close match to the target distribution. In recent years, the differences between the distributions are small and so the scaling is only trying to model a small change.

Appendix 5.I has some additional charts. It has charts of the results with each distribution in a separate chart. It also has charts of the differences between the scaling model and the target distribution. The difference values give further information and again indicate that the scaling models all do well in fitting the target distributions. In Figures 5.1 to 5.5, differences in the peaks of the distributions tend to be more noticeable to the eye than differences in the tails and so the figures in Appendix 5.I are useful in assessing the differences. The differences are small (all less than 0.2% in magnitude) and there are no particularly long BMI intervals where the scaling distribution is continuously higher or continuously lower than the target distribution. Charts of the cumulative differences are also included in Appendix 5.I.

The main conclusion drawn is therefore that the linear scaling transformations are a good representation of the changes in the BMI distributions for each of the time intervals. They are therefore very suitable models for this situation.

Table 5.1 Final values for the scaling models.

Period	1993 to 1997	1997 to 2001	2001 to 2005	2005 to 2009	2009 to 2013
Men					
Start point	18.25	18.25	23.25	26.25	25.25
Scaling factor	1.043	1.061	1.056	1.052	1.021
Fitness value	3.73×10^{-6}	12.15×10^{-6}	6.88×10^{-6}	7.85×10^{-6}	9.06×10^{-6}
Women					
Start point	17.50	20.25	18.00	17.25	22.75
Scaling factor	1.063	1.068	1.024	1.016	1.023
Fitness value	5.85×10^{-6}	8.79×10^{-6}	5.95×10^{-6}	7.69×10^{-6}	5.91×10^{-6}

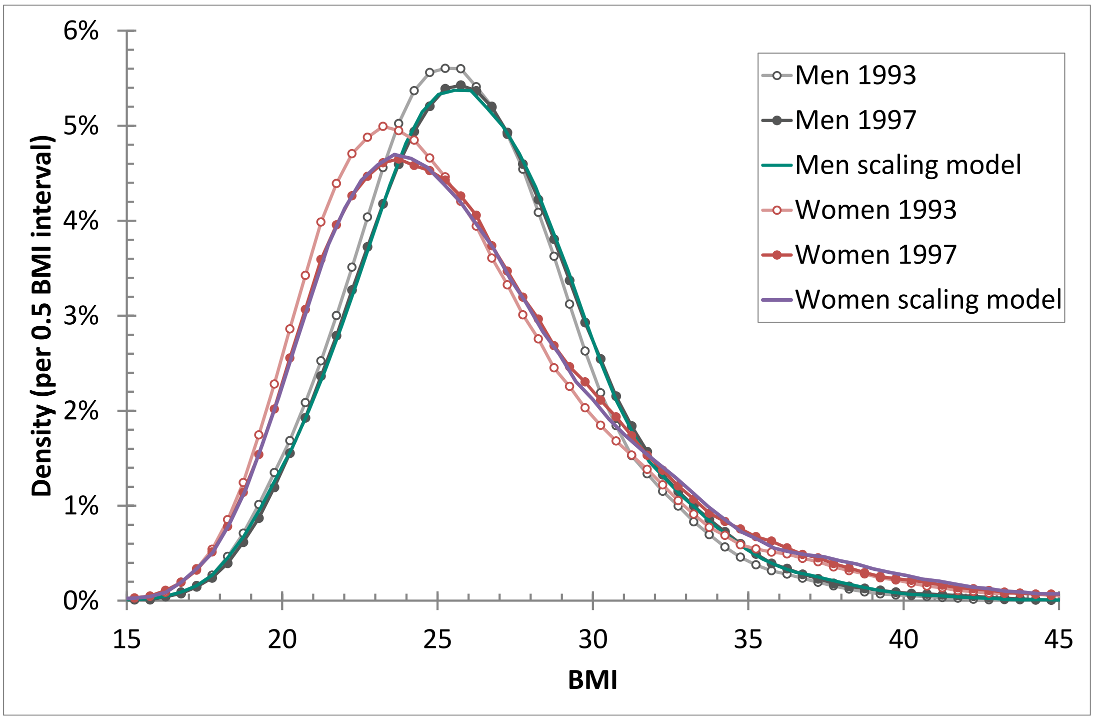


Figure 5.1 Scaling models for men and women for 1993 to 1997.

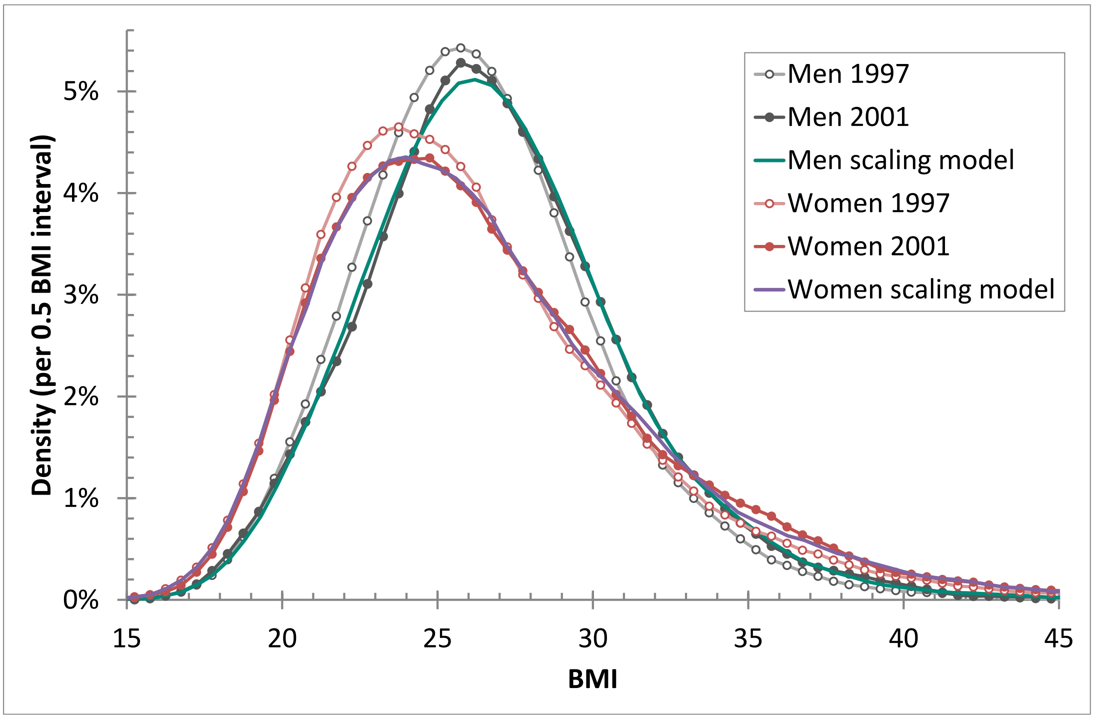


Figure 5.2 Scaling models for men and women for 1997 to 2001.

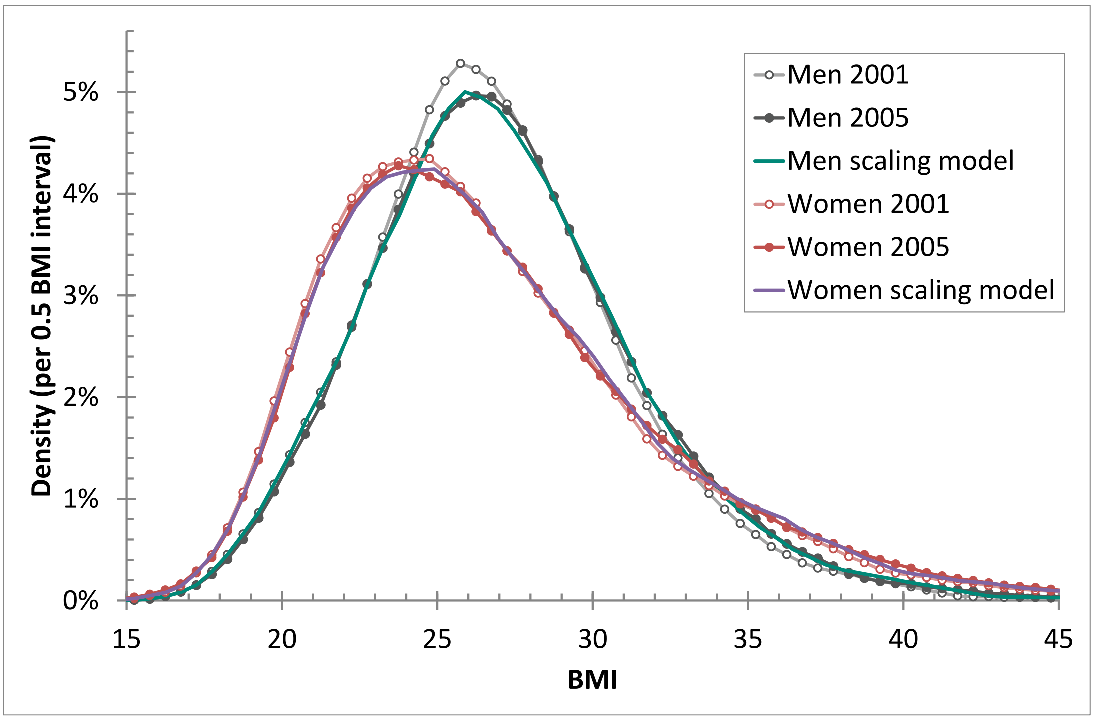


Figure 5.3 Scaling models for men and women for 2001 to 2005.

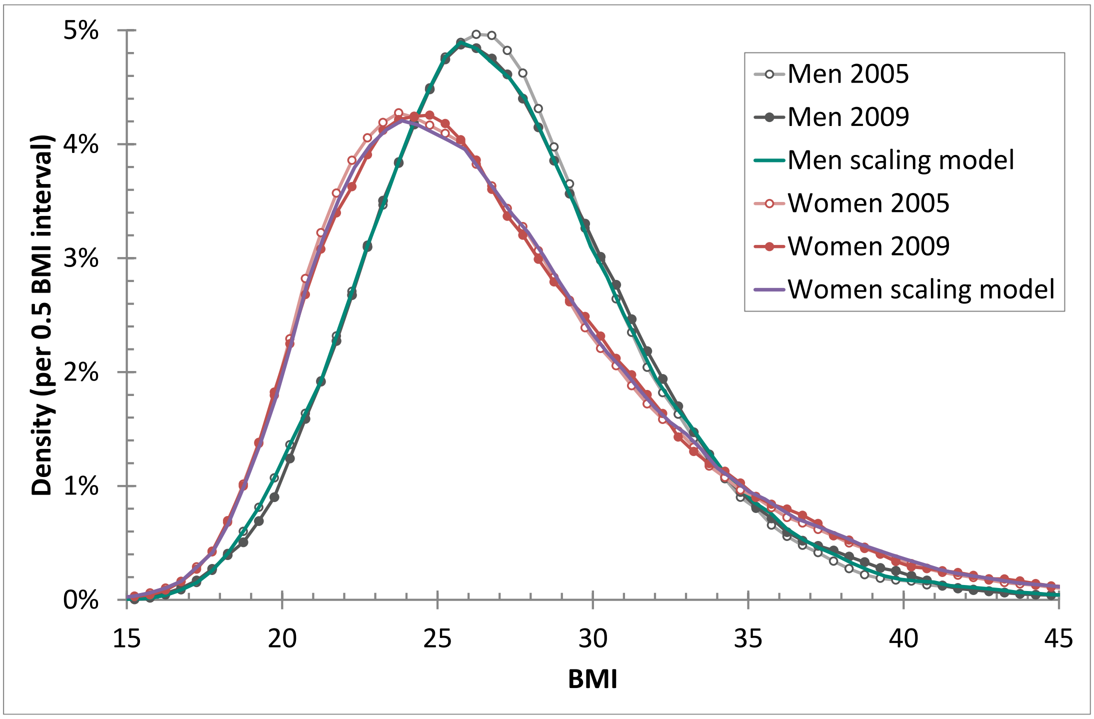


Figure 5.4 Scaling models for men and women for 2005 to 2009.

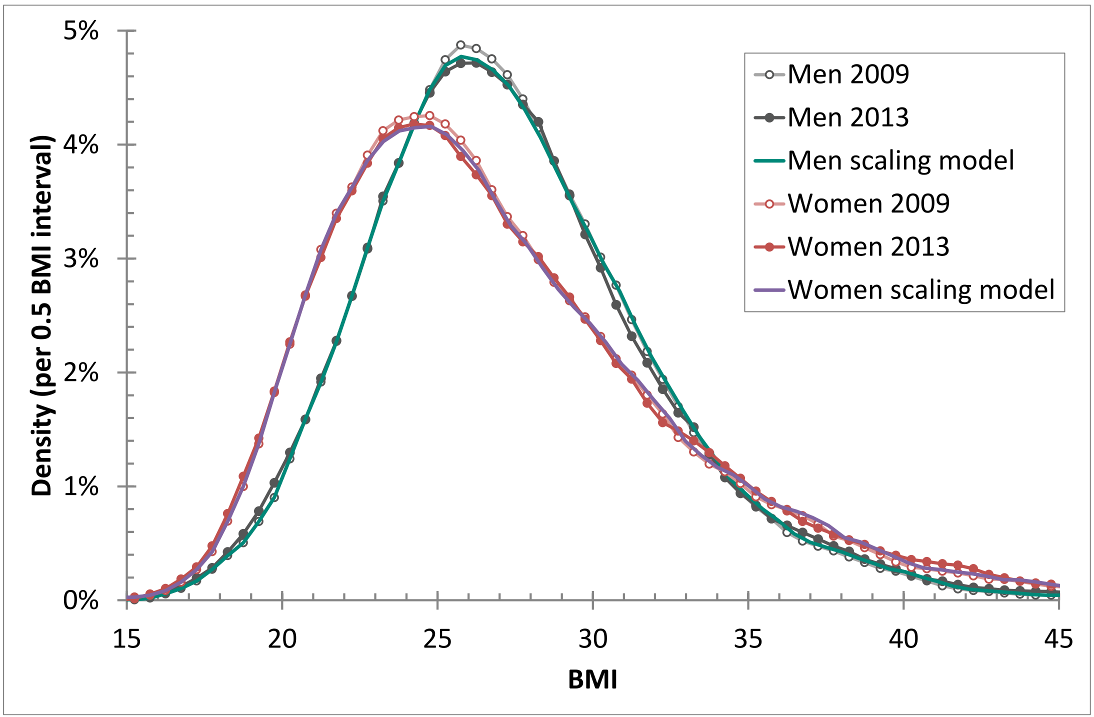


Figure 5.5 Scaling models for men and women for 2009 to 2013.

5.5 Two alternative models for men for 2009 to 2013

The one exception in the results is the model for men for 2009 to 2013. For this scenario there are two scaling models that give very similar fitness values. The model that was chosen and is shown in Table 5.1 matches the peak of the 2013 distribution pretty well. However, it does not quite follow the shape of the right tail being mostly above the actual right tail between BMI values of 30 and 36 and below it for BMI values higher than that (seen most clearly in Appendix 5.I: Figures 5.I.9 and 5.I.19).

The alternative model has a scaling starting point of 30.25 and a scaling factor of 1.077. This model actually has a slightly better fitness value of 7.16×10^{-6} compared to 9.06×10^{-6}. With its high starting point value the alternative model only changes the right end of the distribution and leaves the peak values unchanged at the original 2009 values. Consequently, it does not match the 2013 peak well. However, it gives a good match with the right tail which is the reason for the slightly lower fitness values. The charts for both models are in Appendix 5.II.

The change from 2009 to 2013 is relatively small. The shape of the 2013 distribution is slightly different to 2009 in having a peak that is a bit narrower and so neither transformation fits perfectly. The model chosen and shown in Table 5.1 was preferred as a more plausible model in giving a more even match across the whole distribution.

5.6 BMI increases and comparisons of the scaling models

The scaling model parameters are given in Table 5.1. However, the models can probably be understood most easily by plotting the BMI increases and these are shown in Figure 5.6 for men and Figure 5.7 for women. Since the models do well in matching the target distributions the differences in the models naturally reflect the discussions and results in Chapter 4. In particular, the BMI increases for the scaling models generally match the percentiles calculated previously (Section 4.6) and comparison charts of the scaling model increases and the percentile increases are in Appendix 5.III.

The BMI distributions should be borne in mind in interpreting Figures 5.6 and 5.7, in that the distribution frequency varies a lot across the BMI values. Most people in the population distributions have a BMI between 20 and 35, with the highest frequencies at around 25. Even though Figures 5.6 and 5.7 go up to a BMI of 45, there are relatively few people with a BMI over 40 and so the changes for over 40 will only apply to a small percentage in the population. The models will therefore be quite insensitive to the specific increases in high BMI values and hence the changes for high BMI values should be considered much more uncertain and less accurate than the increases for lower BMI values.

For men, Figure 5.6 shows that the model for 1993 to 1997 has quite large increases in BMI. The increases are even greater in 1997 to 2001, and then generally reduce in each of the next three time intervals. Taking the example of a BMI of 30, the BMI increases for the five time intervals are 0.51, 0.72, 0.38, 0.20, 0.10, respectively.

The gradients for the models for men are similar, particularly for the middle three time intervals, but the scaling starting point tends to get higher. In particular, the models for the last two time intervals have starting points above 25 which means that there will be no change in the model for the proportions in the healthy category. In both cases, the starting points are approximately at the mode of the initial distribution and at or only just before the median value. This means that the left tail and the first half of the distribution has little or no change in the scaling model from 2005 to 2009, and from 2009 to 2013. In other words, increases in BMI occur mainly for the top half of the distribution only.

For women, Figure 5.7 shows that the models for 1993 to 1997 and 1997 to 2001 are similar in giving high BMI increases. The models for the last three time intervals have much lower increases, and are similar to each other. Again taking the example of a BMI of 30, the BMI increases for the five time intervals are 0.79, 0.66, 0.29, 0.20, 0.17, respectively. In this case the starting points are reasonably close to each other, but the gradients are less for the last three time intervals.

Comparing Figures 5.6 and 5.7, there is a considerable degree of similarity between the models for men and women, despite some differences in the details of the patterns.

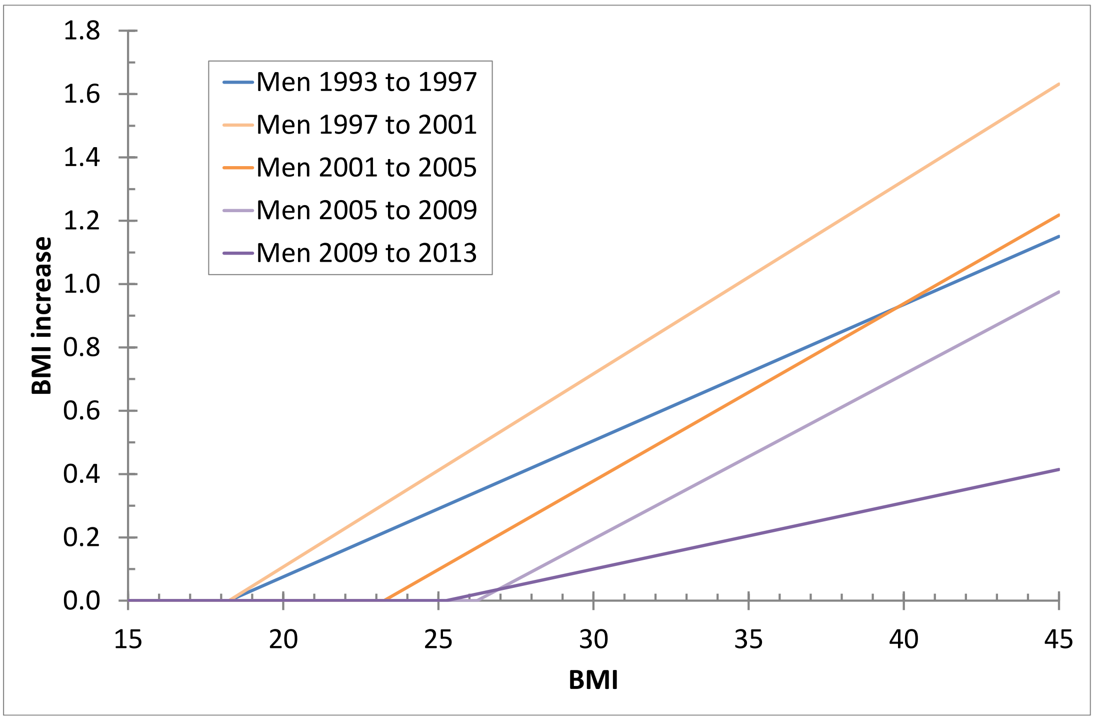


Figure 5.6 BMI increases for the scaling models for men.

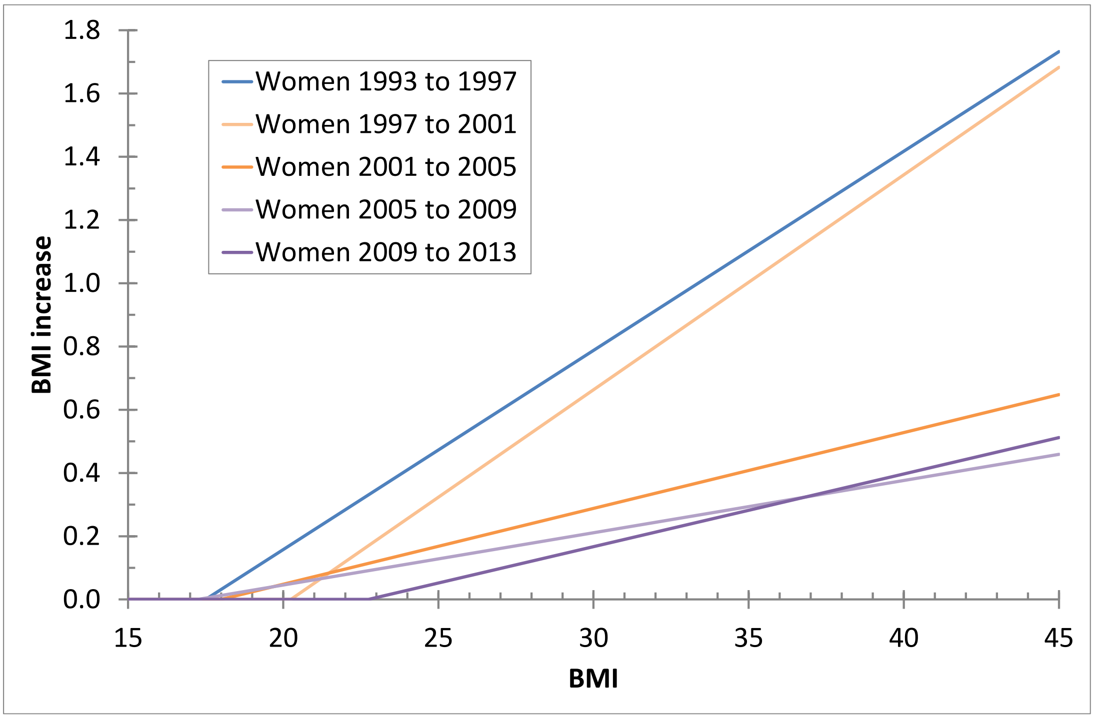


Figure 5.7 BMI increases for the scaling models for women.

5.7 Statistics of the scaled distributions and comparison with actual

Descriptive statistics were calculated for the scaled distributions and these are in Appendix 5.IV. The statistics are all quite close to those for the actual distributions given in Tables 4.1 and 4.2.

The scaled distributions model the change from the previous actual distribution, and the models each start with the actual distribution. Therefore, comparing the changes from the previous actual is the best way to assess the scaling models. Table 5.2 shows the increase in mean and median for the scaled and the actual distributions from the previous actual distribution for men and women. The values for the scaled distributions match the actual values well.

Table 5.2 Increase in the mean and median BMI compared to the previous actual distribution.

	1997	2001	2005	2009	2013
Men - mean					
Actual distribution	0.36	0.49	0.28	0.20	0.03
Scaled distribution	0.34	0.50	0.23	0.11	0.06
Men - median					
Actual distribution	0.34	0.46	0.24	0.09	-0.01
Scaled distribution	0.32	0.48	0.18	0.03	0.03
Women - mean					
Actual distribution	0.41	0.48	0.22	0.15	0.10
Scaled distribution	0.52	0.41	0.21	0.15	0.11
Women - median					
Actual distribution	0.46	0.38	0.18	0.10	0.06
Scaled distribution	0.47	0.35	0.19	0.14	0.07

The changes for the percentages in the BMI categories can also be compared. The values are in Appendix 5.IV but are shown in charts in this section. For men, Figure 5.8 shows the changes in category percentages for each actual distribution (the same as Figure 4.12) and Figure 5.9 shows the changes in each scaled distribution from the previous actual. The scaled distribution can be assessed by comparing the two charts to see how similar the values are. They do look much the same which is further evidence of the scaled distributions doing well in modelling the changes. For women, the corresponding charts are Figures 5.10 and 5.11. Again, the changes for the actual distribution (Figure 5.10, which is the same as Figure 4.18) are very similar to those for the scaled distribution (Figure 5.11).

Along with the charts in Section 5.4, the main finding is that a simple linear scaling transformation is a very good model of the changes in the BMI distributions for each time interval for both men and women.

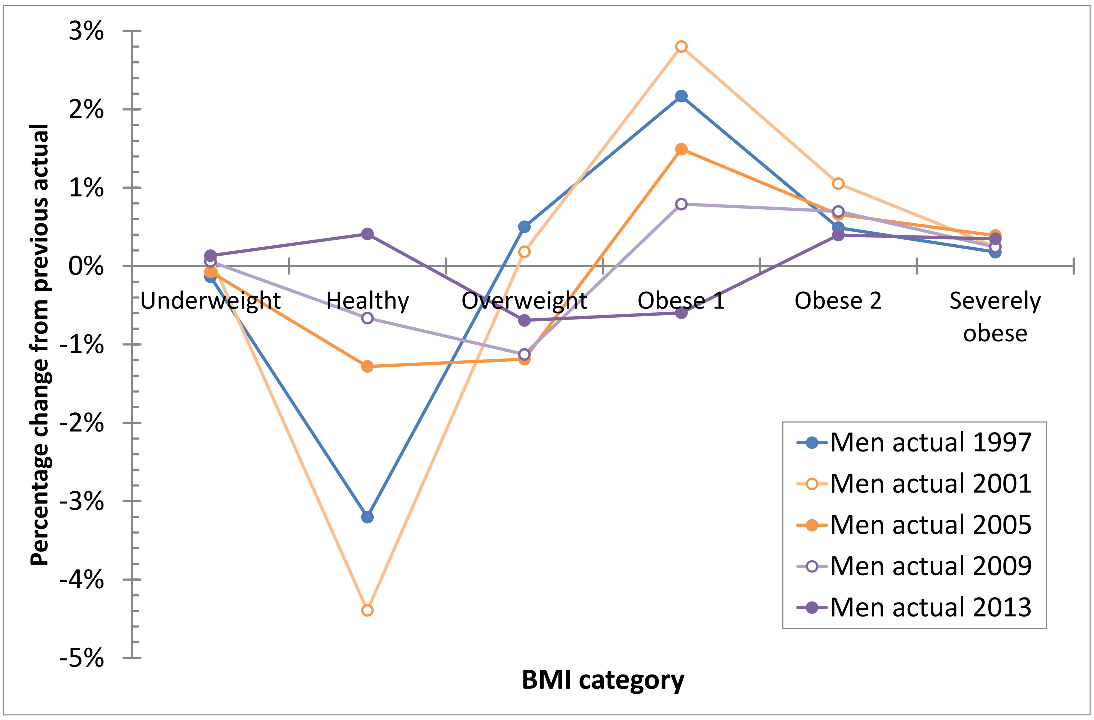


Figure 5.8 Changes in the category percentages for the actual distributions for men.

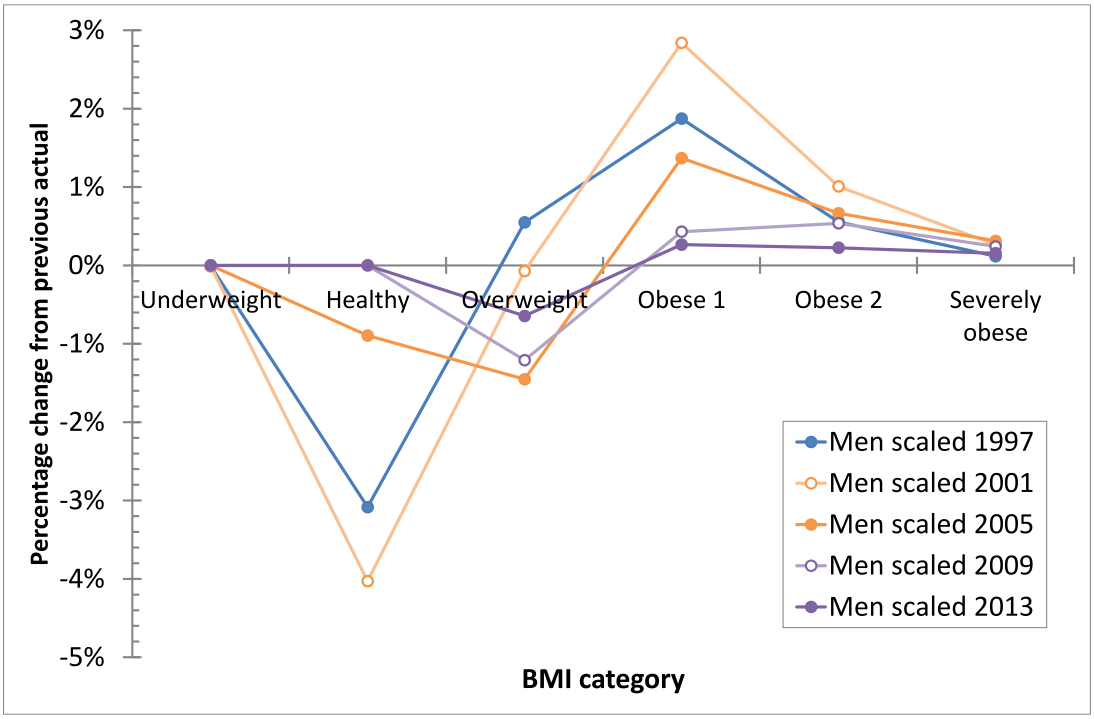


Figure 5.9 Changes in the category percentages for the scaled distributions for men.

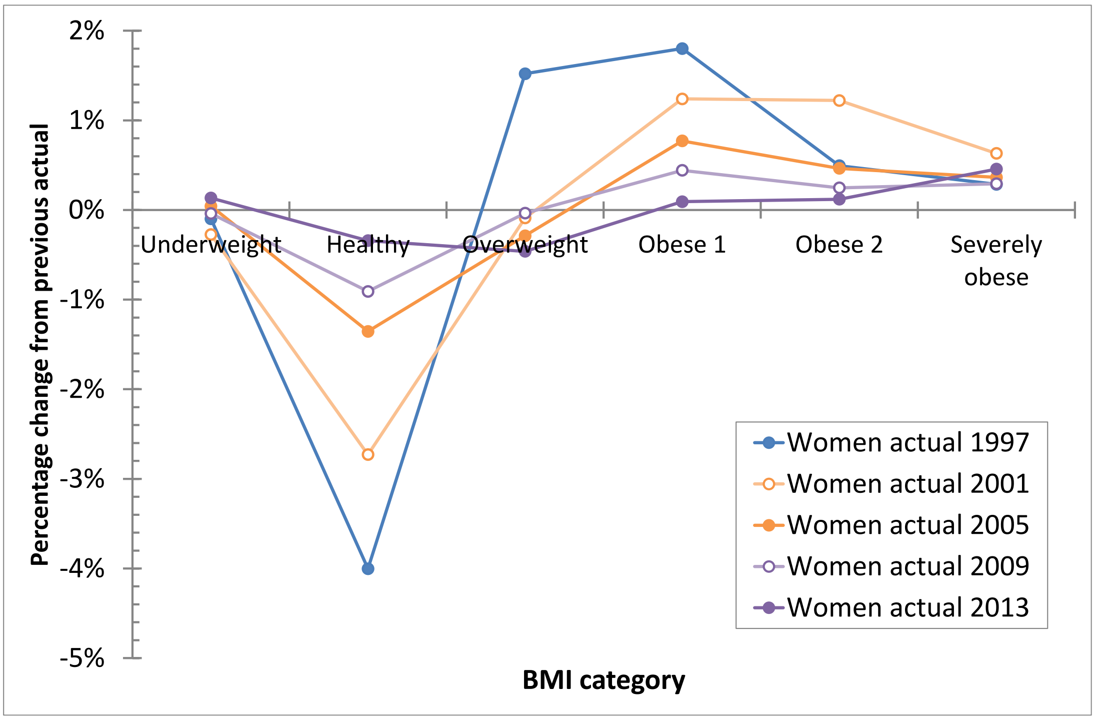


Figure 5.10 Changes in the category percentages for the actual distributions for women.

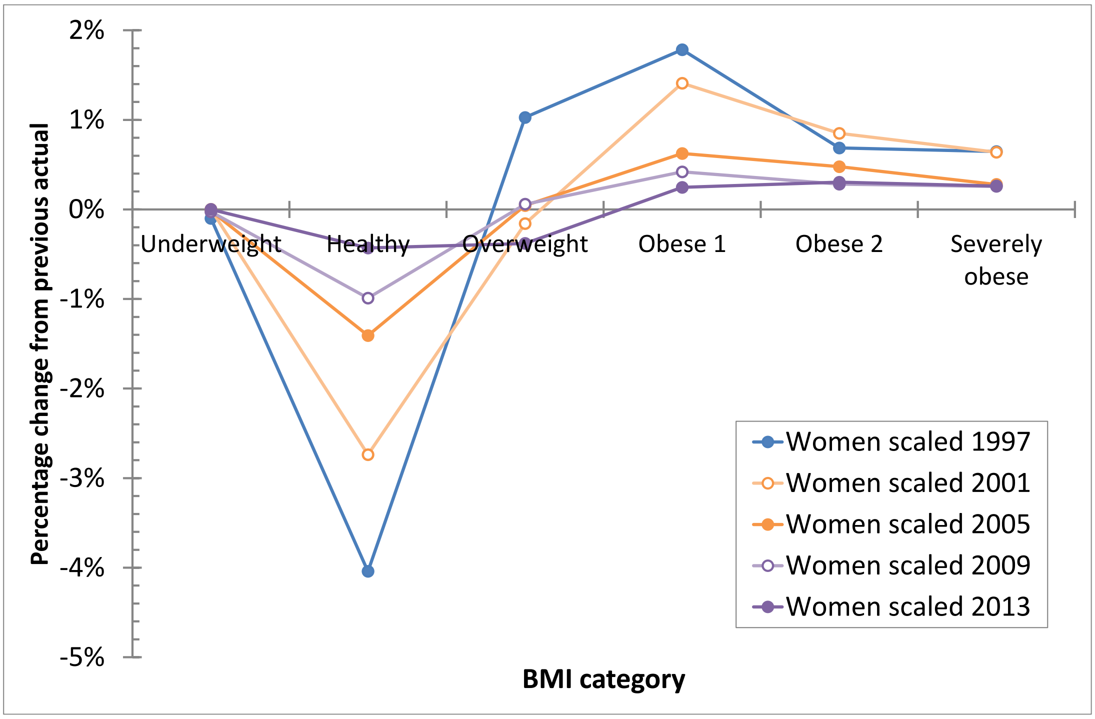


Figure 5.11 Changes in the category percentages for the scaled distributions for women.

5.8 Scaling models for 1993 to 2013

The scaling model approach can be applied to model the whole time interval being considered from 1993 to 2013. This can be done in two ways. One is to combine the scaling models derived for each distribution as set out in Section 5.4. The other is to fit a single scaling model from 1993 to 2013. Both of these were done and are described here in Sections 5.8.1 and 5.8.2. Some other more complex models are considered in Section 5.8.3.

5.8.1 Combined model for 1993 to 2013

The combined model takes the scaling for each time interval in Table 5.1 and combines them together. This produces a model for the overall time interval from 1993 to 2013. Comparing the result against the actual 2013 distributions gives an additional assessment of how well the individual scaling models are doing in modelling the changes.

The 1993 actual distribution has 0.5 BMI intervals and a percentage for each interval. To produce the overall model for 2013 the scalings were applied in turn to the interval boundary values. To give an example, consider the 1993 BMI value of 20 for men. This needs to be scaled up by successively applying the scaling models. From Table 5.1 the 1997 value becomes 18.25 + (20 – 18.25) × 1.043 = 20.075 to 3 decimal places. The next scaling model is then applied to this value as 18.25 + (20.075 – 18.25) × 1.061 = 20.187 to 3 decimal places. This value is below the start point for the next three scaling models and so is the final 2013 value. Applying the scaling models in this way to all the boundary values produces revised BMI intervals. The percentage for the interval from the 1993 distribution stays the same and the density value (the height of the curve) is adjusted for the revised width of the interval after scaling.

The resulting models for men and for women are shown in Figure 5.12. The fitness values are 18.42×10^{-6} for the model for men and 5.51×10^{-6} for the model for women. Both models do pretty well in matching the 2013 actual distributions, particularly given the large changes from 1993 to 2013. The model for women gives a slightly closer match and a lower fitness value than the model for men.

The BMI increases for the combined models are shown in Figure 5.13. The increases are very close to a linear pattern for women since the scaling starting points are similar for each model (Table 5.1). The variation in starting points for the models for men (Table 5.1) produces a piecewise linear curve for lower BMI values.

The increases are large for high BMI values. However, it is important to bear in mind that most of the 1993 BMI values are between 20 and 35 with very few BMI values over 40. Therefore, the BMI increases for the higher BMI values in Figure 5.13 are quite uncertain.

One interesting aspect is the similarity between the BMI increases for men and for women. This is despite the distributions for men and women being very different. As also noted in Chapter 4, this indicates that the changes in obesity factors over time affecting BMI seem to have had a very similar effect on the BMI of both men and women. The interpretation of the changes is discussed further in Chapter 6.

Overall, the combined model works well in modelling the changes from 1993 to 2013.

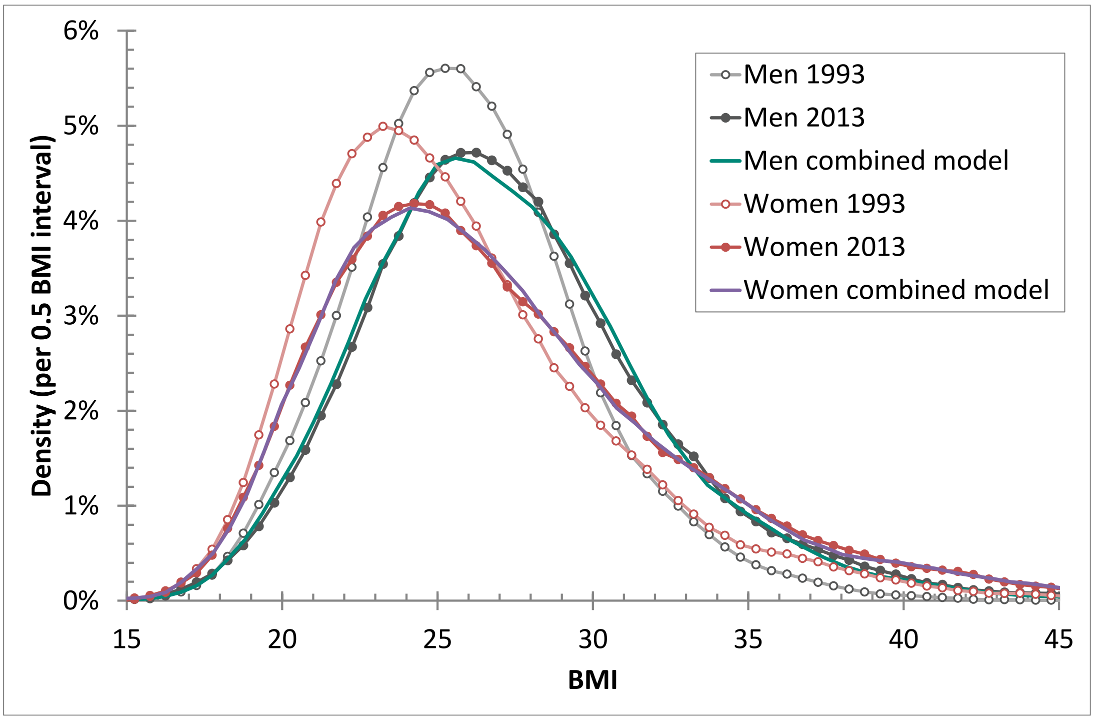


Figure 5.12 Combined scaling models for 1993 to 2013 for men and women.

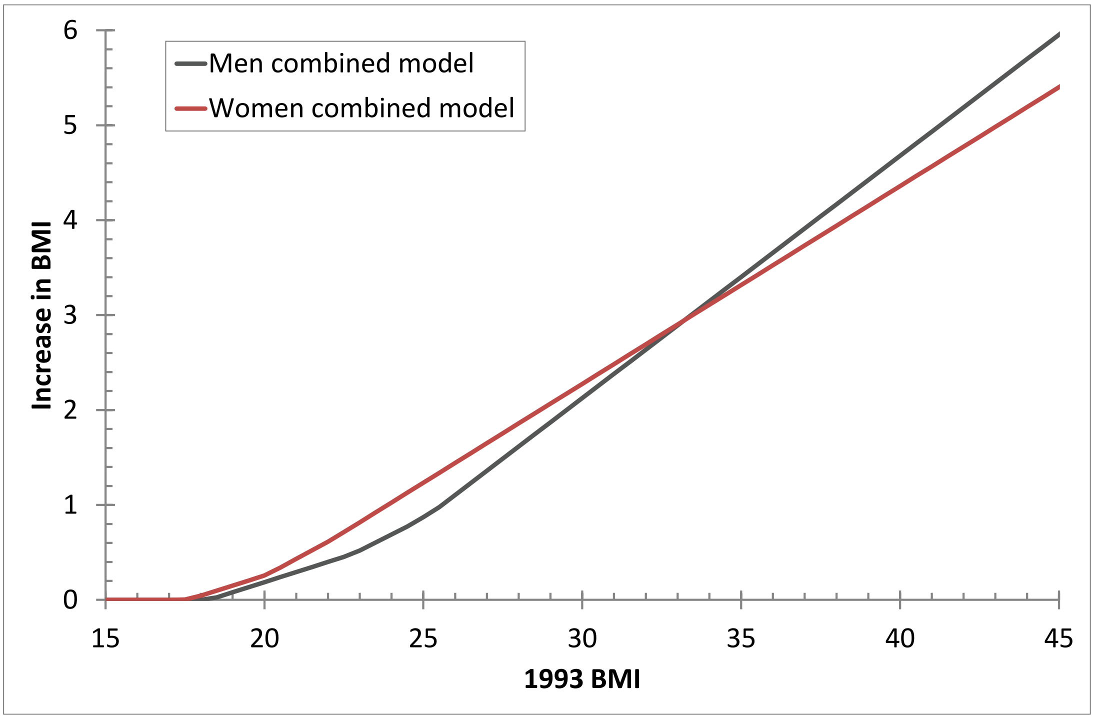


Figure 5.13 BMI increases for the combined scaling models for 1993 to 2013 for men and women.

5.8.2 Single linear scaling model for 1993 to 2013

A linear scaling can be applied from 1993 to 2013 in exactly the same way as the models described in Section 5.4. There are two main differences in this approach to the combined model described in Section 5.8.1. One is that the optimisation process is applied and so the model chosen is the one with lowest fitness value. In other words, the model gets the best match it can to 2013. This is not the case for the combined model which is just a combination of the individual models with no direct reference to the 2013 distribution. This aspect would be expected to result in the single model having a better match to the 2013 actual distribution. The other difference is that the single model is restricted to a linear scaling. The combined model is effectively piecewise linear by using the separate individual models which makes it more flexible. This aspect would be expected to result in the single model having a worse match to the 2013 actual distribution. Hence, the model could conceivably be better or worse depending on the relative importance of these two aspects.

The single scaling model produced for men has a scaling start point of 20.25, a scaling factor of 1.190, and a fitness value of 34.83×10^{-6}. The single scaling model for women has a scaling start point of 19.00, a scaling factor of 1.202, and a fitness value of 7.54×10^{-6}. The models are shown in Figure 5.14. The models do quite well in matching the 2013 distributions, particularly for women, although the fitness values are higher than for the combined models in the previous section. Both models have a reasonably high scaling start point and before that point they just use the 1993 values. This creates a noticeable bump in the curves, particularly for the model for men.

The bump could be much reduced by fixing the scaling start point at, say, 18.50 and just optimising the scaling factor. This option results in scaling factor values of 1.152 for men and 1.189 for women, with fitness values of 50.40×10^{-6} and of 10.32×10^{-6}, respectively. These models still fit quite well, although for men the peak of the scaling model distribution is a bit higher than the actual distribution.

The model for men in Figure 5.14 is not quite able to follow the actual curve for 2013 to the right of the peak of the distribution. This is consistent with the analysis of the percentiles in Figure 4.32 where the values for men are a little different to a linear pattern in having a noticeable curve or approximately a 2-piece linear pattern. Even so, the model does quite well and therefore a linear scaling explains most of the changes in the BMI distribution from 1993 to 2013.

The BMI increases for the single linear scaling models are shown in Figure 5.15, as the dashed lines. The values for the combined models from Figure 5.13 are included for comparison. The percentile increases from Figure 4.33 are also shown. As would be expected, for men and for women the increases are similar for the two models and the percentiles. This is particularly the case for women where the combined model values are close to a straight line and very similar to the linear scaling model increases. As mentioned in Section 5.8.1, the increases for men and for women are also about the same.

As noted in Section 5.8.1, the increases for high BMI values such as those above 40 are uncertain because there are not many values in this range in 1993.

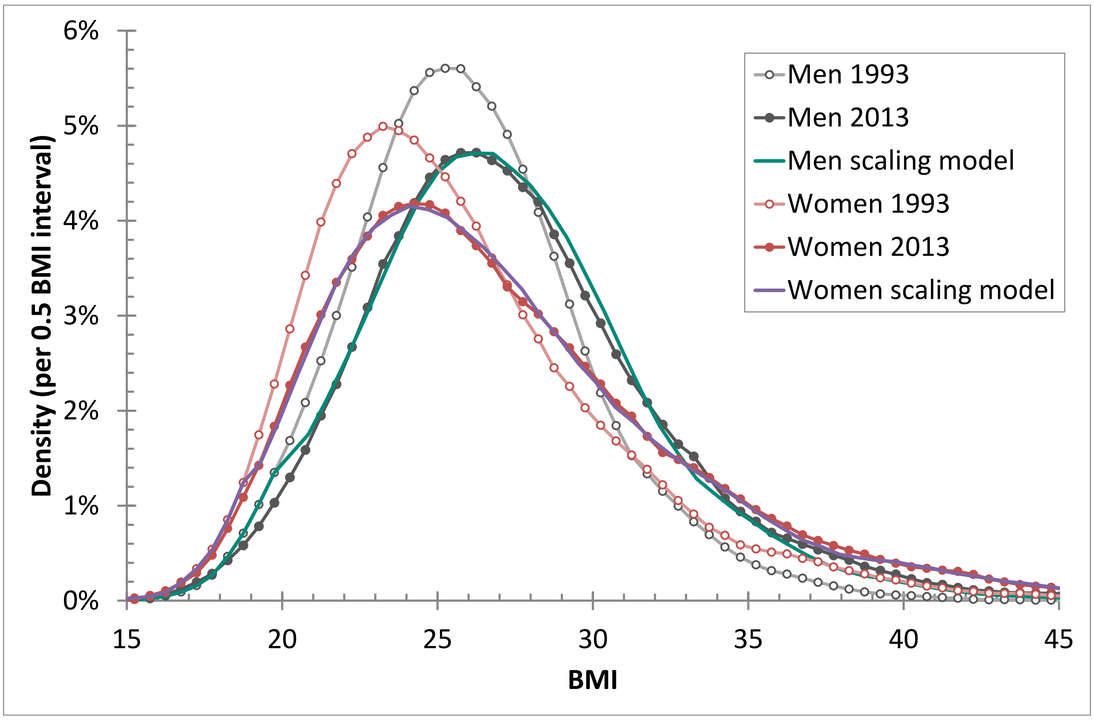


Figure 5.14 Single linear scaling models for 1993 to 2013 for men and women.

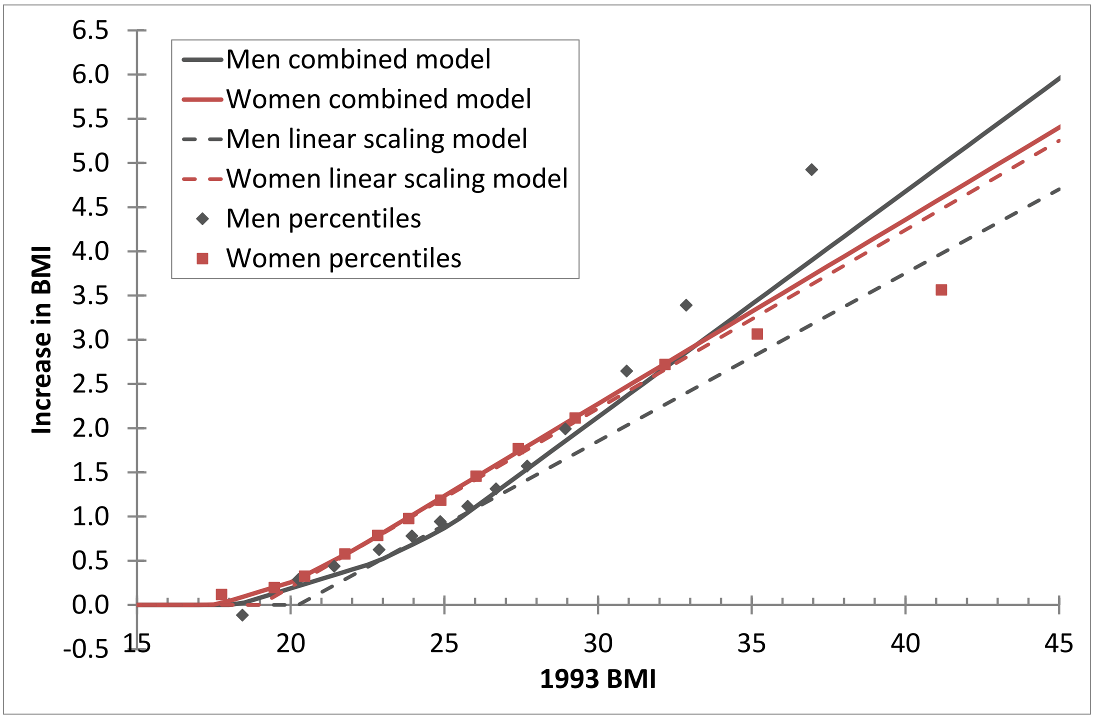


Figure 5.15 BMI increases for the models for 1993 to 2013 for men and women.

5.8.3 Other models

Other related models are possible for this situation. One option is to allow a different pattern in the increase in BMI. A next level of complexity is a 2-piece linear pattern. This will inevitably give at least as good a fitness value as the single linear scaling since it reduces to a linear scaling if the scaling factors for the two pieces are the same. This does give a reasonable improvement in the model for men in that it reduces the fitness value to 13.70×10^{-6}. The model has a scaling factor of 1.168 between BMI values of 19.25 and 26.75 BMI and then a scaling factor of 1.296 above 26.75. The linear model for women is already good and the 2-piece linear model only improves it slightly to a fitness value of 5.04×10^{-6} with a scaling factor of 1.194 between BMI values of 18.75 and 24.75 and then a scaling factor of 1.230 above that. Both scaling factors for women are similar to the single linear scaling factor of 1.202 and therefore it does not change the model much.

Another option that was explored is to use a probability distribution for the BMI increases. In this model, proportions of the population increase in BMI by different amounts following the chosen probability distribution. This is a more detailed model and represents the situation that changes over time will affect people in different ways and by different amounts. From some initial work, the binomial distribution was the best model found in being able to model the changes well and giving a plausible pattern for the changes. It has a feasible interpretation as the binomial trials could represent the applying or not of separate groups of obesity factors. This model was implemented for whole number values of BMI to make the calculations easier. The model applied uses a maximum BMI increase of 10 (i.e., the parameter, n, for the "number of trials" in the binomial is 10). The mean change was modelled in the same way as before as a linear increase, and the binomial probability parameter, p, calculated from this. The model starts with each whole number BMI value and the 1993 proportion for that value (e.g., the proportion for 19 is for the interval $18.5 \leq \text{BMI} < 19.5$). It then uses the binomial with the given mean to determine the percentages of that proportion that increase in BMI by 0, 1, 2, … , 10. So, for a BMI of 19 this gives the percentages that change to BMI values of 19, 20, 21, … , 29. The model parameters used in searching for the lowest fitness value are the values for the linear mean increase: the starting point, the ending point, and the rate of increase. A slight difference here is having an ending point for the linear increase above which the increase stays constant. The model is not very sensitive to this compared to no limit, mainly because there are not many high values in the 1993 distributions.

For men, the binomial model derived has starting and ending BMI points of 18.75 and 35.00, with an increase per unit BMI of 0.157 (i.e., equivalent to a linear scaling factor of 1.157). The fitness value is 16.62×10^{-6} and so a good value. For women, the binomial model has starting and ending points of 18.00 and 34.50, with an increase per unit BMI of 0.174. The fitness value is 4.06×10^{-6} which is even better than the fitness values for the linear and 2-piece linear scaling models. The fitness values are calculated in an equivalent way to the scaling models using density values per 0.5 BMI and multiplying the squared differences by the width of the intervals (which is simply 1 unit of BMI here).

These two ideas can be combined with a binomial model using a 2-piece linear increase in the mean. This gives a particularly good fitness value for men of just 3.57×10^{-6} with the model having a linear increase in the mean of 0.138 per unit BMI between 17.5 and 26.75, then an increase of 0.263 up to BMI of 40. For women the model has a linear increase of 0.165 between 17.5 and 24.5 and then an increase of 0.201 above that with no upper limit. The fitness value is very low at 2.72×10^{-6}.

More details on the results from these models are set out in Appendix 5.V. This includes charts of the model distributions, the BMI increases, and examples of the binomial distributions. Each of these model types could also be applied to the shorter four year time intervals covered earlier in this chapter. However, the changes over the shorter time intervals are smaller and the linear scaling models already do well. Using a more complex model can, to an extent, be overfitting the data. Therefore, it was not considered worthwhile to apply these models to the shorter time intervals. The BMI increases from 1993 to 2013 for all the models are shown in Appendix 5.V in Figure 5.V.6 for men and Figure 5.V.7 for women. In each case, the increases are similar for all models particularly in the BMI range of 20 to 35 where most of the 1993 values are.

Overall, therefore, these alternative models all give good results and the increases in BMI are similar to those in Figure 5.15 for the combined model, linear scaling model and percentiles.

Chapter 6: Obesity factors and interpretation of the changes in the BMI distributions

Key points from this chapter:

- The changes in the population BMI distributions over time can be interpreted as the combined effect of how BMI is different in equivalent people in the two populations. For example, suppose we are comparing the 1993 and 1997 distributions. In theory, people in the 1997 population could be matched with people in the 1993 population with the most similar characteristics (the same age and the closest characteristics for other demographic variables). The difference in the two BMI distributions then represents the combined total of how the BMI of each person in 1997 differs from that of the matched person in 1993.
- The difference in the population at one year compared to another year is that the people have lived in different times. For example, someone of age 35 in 1993 has lived in the time period 1958 to 1993, whereas someone of age 35 in 1997 has lived in 1962 to 1997. The different time periods result in altered lifestyles and people having experience of a different range of conditions over their lives. Obviously, this applies in a similar way to each age: for example, still comparing the 1993 and 1997 distributions, a person of age 20 has lived in 1973 to 1993 compared to 1977 to 1997, a person of 50 has lived in 1943 to 1993 compared to 1947 to 1997, etc. The difference in BMI over a population is the combined effect of the changes in experiences and lifestyles for the different time periods across the whole population over the full range of ages.
- Obviously, the difference in time periods is greater when considering distributions further apart such as 1993 and 2013. In that case, people of age 20 have lived in 1973 to 1993 compared to 1993 to 2013. Those of age 35 have lived in 1958 to 1993 compared to 1978 to 2013. Those of age 50 have lived in 1943 to 1993 compared to 1963 to 2013. This will cause considerable differences in lifestyles and experiences.
- Since changes in BMI are interpreted as due to living in a different time, the assumption is that the 2013 population would have had the 1993 BMI distribution if they had all lived 20 years earlier.
- There are many lifestyle and environment factors that change over time and may have affected BMI. Some of these are discussed in this chapter and a suggested list is provided including factors to do with economics, the food industry, work, leisure, transport, and society.
- The results from Chapters 2-5 are that the increases in BMI are greater for those with a higher BMI in an approximately linear pattern. The important implication is therefore that changes in the obesity factors and the changes in lifestyles through living in different times have affected those with a higher BMI much more than those with a lower BMI.

6.1 Overview

The analysis presented so far in the previous chapters has derived population BMI distributions for England for 1993, 1997, 2001, 2005, 2009, and 2013. The differences between the distributions have been examined in detail and have been modelled as linear scaling transformations. This chapter discusses how to interpret the distributions and, particularly, the changes between them. This requires considering what the distributions represent, and the years and conditions experiences during the lives of the people in the populations.

6.2 Interpretation of the changes over time in the BMI distributions

The BMI distributions in Chapter 4 give the estimated proportions of different BMI values across the whole population of England. The data used is cross-sectional as it is a different random sample of the population each year, as opposed to being longitudinal where the same individuals are followed over time.

We need to consider how best to interpret the changes in the population BMI distributions over time in order to try and understand the reasons for the changes. These are changes in BMI for the whole population. One important point is that a constant age distribution was used in deriving the BMI distributions (Section 3.3) and so the population age profile is always the same. Therefore, age is not a factor in the changes in the population BMI distributions.

In particular, the changes do not represent how much BMI increased on average for individuals over time as they got older. BMI does tend to increase with age (as discussed in Chapters 7-10 in this report) but this is not what the changes in the distributions are due to. Comparing two populations four years apart, new people will enter the later distribution by reaching the age of 18, and some people will leave the later distribution as they have died between the two time points. Everyone who is in both populations will have increased in age by four years. However, the proportion of each age is exactly the same in both populations because of the constant age profile.

Instead, what the changes represent is how BMI has changed in equivalent people. For example, how BMI is different for those aged 18 in the later population compared to those aged 18 in the earlier population. Similarly for each age. The overall change is then the combined effect across all ages.

Theoretically, the people in the two populations could be matched up as closely as possible by other demographic and individual characteristics as well as age. If the mix of population demographics change over time (such as the proportions of people with different educational levels or in different job types) then the theoretical matching would need be on some sort of ranking basis. The idea would be to match people who would have had the same characteristics, such as job, or educational level, if they had been born in the same year. Similarly, they would have lived in the same area, had the same income, etc. However, due to living in different times, their life experiences are different in various ways. This is potentially a combination of different personal circumstances and also experiencing different conditions in society such as changes in the food environment (which could be a variety of changes in food production, retailing, restaurants, etc.).

To give an example, suppose we compare the population distributions for 1993 and 1997. A woman or a man of a particular age and characteristics might have a BMI of, say, 26.2 in 1993. They might be aged 35, for example. Theoretically, we could find the nearest equivalent person in 1997 who is aged 35 and has the same or very similar characteristics. This equivalent person might have a BMI of, say, 26.7. The main reason for the difference in BMI is that the person has lived in a different time period. So, the first person has lived in the time period of 1958 to 1993 for the 35 years of their life. The second person has lived in the time period of 1962 to 1997 for their 35 years. The lifestyles of the two people will be different because of living in the different time periods and hence their BMI is different. The lifestyle factors that might cause the change in BMI are considered in the next section.

The change in the BMI distributions is the combined effect of the different time periods across the whole population. The time periods are offset by the same amount but the specific dates vary by the age of the person. So, comparing the 1993 and 1997 distributions, a person of age 20 has lived in 1973 to 1993 compared to 1977 to 1997. A person of 50 has lived in 1943 to 1993 compared to 1947 to 1997. The same applies to every age from 18 upwards.

Obviously, the difference in time periods is greater when considering distributions further apart such as 1993 and 2013. In that case, people of age 20 have lived in 1973 to 1993 compared to 1993 to 2013. Those of age 35 have lived in 1958 to 1993 compared to 1978 to 2013. Those of age 50 have lived in 1943 to 1993 compared to 1963 to 2013. This will cause considerable differences in lifestyles.

As already discussed, suppose we compare someone in the population for a given year with an equivalent person from an earlier year. The equivalent person might have a different type of job, have a different level of income, eat a different mix of food, live in a different place, do different leisure activities, etc., just from living in a different time period. If the time difference is small then many of the factors will be the same for a lot of the population. Comparing distributions four years apart then an equivalent person may well live in the same area, and have the same type of job and similar income, for example. However, they will still be affected by broader societal and environment changes.

There are obviously wide variations in BMI between different people in the population as can be seen in the distributions in Chapter 4. The reasons for the differences within the population at an individual level are not considered in this report. However, height, weight, and body type and hence BMI will be dependent to some extent on the genes and the natural physiology of the person – for example, Naukkarinen et al. (2012) discuss genetic and lifestyle effects on obesity from studies of twins in Finland. However, at a population level it is generally assumed that overall the population stays basically the same over time in physiology and genetic make-up (Section 6.3.1). Hence, across the population at different times it is assumed that there is the same mix of natural innate tendencies to have high or low BMI. The difference over time at a population level is then due to the different conditions experienced.

In the analysis in this report, an important part of the comparison of distributions over time is to look at percentiles. This is a simple way to match up the two populations. It is assumed that a given percentile represents equivalent people (or an equivalent mix of people with that level of BMI) who would therefore have had the same BMI if they had lived at the same time. Hence, the change in BMI for a given percentile represents the effects of the different time periods on that level of BMI. The results for the percentile changes over time in Sections 4.6 and 4.8.3 are that the BMI increases are higher for higher BMI levels in an approximately linear relationship. Hence, the effect of living in later times and experiencing different conditions has had a much greater effect on those with a high BMI than on those with a low BMI.

Overall, the general assumption is that the 2013 population would have had the 1993 BMI distribution if they had all lived 20 years earlier. The interpretation suggested for the changes in the BMI distributions is that it is due to different lifestyles because of living in different times.

6.3 Lifestyle factors for the different time periods

As explained in the previous section, the changes in BMI are interpreted as being due to the population living in a different time period and hence having a different lifestyle experience over the course of their lives. This section considers and discusses some of the possible factors. Some references are made to reports and academic papers, although this is not intended as a comprehensive literature review.

6.3.1 General discussion of obesity factors in other reports

Foresight project

A major large scale obesity study that was published in 2007 is the Foresight Tackling Obesities project. There are a number of reports from the project and these are available from the links on the Foresight web page on the U.K. Government website [21]. A comment about the project on the web page is: “The project involved over 300 experts from a wide range of disciplines and was overseen by a high level stakeholder group. The project was sponsored by the Department of Health”. The main report (Butland et al., 2007) explains that this was an independent project that was commissioned by Sir David King, Chief Scientific Adviser to the U.K. Government, and Head of the Government Office for Science.

The executive summary on page 5 of the main report says “People in the UK today don’t have less willpower and are not more gluttonous than previous generations. Nor is their biology significantly different to that of their forefathers. Society, however, has radically altered over the past five decades, with major changes in work patterns, transport, food production and food sales. These changes have exposed an underlying biological tendency, possessed by many people, to both put on weight and retain it.” This is consistent with the comments above in Section 6.2 that changes in the BMI distributions are a consequence of people living in different times and experiencing different conditions.

The causes of obesity are discussed in section 3 of the main Foresight report. At the start of that section on page 43 the report says: “Although there are many reasons why an individual may become obese, it is now generally accepted by health and other professionals that the current prevalence of obesity in the UK population is primarily caused by people’s latent biological susceptibility interacting with a changing environment that includes more sedentary lifestyles and increased dietary abundance.” Again, this indicates that increases in obesity levels are due to conditions changing over time. Section 3 discusses five main categories of causal factors: the natural biology of the human body; the effect of the first few years of life; behavioural aspects of eating and physical activity including psychology; the environment and living conditions including the effects on exercise and the extent of sources of food; economic impacts on the amount and type of food and drink consumed including the effects of prices and marketing. The report makes a number of comments that the importance of the various factors and their relationships are not understood well.

The report argues that increasing obesity levels are due to interactions between human biology, changes in society, and human behaviour. A comment on page 46 is: “Food is a fundamental biological necessity and the body has evolved to make sure that its needs are met. The hunger drive is very powerful and compels humans to search out food. By contrast, there is limited sensitivity to abundance.” In other words, human biology developed through evolution has strong responses for hunger and shortage of food but is not adapted so well for situations where food is plentiful, which has occurred in recent years. Hence, modern lifestyles are behind the rise in obesity in creating an environment that tends to lead to weight gain. This is often referred to as an “obesogenic environment”. Some of the changes in lifestyle considered in section 3 of the report are: the trend of technology reducing physical tasks; environmental effects on levels of physical activity and on how easy it is to obtain food; economic effects such as cheaper food prices, lower priced foods perhaps tending to be less healthy, and longer working hours. The report also discusses other elements including the human biological mechanisms and the psychology of human decision making including motivation. It also mentions a “generational dimension” (page 8) with childhood obesity being related to parental obesity.

[21] https://www.gov.uk/government/collections/tackling-obesities-future-choices

One of the component pieces of work in the Foresight project was to compile a list of variables that affect obesity and link them in a system map and submaps. This is described in section 5 of the main report (Butland et al., 2007). The report presenting the collection of the maps is called the "obesity system atlas" (Vandenbroeck et al., 2007a). There is also a report on the mapping process and the results (Vandenbroeck et al., 2007b). A total of 108 variables were identified and included on the maps, and these are listed at the end of the system atlas (Vandenbroeck et al., 2007a). The focus was on factors that affect the situation at the current time of the report. However, the variables could also be reviewed to identify how each one changed over a particular time interval which would produce a detailed list of lifestyle changes.

The maps are causal loop diagrams, which is a technique that was devised originally by Maruyama (1963). Causal loop diagrams are a tool for helping to visualise and understand the interactions in a complex system. They are sometimes used in the process of developing quantitative system dynamics models, as discussed for example in chapter 7 of Pidd (2009). However, this is not the usage in this case where the focus is on the diagram itself to show the system structure and the qualitative relationships.

A causal loop diagram shows the factors or variables in the system and the causal relationships between them that are considered to exist. The relationships have a direction and can produce reinforcing or balancing loops that can indicate some of the potential dynamics in the system. Reinforcing loops generally mean that the value or level of each variable will either increase or decrease continuously, whereas balancing loops tend to control the system and put limits on the increases or decreases in the values of variables. Loops can interact with each other and so the overall dynamics of the system can be very complex.

There are various maps, but the main obesity system map is the "Full Generic Map" and is Figure 5.2 in main report (Butland et al., 2007) and map 5 in the system atlas (Vandenbroeck et al., 2007a). The map has a central area with six interacting variables representing "energy balance". This combines loops for the natural biological impulses for eating food to balance the energy used, habits of behaviour that might develop, and deliberate decision making regarding eating choices. The other factors that surround and connect into the central area are divided into seven clusters. Four of these are next to the central area and are mainly about the nature, behaviours, and actions of the individual or group being considered: "physiology", "individual physical activity", "individual psychology", "food consumption". The other three clusters are at the outer edge of the map and refer to the external environment: "physical activity environment", "social psychology", "food production". Each one of these particularly affects and connects to the corresponding related individual cluster (the last three of the four individual clusters listed above).

The food consumption cluster is perhaps the most obvious in terms of factors that have changed over time and are likely to have a clear direct influence on obesity. This has 16 variables that include aspects to do with the availability of food, the nature of the food eaten in terms of both quantity and composition, the amount of alcoholic drinks in the diet, and purchasing, eating and cooking habits. Hence, it is largely about individual behaviours and decisions, but these will be heavily influenced by the external environmental situation, particularly variables from the food production cluster. Four "key variables" were identified in the map (Butland et al., 2007 section 5.2.1; Vandenbroeck et al., 2007b section 3.5) that are the main links from the individual clusters to the central area. The one for food consumption is called "force of dietary habits". One of the variables included in the related food production cluster is "purchasing power", which depends on some economic factors and on food price levels. The changes in the affordability of food over time are discussed further in Section 6.3.3.

The obesity system map highlights the complexity of the obesity issue and the wide range of factors that affect it and interact with each other. It is likely to require many different approaches to tackle different aspects of the system over a long timescale in order to make significant changes to the overall outcome. The obesity system map might be used to indicate good areas where initiatives can be directed and also how they might affect other parts of the system (Butland et al., 2007 section 5.2.2).

Scientific Advisory Committee on Nutrition (SACN) report

The Scientific Advisory Committee on Nutrition (SACN) [22] is an advisory group to the U.K. Government. The 2020 annual report for SACN explains their role as follows (page 3 of the report [23]): "The role of SACN is to provide scientific advice on, and risk assessment of, nutrition and related health issues. It advises the four UK health departments and other government departments and agencies. Members are appointed as independent scientific experts on the basis of their specific skills and knowledge. The committee also includes 2 lay members."

In 2011, SACN looked at energy requirements in diet. The report on this is called "Dietary Reference Values for Energy" and was published in 2012 (SACN, 2012 [24]). Some of the results from the report are used in the analysis on calories in Chapter 13 of this report. The work was done in the context of increasing rates of obesity. The aim was to derive reference values, and for adults these are based on a BMI of 22.5. The idea is that following the recommended energy amounts will tend to result in moving towards a BMI of 22.5 whether a person is underweight or overweight (report preface on page 1).

Regarding the causes of obesity, the report has a comment in paragraph 50 on page 33: "Obesity results from a long-term positive energy imbalance. The increasing prevalence of obesity must reflect temporal lifestyle changes, since genetic susceptibility remains stable over many generations, although inter-individual differences in susceptibility to obesity may have genetic determinants (Maes et al., 1997)." Again, this is saying that, at a population level, changes in obesity levels are due to alterations in lifestyles over time. The Maes et al. (1997) paper is a literature review looking at how much the variation in BMI levels is due to genetic or environmental effects from results in twins, family, and adoption studies.

Wardle and Boniface (2008) paper

A paper by Wardle and Boniface (2008), that analyses changes in percentile values over time from the HSE data, is described in Section 4.8.4 in Chapter 4 of this report. In their results and also found in the work in this report, higher percentiles at higher BMI values tend to have much greater increases in BMI over time. In discussing their results, they make similar points to those above in this section that changes in BMI over time are due to experiencing and living in different environments. As also mentioned in the SACN report quote in the previous paragraph, Wardle and Boniface (2008) comment that the difference in levels of weight and hence BMI in the population is likely to have genetic causes to some extent. There would then be an interaction between genes and the environment, since the percentile results indicate that those in the population with higher BMI tend to have been affected more by the environmental changes over time.

[22] https://www.gov.uk/government/groups/scientific-advisory-committee-on-nutrition
[23] https://www.gov.uk/government/publications/sacn-annual-report-2020
[24] https://www.gov.uk/government/publications/sacn-dietary-reference-values-for-energy

6.3.2 Timescale experienced by people in the HSE

We can consider the timescale covered by the distributions from the HSE data. The earliest distribution produced is for the 1993 population, derived from the 1991-1994 data. The latest is the distribution for the 2017 population, using the 2015-2018 data, and the results for this are in Chapter 14. People in their 70s and 80s in the 1993 population grew up in the 1910s and 1920s. People in their teens or early 20s in the 2017 distribution grew up in the 1990s and early 2000s. All the people in the 2017 distribution have obviously experienced current modern lifestyles in the 2010s as their most recent years of life. Hence, the distributions reflect experiences over the past 100 years and this will be a great variety of conditions. Clearly, there have been enormous changes over that time in many areas of life including food, work, leisure, transport, the economy, and home life.

This report is considering BMI at a population level and how this has changed over time, rather than differences between people within the population. Hence, the interest is in the general changes in factors over time and their overall effects on the population.

Obesity can also be considered at an individual level and differences in the factors could be examined to try and explain differences between individuals. Clearly, at an individual level personal circumstances will vary extremely widely in many different ways. The obesity system in the Foresight obesity map discussed in Section 6.3.1 is described as "operating 'around' an individual or a group of people" (Vandenbroeck et al., 2007b, page 3) and so the map can potentially be applied to either individuals or to small groups or to larger populations. Each of the 108 variables in the map could vary in value considerably from one person to another. However, the discussion here is concerned with the reasons for changes in the population distribution of BMI and so it is the general or average effects on the population that are of most relevance and that are considered here.

The HSE data used in this report goes up to 2018. Therefore, the discussion here is focussed on potential changes in obesity factors in the time period to 2018. Since then, there has been the Covid-19 pandemic and also economic changes with a high inflation rate in 2022 reflecting significant increases in many costs for households. The time after 2018 is outside the scope of this report but these factors could have various effects on BMI and obesity levels in the population. It will be interesting to see exactly how the BMI distributions change in the future as new data becomes available.

The populations in the data considered in this report between them have life times that include more than 100 years up to 2018. Therefore, changes in circumstances over that time are relevant in considering factors that have affected obesity levels in these populations. Sections 6.3.3 – 6.3.5 consider some of these factors.

6.3.3 Economic factors

It is difficult to say which population lifestyle changes have been the greatest over the past 100 years in terms of their impact on obesity. Just as an example of one general area, this section considers various economic factors including income levels and food prices, and the average economic prosperity of the population. This topic is not necessarily the most important obesity factor, but it is considered in more detail here because it is one where there is quite a lot of data available. There will be many other aspects of lifestyles and society (work, leisure, transport, housing, etc.) that have changed a great deal over the past 100 years and these may have had a greater effect on obesity. There will also be food industry factors, although the economics of food production and retailing will be part of this and some aspects of this are mentioned in this section.

Proportion of household expenditure spent on food

The main Foresight report (Butland et al., 2007) described in Section 6.3.1 has a section on economic factors called "Economic drivers of food and drink consumption" (Section 3.5 on pages 55-58). One aspect that they look at is the proportion of household spending that is on food. Figure 3.2 on page 55 of the report is a chart of this from 1963 to 2002. The y-axis is labelled as the "percentage of consumer expenditure on food", and the source cited is the Office for National Statistics. On the chart, the line of the values for the percentage spent on food reduces approximately linearly from about 23% for 1963 to about 8% for 2002. This is a very large reduction. The report comments that this is an "indicator of economic progress" and also that it will tend to be beneficial for health at the population level. The report says that "In developed economies … there are now plentiful sources of relatively cheap foods". It also points out that there will be variations across the population, and it states that the percentage of income spent on food is greater than 23% for "lower-income households" (page 55 of the report).

Section 3.5 of the main Foresight report also considers some food industry economic factors. This includes the costs of different types of food with comments that cheaper foods tend to be high in calories but low in nutrients. There is also discussion of the development by the food industry of a wider variety of processed foods, with manufacturing by mass production enabling the products to be made cheaply and hence they become more widely accessible. A different economic factor mentioned on page 58 is longer working hours tending to be associated with greater prevalence of obesity.

For more recent data on the relative amount that households spent on food, the Department for Environment Food & Rural Affairs publishes a Food Statistics Pocketbook on their website [25]. The section on prices and expenditure shows data on the share of household spend on food and non-alcoholic beverages in Figure 2.2 on the web page for the report [26] for all U.K. households and for the "lowest 20% by equivalised income". The spreadsheet that is available from the web page in the link below the chart has data from 2010 to 2020/21. There are earlier versions of the pocketbook [27], and the 2014 version has a spreadsheet available with data going back to 2003/04. Combining the data, the values do not change much between 2003/04 and 2019/20 being between 10.2% and 11.6% for all U.K. households and between 14.3% and 16.8% for the lowest 20% by income. This report is concerned with HSE data up to 2018 and the values for 2018/19 are 10.5% and 14.7% respectively.

The source of the data is given on the web page in footnote 26 as the "Living Costs and Food Survey (Defra/ONS)" with a link given on the webpage to the data spreadsheet for the latest year. The calculation of the percentage is the total expenditure at home of food and non-alcoholic drinks divided by total expenditure. It does not include alcoholic drinks or restaurant and hotel expenditure. The percentage values for the last year of data for 2020/21 are higher being 14.4% for all households and 18.3% for the lowest 20% by income. However, section 2.2 of the web page explains that :"In 2020/21, with COVID-19 restrictions imposed for some of the year and households not spending as much on eating out and recreation, the proportion of expenditure spent on food to eat at home increased."

[25] https://www.gov.uk/government/statistics/food-statistics-pocketbook
[26] https://www.gov.uk/government/statistics/food-statistics-pocketbook/food-statistics-in-your-pocket
[27] https://www.gov.uk/government/collections/food-statistics-pocketbook

The Our World in Data website, founded by Oxford University economist Max Roser, has a web page on “Food Prices” by Max Roser and Hannah Ritchie giving a variety of charts and data [28] (Roser and Ritchie, 2021). This includes a chart for the U.S.A on the proportion of expenditure on food from 1929 to 2014 (web page section entitled “Share of expenditure on food” with subsection title “Food consumed away from home”, chart title “Food expenditure as a share of family disposable income, United States, 1929 to 2014”), with the source being the United States Department of Agriculture (USDA) Economic Research Service [29]. This chart has a similar pattern to the U.K. data in the Foresight report discussed above with considerable decreases in the last 50 years. The chart has peaks in the data in 1933 and 1947 with a dip between those dates. There are then considerable decreases after 1947. The percentage value of expenditure on food as a share of income for 1947 is 23% and then this reduces rapidly to a value of 13% for 1968. The reductions since then are at a slower pace down to 10% in 2000 with not much change after that. The chart divides the expenditure between food consumed at home and food consumed away from home. The percentage values as a share of expenditure for food consumed away from home change very little, always being between 3.0% and 4.3%, and it is the percentage values for food consumed at home that reduce over time. These are percentages of the total expenditure. The next subsection looks at income for the U.K., which has increased considerably over time in real terms. This is presumably similar for the United States and total expenditure is likely to increase in a similar way to total income. Hence a constant percentage of expenditure for food away from home would mean increases in the actual real amounts spent. This is also shown in the subsection on the next page on food expenditure and food prices.

Based on the references in this subsection, one big economic change in the last 60 years has been large reductions in the average proportion of expenditure spent on food indicating that on average for the population food is much more affordable.

Economic prosperity

Regarding general economic prosperity at the population level, long term changes are discussed by Max Roser on the Our World in Data web page on Economic Growth [30] in a section headed “From poverty to prosperity: The UK over the long run” (Roser, 2013). The section includes a chart of gross domestic product (GDP) per capita in real terms (at 2013 prices) from a source data spreadsheet published by the Bank of England (Thomas and Dimsdale, 2017) with the original source of some of the data being Broadberry et al. (2015) (presumably from the title of this reference, this source provided the data from 1270 to 1870). The text in the section on the web page indicates that GDP per capita is a measure of average income. The data from 1900 to 2016 from the spreadsheet is shown overleaf in Figure 6.1.

The web page text by Roser (2013) has the comment that living conditions changed little for a thousand years up until the 17th Century. The average of the GDP per capita values from the spreadsheet from 1270 to 1650, as quoted on the web page, is £1051, with values varying between £673 and £1391. The spreadsheet value for the year 1659 is £974, but from then on the GDP values increase rapidly, with some fluctuations, and the web page text notes that the U.K. was the first place to obtain such increases. As shown in Figure 6.1, the values for GDP per capita reach about £5000 in 1900. Between 1900 and 1930 there is little change but after that there are further considerable increases. From Figure 6.1, over the 40 years from 1976 to 2016 the value approximately doubles from about £14000 to about £29000. Therefore, current values of GDP per capita are about six times the values from 100 years ago. This is relevant because, as discussed in Section 6.3.2, some people in the HSE data grew up about 100 years ago. As seen in the data in the previous subsection, such increases in prosperity are likely to make food considerably more affordable on average for the population.

[28] https://ourworldindata.org/food-prices
[29] https://www.ers.usda.gov/data-products/food-expenditures.aspx
[30] https://ourworldindata.org/economic-growth

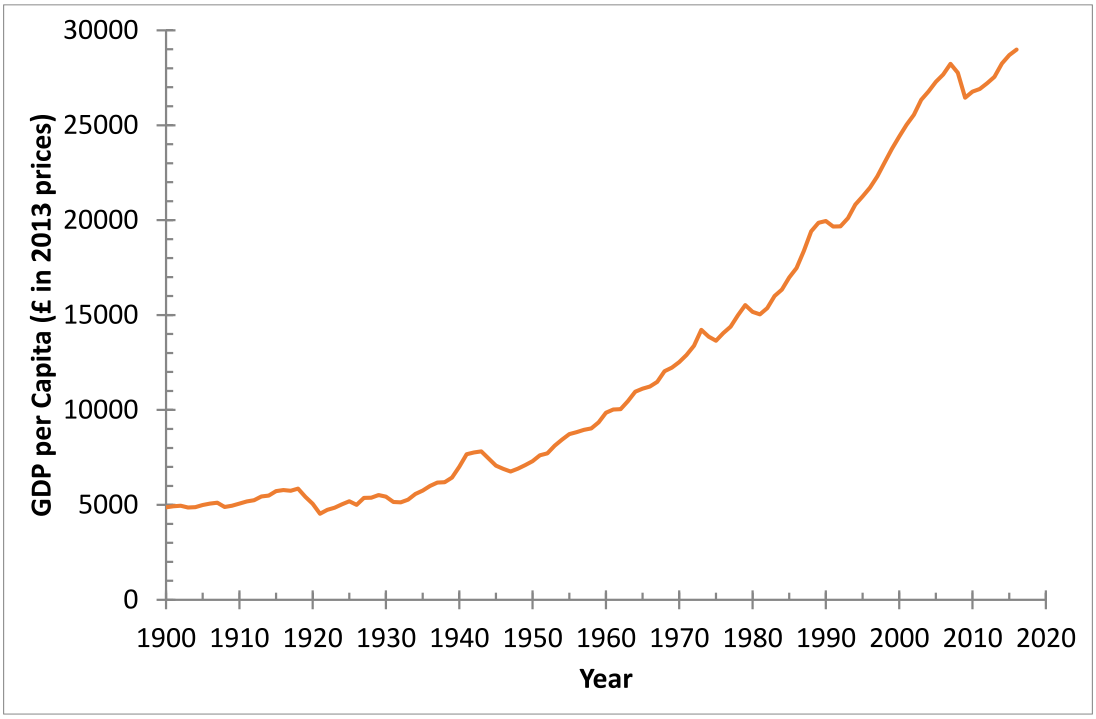


Figure 6.1 GDP per capita for the U.K. since 1900 at 2013 prices (Thomas and Dimsdale, 2017). [Data from the source spreadsheet published by the Bank of England (Thomas and Dimsdale, 2017)].

Food expenditure, prices, and quantity

There is then a question of how much change there has been in absolute terms in the expenditure on food and in the quantity of food purchased. Another chart on the Our World in Data web page for Food Prices [31] by Roser and Ritchie (2021) is food expenditure per person for the U.S.A. in 1988 dollar values from 1953 to 2014. The link for this chart is in the section "Share of expenditure on food" and is the text "Food expenditure in person" immediately below the chart on the proportion of expenditure on food discussed earlier in the first subsection of Section 6.3.3. The data source again is the United States Department of Agriculture (USDA) Economic Research Service [32].

The chart shows food expenditure at home staying roughly constant at about $1100, whereas the expenditure on food away from home increases from about $500 to about $1100. Hence overall expenditure increases from about $1600 to about $2200. The specific values for 1953 are $522 and $1129 giving a total of $1651, and for 2014 are $1125 and $1126 giving a total of $2251. This is an increase of 36%. The category of food away from home is described as "restaurants, cafes, colleges, work etc.". The increase in expenditure on this probably reflects eating at such places more often, although it could also be partly due to eating at more expensive places or due to price increases in real terms. It is difficult to be sure of exactly what this means for the amount of food purchased. The increases in expenditure could obviously be due to buying and consuming a greater quantity of food in total. However, there could be other effects depending on how food prices have changed in real terms and whether the mix of food items has changed, with perhaps more expensive foods being purchased. For example, a lower quantity of food might be eaten at home but with higher prices giving the same overall cost, along with a higher quantity of food away from home.

The Our World in Data web page on Food Prices (Roser and Ritchie, 2021) has several other charts. This includes another way of looking at food affordability by how much can be bought with typical wages. In the web page section entitled "Food prices compared to wages" there are two charts next to each other. On the web page, the chart on the right is a chart of seven food items that shows how much

[31] https://ourworldindata.org/food-prices
[32] https://www.ers.usda.gov/data-products/food-expenditures.aspx

can be bought for one hour working in the manufacturing sector from 1901 to 2003, and the chart on the left shows the actual prices for these items from 1901 to 2003. The calculations for the charts were done by Roser and Ritchie (2021) using U.S.A. data from a report by the U.S. Department of Labor and the U.S. Bureau of Labor Statistics (USDoL, 2006). The note to the charts also says that the idea for the charts came from an online article on The Atlantic, that discusses wages and food costs over the past 100 years and how manufacturing wages have increased much more than food prices (Thompson, 2012). The data from these two charts was used to do some calculations, as given in the next paragraph.

Using the data from the chart of the amount that can be bought for one hour of manufacturing wages, I did some calculations of the amount that could be bought relative to 1950. The time points from that year onwards in the data are 1950, 1960, 1970, 1984, 1996, and 2003. There are generally considerable increases from 1950 to 1970 and from 1984 to 1996, with not much change between 1970 and 1984 and between 1996 and 2003. For the seven items, much more could be bought in 2003 compared to 1950 by a multiple of between 1.9 and 4.7 (for one item, milk, data is not given for 2003 and so the 1996 value was used instead). This is data for the U.S.A. but the main patterns for the U.K. are presumably likely to be fairly similar. For the data from chart on actual prices, the food prices increase between 1950 and 2003 by factors of between 2.1 and 5.0, whereas the wages value increases by a factor of 9.6. These comparisons are limited by only considering a few food items and one sector for wages. However, the pattern is generally of wages rising much faster than food costs again reflecting increasing prosperity and food items becoming cheaper relative to income.

Regarding the quantity of food, the Our World in Data website has a web page called "Food Supply" [33] (Roser, Ritchie, and Rosado, 2013). This has a chart showing the daily supply of calories per day from 1961 to 2019 for various continents, regions, and countries (in the first web page section called "Caloric supply by region" with the chart title "Per capita kilocalorie supply from all foods per day, 1961 to 2019"). The data source is the United Nations Food and Agricultural Organization (FAO) [34]. The chart is interactive and specific countries or regions can be selected. The U.K. data was selected, and there look to be several step changes in the values. From the chart data, the average U.K. values for the time periods with roughly constant values are: 1961-1973 3240 kcal / day, 1974-1984 3136 kcal / day, 1985-1996 3238 kcal / day, 1997-2019 3408 kcal / day. Hence, there is in increase from 1974-1984 to 1985-1996 of 102 kcal / day, and then a further increase to 1997-2019 of 170 kcal / day. The comments on the website point out that this is the food available at retail and that actual food intake will be lower than this because of food that is wasted. However, if the amount of waste is roughly constant then this would indicate increases in food intake in recent years compared to 1974-1984. As discussed in Chapter 13 of this report that looks at calories and BMI, fairly small changes in average population calorie intake of this sort of magnitude may lead to quite large changes in population BMI. This includes the calories calculated for the changes in population BMI from 1993 to 2013 in Section 13.5.1, and from 1993 to 2017 in Section 14.9.

An article in Obesity Reviews (Lee et al., 2013), by members of the INFORMAS group (the International Network for Food and Obesity/non-communicable diseases Research, Monitoring and Action Support), considers the affordability of different diets, particularly comparing the prices of healthy and less healthy diets. However, they find that there are various difficulties in doing this, which include a lack of a clear definition of which foods are healthy and unhealthy, and the choices of methods used for comparing the foods fairly. From their review of the literature they say that "it is not clear whether 'healthy' foods and diets are generally more expensive than 'less healthy' foods and diets on the basis of price per calorie" (page 83). They mention the use of taxes and subsidies which might affect the relative demand of different food items although the effects are unclear (as also discussed in Frew et al., 2018). They also consider the effects of differences in affordability between countries and within countries. The relative amount spent on food was considered at the start of this section and Lee et al., (2013) comment that food expenditure can be as high as 80% of income for the poorest groups in low income countries. This is obviously very different to the figures above for the U.K. and U.S.A. Within high income countries there will be large differences and they say that "in Australia, a 'healthy' diet

[33] https://ourworldindata.org/food-supply
[34] https://www.fao.org/faostat/en/#data/FBS, https://www.fao.org/faostat/en/#data/FBSH

costs between 28% and 40% of the disposable income of a welfare-dependent family compared with 20% for families on the average income". The paper proposes a framework to collect and analyse suitable data to monitor the affordability of different diets.

Economic factors summary and conclusions

Based on the data discussed in this section, the main economic patterns over the past 100 years at a population level for England appear to be substantial increases in income resulting in the proportion of expenditure on food reducing considerably. This economic relationship that the proportion of food expenditure reduces as income increases is well established and is known as Engel's Law, dating back to a paper in German by Ernst Engel in 1857 (Engel, 1857 as discussed for example in a 75th anniversary paper by Zimmerman, 1932). A paper by Houthakker (1957) to mark the centenary of Engel's paper found that data from 40 surveys in 30 countries followed Engel's Law. In other words, the income elasticity of demand for food is less than 1 and so, as income increases, expenditure will increase but at a lower rate than income.

In the INFORMAS paper, Lee et al. (2013) comment that "In high-income countries, greater total spending on food tends to be associated with more nutritious dietary patterns". Similarly, Clements and Li (2018) compared food data for over 150 countries and found that rich countries had a much more diverse diet than poor countries. Hence, the increasing affordability of food over time in the U.K. has significant potential benefits on the nutritional quality of the diet through the ability to buy a greater variety of food and food of a higher quality. Increasing affordability will potentially also result in more food being purchased and so this could be a major obesity factor. This is likely to be an important aspect of the greater food abundance mentioned in a couple of the Foresight report quotes in Section 6.3.1.

Of course, economic conditions in the future could bring changes in the population levels of income, food prices, and other elements of personal expenditure, resulting in either an increase or decrease in the affordability of food. For example, energy and other costs including food increased significantly in 2022 and the inflation rates during the year were high. The HSE data used in this report only goes up to 2018 and so the conditions relevant for the results here are just those up to 2018.

Whilst at a population level prosperity has increased considerably over the last 100 years there will clearly be very wide variations in the circumstances within the population in all decades. Food expenditure as a percentage of income or of total expenditure for some households will be very different to the average population values given in the above data. A thorough analysis of the economic situation over the years and of the variation in household finances across the population is outside the scope of this report. A detailed evaluation would need to consider the range of different types of expenditure and the very different circumstances in the population. The Foresight report (Butland et al., 2007) comments on some of these complexities (page 57), with various factors suggested that will affect exactly how food prices are related to purchasing decisions including income, age, the size of the household, and other costs such as housing.

The HSE data can be used to look at some differences in demographics and circumstances within the population. For example, the HSE data has a variable for the level of deprivation with five categories. Some analysis was done comparing the BMI distributions for the deprivation categories and the results for this are given in Chapter 15. The results in that chapter show that BMI and obesity levels are higher for categories of greater deprivation, particularly for the data for women.

6.3.4 Other factors

As mentioned at the start of Section 6.3.3, although the factor of income and food prices has been discussed in some detail this is not necessarily the most important factor for obesity. From the discussion in Section 6.3.1, there are very many potential factors that affect obesity.

There are many aspects of lifestyles and circumstances have changed considerably over the past 100 years and are likely to have had a significant effect on obesity. For example, other economic and society factors that could be relevant include general economic cycles of booms and recessions, war and post-war time periods including military service and food rationing, changes in patterns of travel and work, changes in the availability and popularity of leisure activities, and the effects of technological developments. A detailed discussion of such changes is not included here, although the effects of such changes on obesity could be very considerable.

One particularly important factor affecting obesity is likely to be changes over time in the food industry. This will include changes in manufacturing such as food composition and the range of products produced. There will also be changes and developments in retailing in both the extent of the availability of food, and the mix of types of retail outlets whether shops or restaurants. There are a very wide range of types of restaurant such as fast food, coffee shops, specific cuisine, luxury dining, etc., with changing trends over time. There are also other food providers such as schools and workplaces and even vending machines, and again the availability and the type of food provided will have altered over time.

Some journal papers consider the number of different types of food outlets and how this might be related to obesity with concepts such as "food deserts" and "food swamps". A blog on the UK Health Security Agency website [35] (Blackshaw, Ewins and Chang, 2019) discusses the food environment and its effect on obesity. One obesity factor that they mention is more eating of meals away from home because such meals tend to have larger portion sizes and have greater amounts of sugar, fat, and salt. Hence, a greater frequency of eating out over time, as discussed in the previous section, could be a factor in increasing obesity levels. The blog has a graphic with the comment that one quarter of the places providing food in England are fast food outlets. It also has a chart of data by local authority showing that the density of fast food outlets tends to increase with greater deprivation. Again, a detailed discussion of the food environment and possible relationships with obesity is not included here in this report.

All possible factors have not been reviewed here in detail. However, a suggested brief list of lifestyle factors is given in the next Section 6.3.5.

[35] https://ukhsa.blog.gov.uk/2019/08/08/health-matters-addressing-the-food-environment-as-part-of-a-local-whole-systems-approach-to-obesity/

6.3.5 List of lifestyle factors

This section lists some suggestions of changes in lifestyles over the long term up to 2018 that might have affected BMI. Each factor could affect BMI in either direction, but most will probably tend to increase BMI. The list is based partly on the literature discussed in the previous sections and also to some extent based on personal recollections of changes in lifestyles observed in society. It is therefore somewhat subjective and speculative. The factors are:

- Cheaper food prices relative to income.
- Availability and changes in the popularity of different food and drinks products. For example, the emergence of new products such as microwave ready meals, certain processed foods, energy drinks, smoothies etc.
- Changes in ingredients and the composition of foods - e.g., sugar and salt content.
- Different number and mix of eating establishments such as greater food availability, and more fast food outlets and coffee shops. Also, greater availability of home delivery from shops and restaurants.
- Different attitudes to eating out with going to restaurants and coffee shops becoming more fashionable and more affordable relative to income.
- Different job types - fewer manual jobs, perhaps longer working hours and higher stress levels, or different types of stress.
- Different choices of leisure and exercise activities – technology providing new less active leisure pursuits such as computer games and using the internet, but also opening up new opportunities through things like fitness devices and apps. Also changes over time in trends and the popularity of different activities such as sport, fitness, dancing, etc. This will include some activities being developed and becoming more popular, such as parkrun.
- Less time available for leisure.
- Differences in transport such as more car use and less walking. Longer commuting times may also be a factor.
- Differences in cooking skills and the time available for cooking with these both perhaps becoming lower.
- Differences in food habits developed as children at school and at home.
- Different societal attitudes to food, work, and leisure.
- Different societal norms for BMI and body type.
- Changes in lifestyles and society perhaps affecting mental health, in turn affecting eating habits.
- Changes in the levels of other factors that might affect BMI such as alcohol consumption and smoking.
- Differences in household composition such as more people being single. If the person who is best at cooking in a household does most of the cooking then cooking quality will reduce if there are more single households. I.e., two single people each cook for themselves whereas if they lived together they can combine their skills or the person with the better skills might do more of the cooking.

6.4 How the different time intervals lead to BMI changes

It is not straightforward to evaluate how the different time intervals for the BMI populations lead to the BMI changes. Consider the example in Section 6.2 of a 35 year old person having a BMI of 26.2 in 1993. The example supposed that the nearest equivalent person in 1997 aged 35 might have a BMI of 26.7. The first person had a BMI of 26.2 from living in the time interval from 1958 to 1993. The second person's BMI of 26.7 is from the time interval 1962 to 1997. The assumption is that the different BMI is due to different lifestyles. One difference is that the second person experienced the years (and lifestyles) from 1993 to 1997 which the first person did not. Another difference is that the first person experienced the years 1958 to 1962 (as a young infant) which the second person did not. Both people experienced the years 1962 to 1993 but this was at ages four years apart and so this is a third difference. The change in BMI will be some mix of these different experiences.

This is of course true right across the population but the dates vary according to age. In all cases one difference is the later population experiencing 1993 to 1997 which the equivalent people in the earlier population did not. The other dates though are different. For example, a 20 year old person in the first population experienced 1973 to 1977 before the second person was born. The time that they both experienced but at different ages is 1977 to 1993. For each different age these dates will be different. For populations further apart in time then the difference in these dates will be greater.

Thinking about these complexities and interactions led to investigating how BMI is related to both age and year of birth, and how the relationship with age has changed over time. The resulting analysis on this is set out in the next Chapters 7-10.

6.5 Conclusions on the interpretation of the results in earlier chapters

The overall conclusion drawn here is that the differences in the BMI distributions are due to living in different times and experiencing different conditions and lifestyles. In particular, the assumption is that if the people in the 2013 population had all lived 20 years earlier then their BMI distribution would have been the 1993 distribution. There are many factors that have changed considerably over the range of time of more than 100 years covering the lifetimes of the people in the HSE data. The difficult issue for interventions is how best to counteract the effect on BMI of these changes. The earlier chapters have shown that BMI has increased considerably since the early 1990s, and so if it is possible to return BMI to what it was in those times then this would be a big change and a big improvement.

The analysis and results in Chapters 2-5 show that the changes in the population BMI distributions since the early 1990s are approximately a linear scaling transformation for both men and women. The results for the percentiles in Sections 4.6 and 4.8.3, and for the linear scaling models in Chapter 5, mean that on average those with a higher BMI have a higher increase in BMI in an approximately linear relationship.

Applying the interpretation in this chapter, the implication is that comparing equivalent people, someone in 2013 who would have had a high BMI in the early 1990s tends to have a high increase in BMI. Someone who would have had a low BMI in the early 1990s tends to have a low increase. Hence, the changes in the obesity factors through living in different times have, on average, affected people with a high BMI much more strongly than those with a low BMI.

More specifically, the effect of the changes in the obesity factors has resulted in an approximately linear relationship between the BMI level and the increase in BMI for populations at different times.

The interaction between age and BMI within the distributions is potentially quite complex. BMI will tend to change with age and is likely generally to increase with age. However, as set out in this chapter, different ages in a given population will also have lived in different times and this needs to be taken into account. The aim of the following chapters (Chapters 7-10) is to examine the relationship between BMI and age, whilst considering the effect of when people were born. In particular, Chapter 7 looks at the relationship between BMI and age within narrow birth cohorts.

Chapter 7: Cohort analysis – variation of BMI with age and birth year

Key points from this chapter:

- Looking at the simple relationship between mean BMI and age in the HSE data is of quite limited use in identifying the effect of aging on BMI. This is because people of different ages have lived in very different times and conditions, as discussed in Chapter 6. Therefore, the effect of aging is mixed with the effects of the different conditions.
- Looking at how mean BMI varies with age within narrow birth cohorts gives a much greater insight by limiting the variation in conditions. The people within a birth cohort have all lived in much the same time period and so the macro environmental conditions will be similar. Hence, the variation with age is mainly due to the effect of aging.
- The cohort analysis indicates that mean BMI tends to increase with age for all ages. Older cohorts are at lower levels, which is presumably because of different experiences earlier in life.
- The increase of BMI with age is greatest at the youngest ages with the rate of increase getting smaller with age.
- An equation for a BMI aging model is proposed based on the data analysis. This has mean BMI increasing towards a horizontal asymptote.
- The model for BMI and age is the suggested aging effect in current modern conditions. The variation of BMI with age may have been quite different in the past, and it is likely to change in the future if circumstances alter.

7.1 Overview

The analysis presented in Chapters 2-5 looks at how the overall BMI distribution for adults in England has changed over time. Distributions were derived in four year intervals from 1993 to 2013. For both men and women, BMI increases considerably over time in a pattern close to a linear scaling. The changes are much greater from 1993 to 2001 than in the years since then.

Understanding the reasons for the changes in BMI over time for a whole population is a difficult task. One aspect of this difficulty is that a population consists of people from a wide range of ages, who have therefore lived in different times and have experienced very different conditions and lifestyles. In the analysis in this report the population of interest is everyone in England aged 18 and over, and so ages will vary from 18 up to some aged over 100.

As discussed in Chapter 6, if we compare the population at two points in time then each age group is offset. As an example, suppose we compare the populations in 1993 and 2005. Considering the 20 year olds in the populations, we are comparing living in 1973-1993 with living in 1985-2005. If we look at those aged 35 we are comparing 1958-1993 and 1970-2005. For 50 year olds it is 1943-1993 and 1955-2005. Similarly for each age. One difference that they all have in common is the later 2005 population experienced life in 1993-2005, as their most recent 12 years, whereas the earlier 1993 population had obviously not yet experienced 1993-2005. However, the differences at specific times in life will vary for each age. For example, the first 10 years of life for 35 year olds in the two populations is 1958-1968 and 1970-1980, for 50 year olds it is 1943-1953 and 1955-1965, and so on.

The lifetime conditions experienced by the different age groups will vary enormously. Section 6.3 discussed various ways in which lifestyles may have changed over time, and over a long time period of the last 50 or 100 years will include major factors such as income levels and the affordability of food, war time and post war food rationing, economic booms and recessions, extensive changes in types of work and leisure activities, and considerable changes in the types and availability of food.

The aim of the analysis in this chapter is to look in detail at the relationship of mean BMI with age and how this has changed over time. This should give additional insights into the reasons why the overall population BMI distribution has changed in the way demonstrated in Chapters 2-5. The BMI distributions for different ages are analysed in Chapters 8-10. The relationships found with age lead on to a framework for modelling BMI and generating future scenarios, and this is set out in Chapters 11 and 12.

The approach taken in this chapter attempts to disentangle the effects of age and the specific time period. It does this by analysing the relationship of BMI with age for narrow birth cohorts. Section 7.2 explains the data used. Preliminary analysis of how BMI varies with each variable of age and birth year on their own is given in Section 7.3. This gives initial context but is of fairly limited use because of the interaction between the two variables. The cohort analysis is set out in Section 7.4, and this gives a much greater insight into how BMI varies with age. A suggested aging model based on this is presented in Section 7.5. The changes over time are examined in Section 7.6 by looking separately at 1991-2002 and 2003-2014.

7.2 Data used

As is the case throughout this report, the main data used is from the Health Survey for England (HSE). As an aside, the “Data table” or “Trend table” spreadsheet produced alongside the HSE data and discussed in Section 1.2.5 and Section 2.5 has some age statistics from the HSE. It gives the mean, category percentages, and standard error, for the age categories of 16-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+. Therefore, these statistics could also be used for some age analysis.

The analysis here uses the actual HSE data for 1991-2014 rather than the smoothed distributions derived in Chapters 2-5. This is because the direct relationship with specific cases is lost when smoothing the distributions and so BMI cannot then be related to the age of the cases. As for the earlier chapters the data for 1991-2014 is used as this is what was available at the start of the work. Also, the HSE data for 2015-2018 only gives the age of each case in five year categories (18-19, then 20-24, 25-29, 30-34, etc.). This limits what can be done, although Chapter 14 has some age analysis for the 2015-2018 data.

The data is analysed by age and birth year. The birth year was calculated using the equation of: birth year = HSE year – age. The HSE data is collected throughout the HSE year and age is of course a truncated value. For example, a person recorded as age 50 in the 2010 HSE survey could be aged from 50 years 0 days to 50 years 364 days. The date of the survey could vary from January to December 2010. Therefore, the person could actually be born any time in a two year time period between January 1959 and December 1960. The birth year calculation above gives a value of 1960. It really represents 1959-1960 and strictly should be interpreted in this way. The middle of the range of dates is the start of January 1960 and so the birth year could be read as the start of the particular year being the mid-point of the range of possible birth dates. Similarly, 1961 represents 1960-1961 and could be thought of as being the mid-point of the two years as January 1961, etc. Slight refinement could have been made using the survey month variable in the HSE data. However, the analysis here does not need a very precise value for the date of birth and the above calculation is considered accurate enough. The analysis here uses the calculated value from the above equation, and the text and charts will also just use this value (e.g., 1960 rather than 1959-1960).

Some of the analysis in this chapter plots BMI values for different ages. As already mentioned, an age in years is really an interval so an age of 18 means 18.0 years ≤ age < 19.0 years. This interval has a mid-point of 18.5 years. However, the charts plot the BMI value for this age against the x-axis value of 18, so as to accord with the common usage and understanding of age in years. In other words, the x-axis gives the truncated integer value (i.e., the lower bound of the interval). Plotting against the mid-points would give exactly the same pattern but just shifted right along the x-axis by 0.5 years.

As discussed in Section 2.1, the data from 2003 onwards has an interview weighting variable that adjusts for various demographic and non-response characteristics of the cases. This is most applicable for using the survey data for a particular year as a whole, which is not what is being done here. Some factors accounted for by the interview weighting such as social or regional aspects are relevant but correcting for the age profile within the HSE sample for the year is not.

Specifically, in the cohort analysis in this chapter, cases are combined across many HSE surveys based on their birth cohort. In the analysis, cases of the same age from different HSE survey years are grouped together and averaged. In this context, it was not considered desirable for cases to have an increased or decreased weighting just because of the relative number of cases of that age in the survey for the year. Also, for some of the analysis cases are combined across all HSE years and cases before 2003 do not have an interview weighting variable.

For example, calculating the average BMI for age 25 for the cohort of people born between 1975 and 1984 combines cases of age 25 in the HSE surveys for 2000-2009. The overall age profile of the whole 2005 survey, for instance, should not affect how an age 25 case from that survey is weighted in the cohort analysis.

Various weighting options were considered. In the end, it was decided to use the interview weights but adjust them so that the mean case weight for each age within each HSE year equals 1. This avoids the above effect of the weighting of cases being partly due to the relative age prevalence in the survey for the year.

As a test of the weighting, the cohort analysis for the 1991-2014 HSE data was repeated using the original HSE weights, the weights from the Chapter 3 analysis, and equal weights for every case (i.e., all case weights equal to 1). The results produced have very little difference in the charts of the data compared to those given in this chapter in Figures 7.3 and 7.4, and the general patterns are the same. The difference that was considered most noticeable is that for equal weights the cohort values for women for 1985-1996 are slightly higher (on average the mean BMI values are 0.2 higher) but with the same basic pattern. This is still a very minor difference.

This indicates that the results are not particularly sensitive to the choice of weighting. Given the good sampling methodology of the HSE and the type of analysis being done here, it is not surprising that the alternative ways of weighting the cases make no real practical difference to the patterns observed and the interpretation of the results.

7.3 Variation of mean BMI with age and birth year

One of the keys to understanding the changes in population BMI is to analyse how BMI varies with age. The analysis in this section takes an initial look at the overall age and birth year pattern in the complete HSE data for 1991-2014. This gives some insight into the patterns but is limited in just looking at the variables separately. The analysis is then developed in the next section by considering the interaction between age and birth year.

Calculating the mean BMI for each age for the whole HSE data for 1991-2014 gives the chart in Figure 7.1. The values for age 90 are for actual age 90, although in the HSE for 2014 age is capped at 90 and so the data for that year is all those aged 90 and over. Figure 7.1 shows BMI increasing with age up to around 60 and then decreasing. Mean BMI increases faster for men than for women at the youngest ages. The curve flattens out for men as age increases. BMI increases at more of a steady rate for women. The result is that mean BMI is similar for men and women at ages of about 20 and about 60, but is higher for men in between those ages.

Mean BMI can also be calculated by birth year and this is shown in Figure 7.2. This has a very similar shape to Figure 7.1, but in reverse. BMI increases as birth year gets earlier to about 1940 then reduces. The similarity in pattern between the two charts is because of the relationship between age and birth year, in that those older in the survey will be born earlier. Similar charts can also be calculated for shorter time intervals of HSE data and this is done in Chapters 8 and 9.

It is well known that weight and BMI tend to increase with age. However, it is important to try and identify the precise details and shape of the aging pattern. Combining the whole HSE data from 1991 to 2014, as in Figure 7.1, is limited and potentially misleading in mixing birth cohorts across a wide time interval. For example, 60 year olds in this data will vary in the dates from birth year to survey year between 1931-1991 and 1954-2014. The 60 year old values that produce the mean value for age 60 in

Figure 7.1 will therefore include cases for 1931-1991 and 1954-2014 and for each time period in between. Each case has experienced the human aging process over 60 years of life but the external environmental conditions will vary considerably between the different cases. In addition, people of different ages will have experienced very different early life conditions. For example, the early life conditions of 30 year olds will be very different to those of 60 year olds. Therefore, they may be starting from a quite different point in experiencing the recent modern conditions of the HSE period.

BMI will be affected by both aging and the time period lived in. Section 7.4 looks at splitting out these effects by considering aging within narrow birth cohorts.

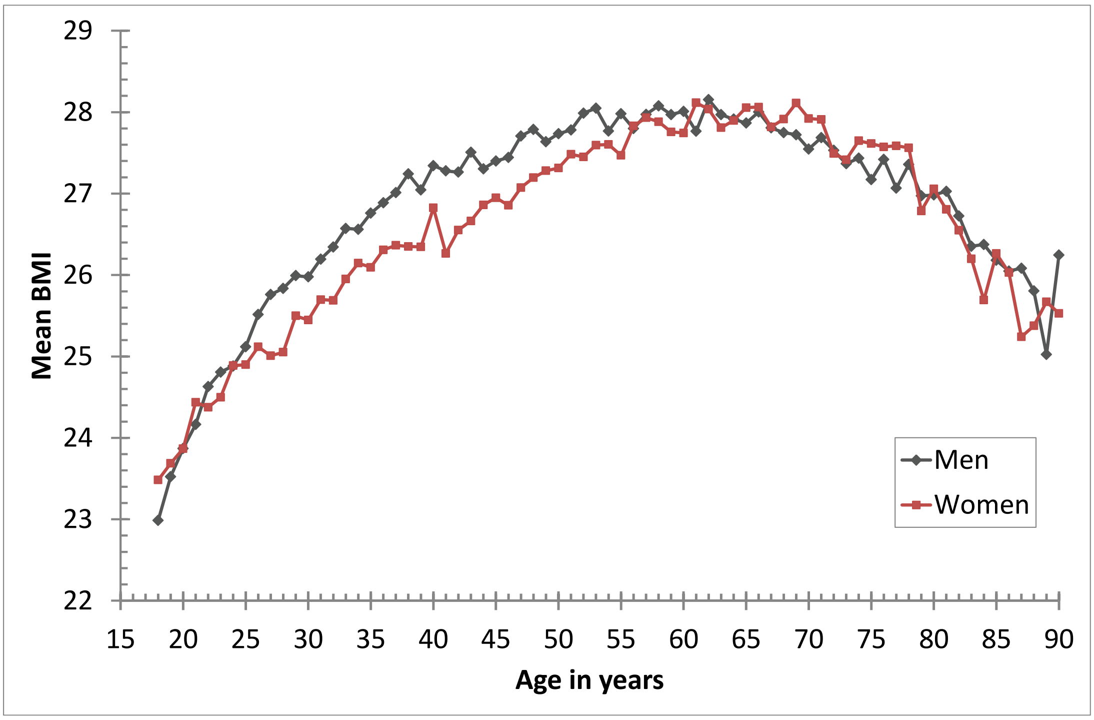


Figure 7.1 Mean BMI by age for the HSE data for 1991-2014.

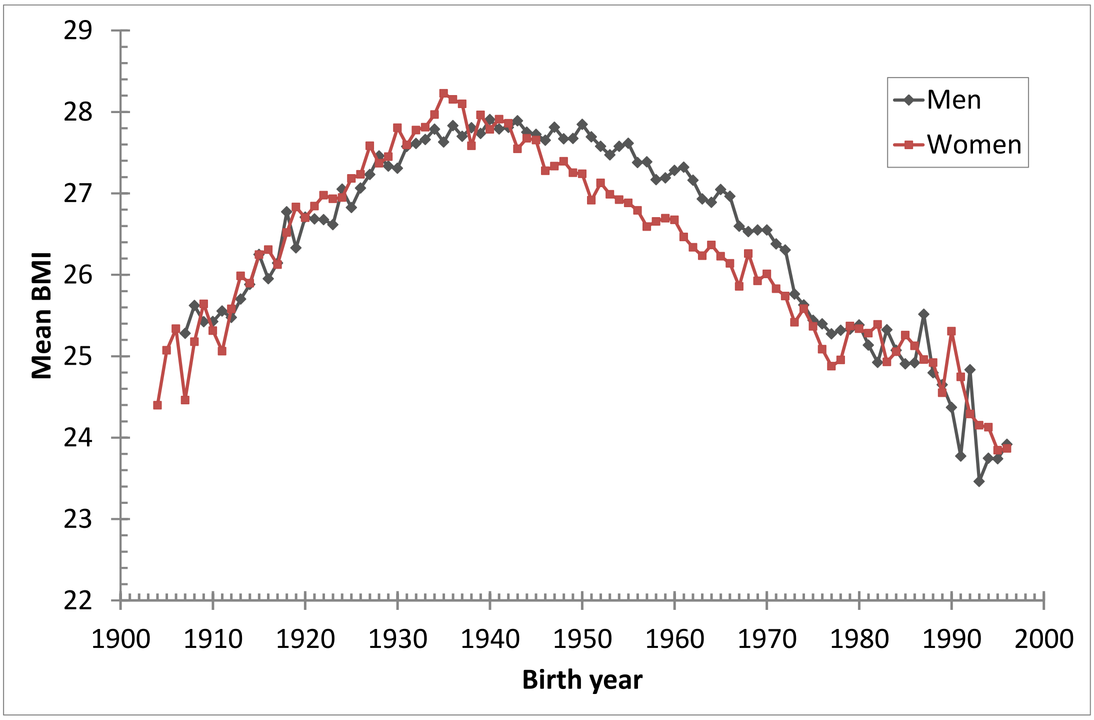


Figure 7.2 Mean BMI by birth year for the HSE data for 1991-2014.

7.4 Cohort analysis: looking at both aging and birth year

7.4.1 Results for 10 year cohorts for the whole HSE data

This section describes the age analysis done by birth cohorts. This looks at both the effects of aging within cohorts and the differences between cohorts. It is considered that the results here give particularly important insights into the BMI patterns and how BMI is changing over time.

The analysis here was done by grouping the HSE data into birth cohorts and calculating the mean BMI for each age in the cohort. Figures 7.3 and 7.4 show the results for 10 year cohorts for men and for women (apart from the most recent group of 1985-1996 which is 12 years) using the entire HSE data for 1991-2014. To avoid using means from small samples, only values where there are at least 50 cases are included on the charts.

Figures 7.3 and 7.4 show a couple of strong and interesting patterns. One is that each series conforms to a general trend of BMI increasing with age, albeit with some fluctuations. The rate of increase is greatest at the youngest ages and gradually gets less as age increases. None of the series show any particularly noticeable decrease in BMI with age, although above age 70 each curve looks roughly flat. Overall, this would imply a pattern of an increasing curve but with a horizontal asymptote.

The other pattern is that the different cohorts are at different levels, with each later cohort (i.e., more recent birth dates) being at a higher level. For successive cohorts the later cohort generally has a higher value for those ages where there are values for both cohorts. For example, the BMI values for 1915-1924 are higher than 1905-1914 for most points. Then, the BMI values for 1925-1934 are mostly higher than those for 1915-1924. Similarly for all the pairs of cohorts right up to the most recent one for 1985-1996 being higher than 1975-1984.

The differences between adjacent cohorts can be calculated for the ages where there are BMI values for both cohorts. Using the data plotted in Figures 7.3 and 7.4 (i.e., ages with at least 50 cases) the mean was calculated of the difference in values for each age. The mean values are a measure of how much a cohort is higher than the next earlier cohort (i.e., the cohort with a birth year range of 10 years earlier). Starting with the most recent cohort of 1985-1996, the differences in values for the cohorts for men are: 0.40, 0.14, 0.68, 0.75, 0.63, 0.78, 0.87, 0.91. For women the differences in values are: 0.61, 0.38, 0.80, 0.59, 0.27, 0.62, 0.82, 0.78. There are therefore sizable increases in mean BMI for each cohort over the next one.

Similar to the discussion in Section 7.1, when comparing different cohorts at the same age the time periods are offset. The data will come from different HSE years and so the lifetimes experienced will be offset. The amount of difference in the HSE years will vary. For the middle of the overlapping chart values for the two cohorts it will be 10 years (e.g., HSE 2003-2012 compared to HSE 1993-2002). For the end values it could be 5.5 years (e.g., HSE 2014 compared to HSE 2004-2013), depending on which ages have at least 50 cases. On average the difference is about 8 years. The impression from Figures 7.3 and 7.4 are that the cohort series are at different levels but are also increasing in level over the course of the series (i.e., as the HSE dates get later). Hence the differences in BMI values are likely to be due to some combination of changing recent conditions over the time span of the HSE and different experiences earlier in life.

Hence, within the time span of the HSE from 1991 to 2014 the trajectory of the cohorts may well have shifted in level during that time and this may be one reason why the later cohorts are higher. In other words, some cohorts may be on a higher level in 2014 than they were in 1991.This particularly looks like it might be the case for 1935-1944 and 1945-1954 for both men and women.

For men, the differences are smaller for the two most recent cohorts. This may indicate the effect of modern lifestyles tending to make recent cohorts converge. This is compared to older cohorts who possibly had a wider variety of different early life experiences resulting in greater differences.

One aspect from the charts that might be a positive pattern is the cohort for men for 1975-1984. Most of the BMI values for the ages in the 30s are lower for this cohort than was the case for the earlier cohort of 1965-1974. This could be sampling variation but perhaps could be due to the current generation taking on board in their 30s some of the messages regarding obesity and exercise and therefore having a slightly healthier lifestyle than the previous generation. It will be interesting to see whether this trend continues in future data. Also, to compare the trajectory of the 1985-1996 cohort when they get into their 30s to see how this cohort compares (assuming such data is available).

The pattern for mean BMI for all HSE data that was plotted in Figure 7.1 is shown as the dotted lines in Figures 7.3 and 7.4. The reason that mean BMI decreases at older ages on this curve is because the older cohorts are at lower levels, not because BMI is decreasing with age within a cohort.

Some slight decrease of cohort BMI might be expected at older ages due to those with a lower BMI perhaps tending to be healthier and living longer. There can also be an effect of reduced muscle mass and bone density reducing the BMI value for older ages, although height might also reduce. There is no especially clear pattern of this in these charts, although perhaps a hint of it in the data for women over 80 in Figure 7.4.

The 10 year cohorts do mix slightly different life experiences. Take, for example, the series of values for men born in 1955 to 1964 in Figure 7.3. There is no value for age 27 as the sample size is not large enough (just 30 cases). The left hand point is for age 28 and is people living in 1963-1991 and 1964-1992 (75 cases). The next point for age 29 is 1962-1991, 1963-1992, 1964-1993 (239 cases). Points in the middle cover a wider range of time periods and have much larger sample sizes. For example, age 39 varies between 1955-1994 and 1964-2003 and has 1092 cases. The right hand point for age 58 is just 1955-2013 and 1956-2014 with 97 cases. Age 59 only has 43 cases and so is not included. So, the left hand points in the series come from the earliest HSE surveys and the right hand points from the latest HSE surveys. There is therefore some variation in time periods within the cohorts, but the assumption is that the general lifetime environment (at a macro population level rather than a micro individual level) for each cohort will be reasonably similar.

The points at each end of the series are less reliable than those in the middle being smaller sample sizes. Some of the series in Figure 7.4 have some quite big fluctuations at the ends of the series (such as 1935-1944, and 1955-1964). This is likely to be due to sampling variation with the small sample sizes rather than indicating a change in the pattern.

7.4.2 Alternative choices for the cohorts and age ranges

The choice of the length of the cohort and the specific dates is arbitrary. Various choices were tried including 5 year cohorts. All gave similar patterns. The charts for 5 year cohorts are included in Appendix 7.I. Using 5 year cohorts compared to 10 year cohorts has the effect, of course, that the birth ranges are narrower and the life experiences within the cohort will consequently be less varied. There are more cohorts and so more series on the 5 year cohort charts. However, the number of cases is much lower at about half the number, and so there is more variability and irregularity due to the smaller sample sizes. The number of points in each series is also slightly less due to using a minimum of 50 cases, being generally 5 fewer.

Cohort charts were also produced for wider age categories to smooth the data. These are the same as Figures 7.3 and 7.4, except using ages in the categories of 18-19, and then five year intervals of 20-24, 25-29, 30-34, etc. The whole HSE data for 1991-2014 was used. This gives larger samples compared to the previous analysis of looking at single year ages. It therefore smooths out the data, although some details of the patterns are lost. These are also included in Appendix 7.I.

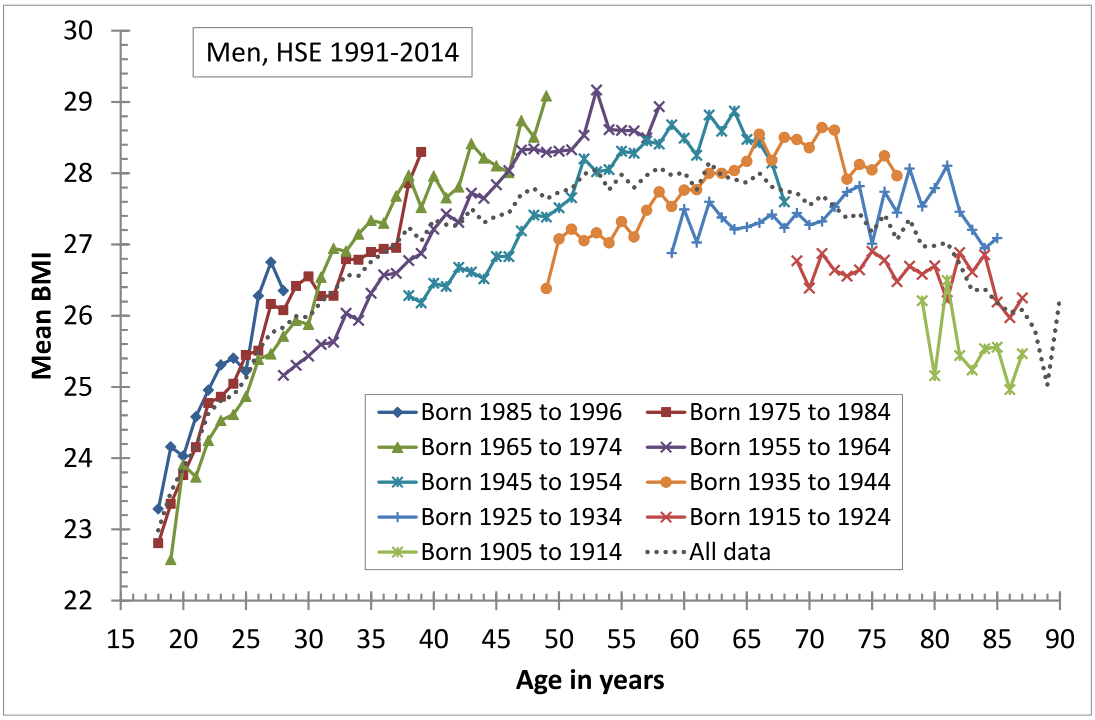


Figure 7.3 Mean BMI by age for 10 year birth cohorts for men.

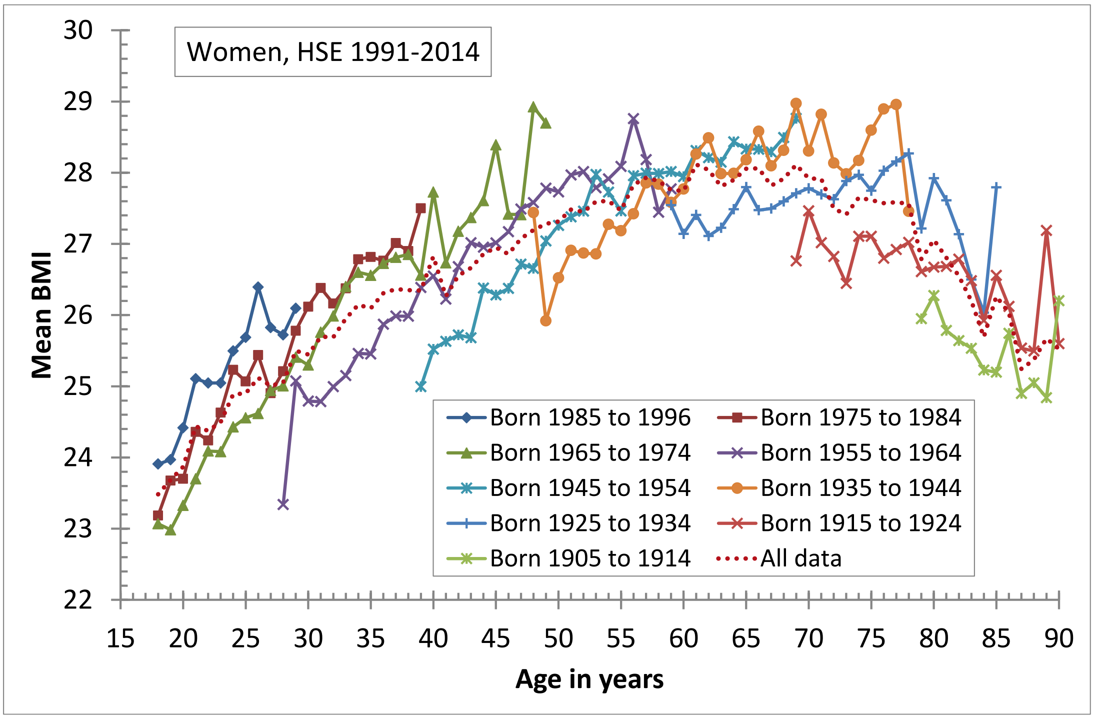


Figure 7.4 Mean BMI by age for 10 year birth cohorts for women.

7.5 Modelling the aging pattern

The aim here is to select a plausible model to represent the current aging profile for mean BMI for adults aged 18 and over. As discussed in Section 7.4, one key aspect of the patterns in Figures 7.3 and 7.4 is that within the cohorts mean BMI increases with age, but towards a horizontal asymptote. The rate of increase reduces with age and is roughly flat for the oldest cohorts. Each older cohort is at a lower level. A proposed pattern for the current aging profile is to follow the most recent cohort for the early years and then continue the curve with the general pattern of the increase of the other cohorts towards a horizontal asymptote.

A curve with these characteristics can be obtained using the equation:

$$\text{Mean BMI} = P_1 - P_2\exp(-P_3 \times \text{age}) \qquad (7.1)$$

for parameters P_1, P_2, and P_3. The parameters are used to shift and change the precise shape of the curve. The parameter P_1 is the horizontal asymptote, and so is the limiting value for mean BMI (i.e., in the equation, mean BMI $\to P_1$ as age $\to \infty$). The other two values can be changed to alter the starting level and the way the slope changes.

When choosing a curve, the values actually selected were P_1, P_2 and the mean BMI value at age 18, B_{18}. The value of P_3 was then calculated from these values, using Equation 7.1, as:

$$P_3 = -\text{LN}((P_1 - B_{18})/P_2)/18 \qquad (7.2)$$

The mean BMI at age 18 was used rather than P_3 as it has a clearer meaning and enables the starting value of the curve at age 18 to be specified directly.

Parameter values were chosen to produce a curve of the estimated current aging profile using Equation 7.1. This was a subjective process of trial and error where the aim was to follow the general trend of the most recent upper values. The curve was selected to try and follow the values of most recent cohort (i.e., 1985 – 1996) at the lowest ages and then follow the general shape of the other cohorts. The older cohorts are at lower levels and so the curve is designed to be roughly parallel to their shape, focussing more on the later values towards the right end of each series. In other words, the curve is intended to represent the path that the most recent cohort might follow if their pattern in the future is similar to the shape of the aging profile for older cohorts. The resulting curves are the dashed lines in Figures 7.5 and 7.6.

The parameter values of the curve for men are $B_{18} = 23.4$, $P_1 = 29.5$, $P_2 = 20.0$.

The parameter values of the curve for women are $B_{18} = 24.0$, $P_1 = 28.9$, $P_2 = 13.7$.

The aging curve in each of the charts in Figures 7.5 and 7.6 is a hypothesised current trajectory for the younger generations in the population. It can be used to construct scenarios of what might happen to BMI in the future, and this is done in Chapters 11 and 12. It also provides a benchmark against which future data can be compared so as to assess whether the situation is improving or getting worse.

Three other curves were also chosen to show similar paths but at lower levels. These are plotted in Figures 7.7 and 7.8. These, to some extent, represent levels that other cohorts might currently be on. So, in Figure 7.7 the three curves might be current paths for the 1935-1944 cohort, 1925-1934 cohort, and 1915-1924 cohort respectively. The 1905-1914 cohort is at an even lower level. The values used for all four curves are given in Table 7.1.

The lower three curves can also be considered as better paths to aim for that would end up where some of the older cohorts currently are. Scenarios can be created based on these curves (Chapter 12). They are not intended to try and show the historical path of the older cohorts. Indeed, given some of the major events in the past 100 years the older cohorts may well have followed a path with a different shape. It is not considered that there is one inevitable shape for the aging curve. Instead, the variation of BMI with age could follow quite different shapes depending on the external environmental conditions during the lifetime of a population. There will be an interaction of these factors with natural biological changes that happen as the body ages.

For example, someone born in 1920 will have experienced World War II (WWII) in 1939 to 1945, including active military service for many, from age 19 to 25. Food rationing occurred during and after the war from 1940 until 1954 and so would be from age 20 to 34. Therefore, the cohort born around this time might have had a path where BMI did not increase much during their 20s and early 30s but increased afterwards and so the pattern may be more of an S shaped curve for this group (speculatively perhaps BMI around 22 for ages in the 20s then rising afterwards). Cranfield et al. (2015) calculated mean BMI by age from 18 to 40 for a sample of 9799 male Canadian soldiers in WWII. The results are in Appendix Table A of their paper. The sample sizes tend to get smaller with age in the data being less than 200 for ages over 30 compared with over 1000 for each age from 18 to 20. The mean BMI values increase with age between 18 and 30 from 21.1 to 22.5. There is some variability for ages 31 to 40 due to the small sample sizes with BMI values being between 22.5 and 24.1. These are much lower values than for the HSE data in Figure 7.3. Cranfield et al. (2015) also find that these BMI values are lower than for soldiers in World War I. Hence, this may perhaps indicate that obesity levels in Canada did not start to increase until at least the late 1940s, although it depends on whether the patterns were similar for the general population to those for the sample of soldiers. I am not sure if any analysis has been done of BMI for any U.K. service records during and after WWII but the results would be interesting. Figure 7.9 shows speculatively what an S shaped BMI curve might perhaps look like for men born about 1920, who are likely to be similar to the cases in Cranfield et al. (2015).

Food rationing is likely to have resulted in lower BMI but will have had a variety of health effects, and this presumably may well have included some negative effects. Considering the range of impacts of factors like this on health is beyond the scope of this report.

As already mentioned, the aging curve in Figures 7.5 and 7.6 is based on subjective judgement. One aspect of this is how close the cohorts for ages 30s, 40s, and 50s are likely to be to the aging curve for current conditions. The hypothesis is that the cohorts are approaching the curve and so are converging. This is explored further in Section 7.6 by splitting the HSE data and particularly looking at the more recent data.

The charts in Appendix 7.I for the five year cohorts and the wider age range also have the curves for the aging model. These provide an additional comparison of how the data compares with the model. The patterns are similar to the charts here for the 10 year cohorts and so the assessment of the aging model from the charts in the appendix is also that it is a reasonable hypothesised trajectory for current conditions.

Table 7.1 Parameter values for the asymptotic curves in Figures 7.7 and 7.8.

Curve	B_{18}	P_1	P_2
Men, upper (aging model)	23.4	29.5	20.0
Men, 2nd	23.0	28.6	15.5
Men, 3rd	22.6	27.8	13.5
Men, lower	22.2	26.9	11.5
Women, upper (aging model)	24.0	28.9	13.7
Women, 2nd	23.6	28.1	11.6
Women, 3rd	23.2	27.4	10.2
Women, lower	22.8	26.7	8.9

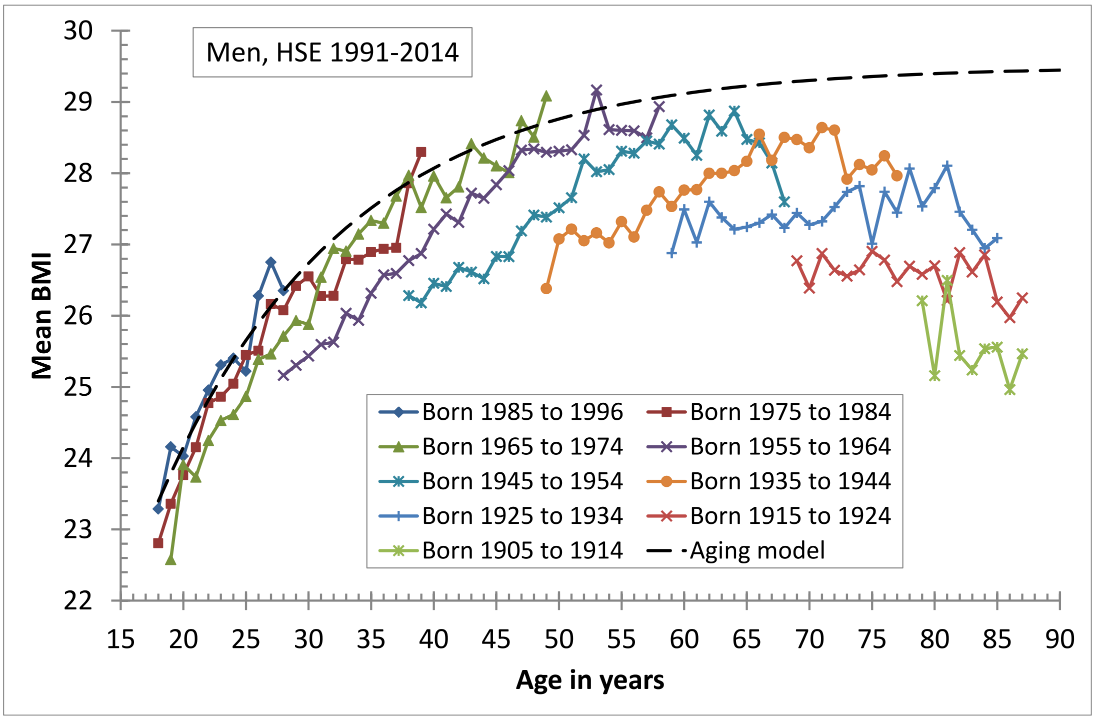


Figure 7.5 Cohort series for mean BMI for men from Figure 7.3 with the aging model curve.

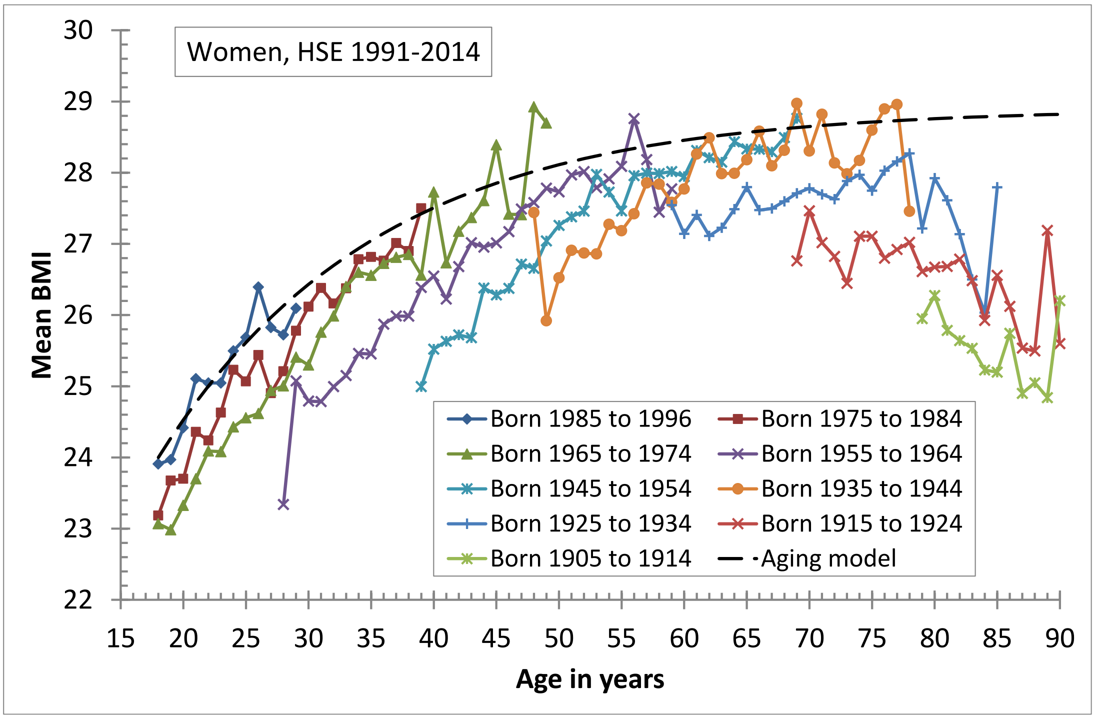


Figure 7.6 Cohort series for mean BMI for women from Figure 7.4 with the aging model curve.

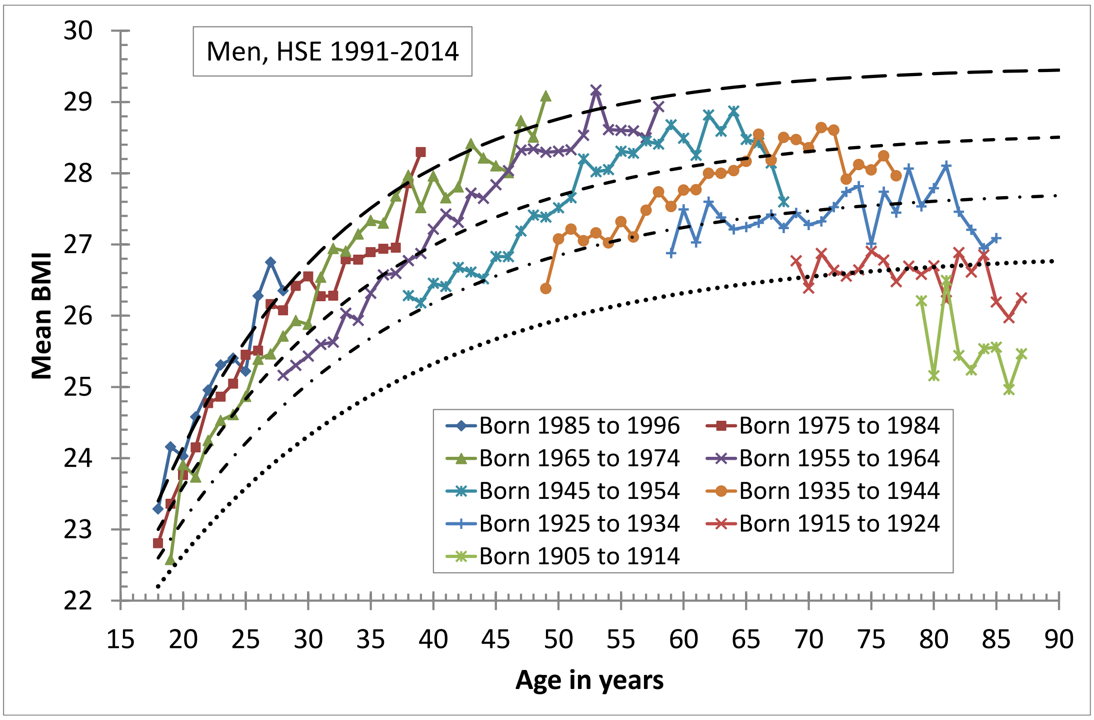


Figure 7.7 Cohort series for men from Figure 7.3 with four asymptotic curves.

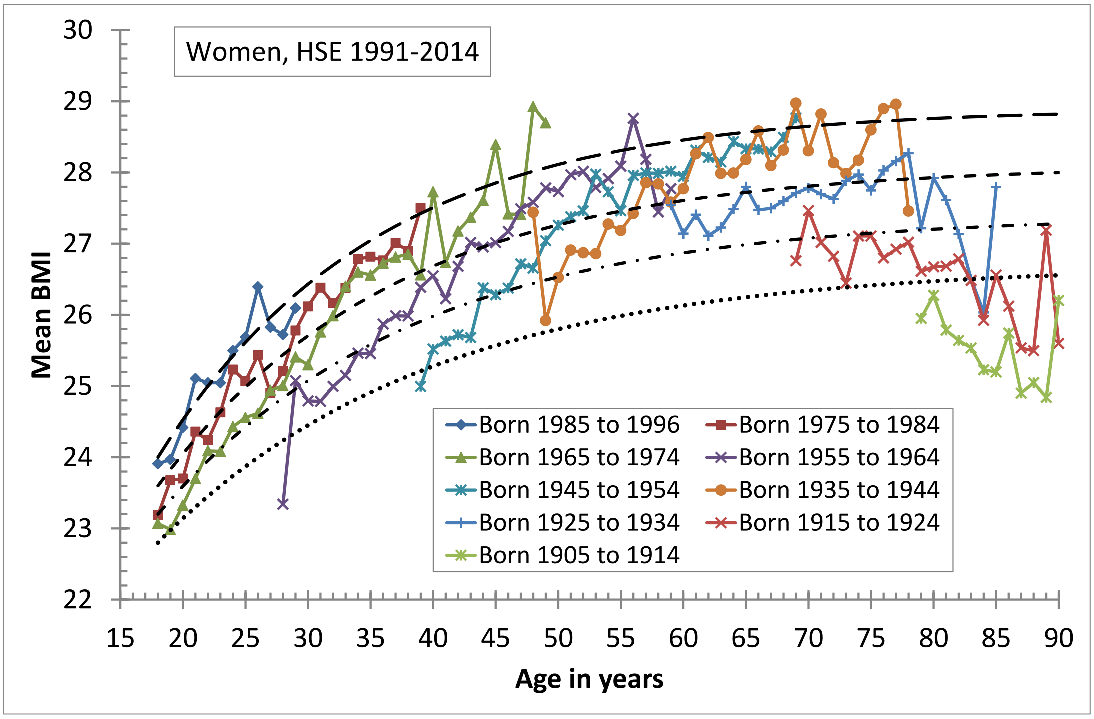


Figure 7.8 Cohort series for women from Figure 7.4 with four asymptotic curves.

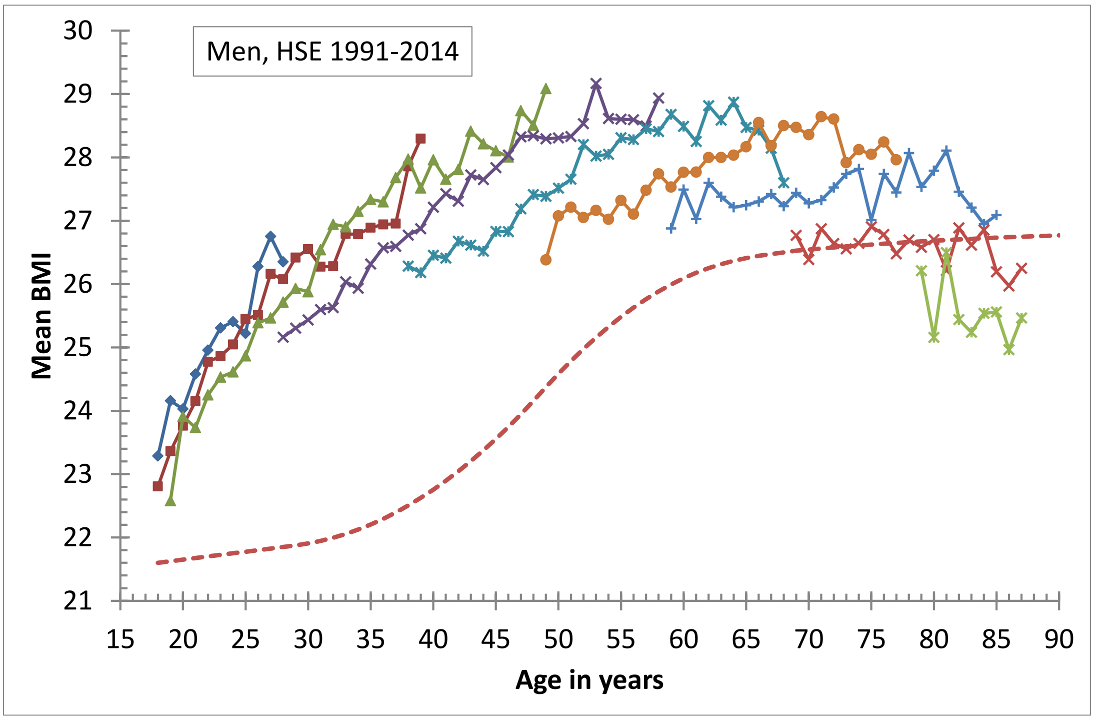


Figure 7.9 Cohort series for men from Figure 7.3 with the addition of a speculative 1920 curve. [The red dashed line is the speculative mean BMI age curve for those born in about 1920].

7.6 Cohort analysis for different periods of HSE data

Charts were also produced for shorter periods of HSE data. This helps to focus on the recent patterns and how things have changed. In order to have a reasonable amount of data the overall period of 1991-2014 was just split into two halves: 1991-2002 and 2003-2014. The resulting charts are approximately the curves from Figures 7.5 and 7.6 split into two. However, the 2003-2014 chart is a useful check that the asymptotic curves generally follow the shape of the later data. Comparing the charts for the two time periods shows how the patterns of the data have changed over time.

Figures 7.10 and 7.11 show the cohort values for men for the HSE data for 1991-2002 and 2003-2014, respectively. The asymptotic curves are provided as a reference. The charts for women are Figures 7.12 and 7.13. In the charts for 1991-2002 there are no values for the first cohort of 1985-1996 because the earliest that people in this cohort reach age 18 is 2003. In the charts for 2003-2014 there are no values for the cohort 1905-1914 because the only possible ages are 89 and 90 and there are fewer than 50 cases for these ages.

7.6.1 Cohort data for men for 1991-2002 and 2003-2014

For men, the chart for 1991-2002 in Figure 7.10 shows each cohort being on noticeably different trajectories to each other. For 2003-2014 in Figure 7.11 the cohorts have moved closer to the upper curve and are much closer together.

As in Section 7.4.1, the average differences in BMI can be calculated for successive cohorts. As before, the differences are the amount each cohort is higher than the adjacent earlier one and are listed starting from the most recent cohort of 1985-1996 through to 1915-1925. The values for 1991-2002 are: N/A, 0.38, 0.76, 0.75, 0.65, 0.51, 0.62, 0.83. The values for 2003-2014 are: 0.29, -0.18, 0.27, 0.21, 0.15, 0.47, 0.69, N/A. As explained above there are no values for 1985-1996 for the first time interval and none for 1905-1915 in the second, hence the not applicable (N/A) values. This data confirms the impression from the charts that the differences are much greater for 1991-2002. The average of the differences for 1991-2002 is 0.64, compared to 0.27 for 2003-2014.

The difference is negative for the second cohort of 2003-2014 and in Figure 7.11 it can be seen that most values for 1975-1984 are lower than the values for 1965-1974. As noted in Section 7.4.1, this may indicate a positive step of this cohort having a better lifestyle in their 30s for their BMI compared to the earlier cohort. It could, of course, just be sampling variation.

7.6.2 Cohort data for women for 1991-2002 and 2003-2014

Figures 7.12 and 7.13 show the charts for women for 1991-2002 and 2003-2014. Again, as for the data for men, there is a big difference between the two charts with the cohorts being at higher levels in the chart for 2003-2014. Figure 7.13 for 2003-2014 also shows considerable convergence with the cohort series approximately joining on from one to the next in a fairly smooth overall pattern, apart from the two oldest series being at lower levels. The most recent cohort appears to be at a slightly higher level which is why the top asymptotic curve is set a little above the general trend of the other series. However, the suggestion here is that the cohorts are close to converging to a trend reflective of current lifestyles.

In the same way as explained in the previous section, the differences between successive cohorts for women were calculated. The values for 1991-2002 are: N/A, 0.39, 0.78, 0.64, 0.33, 0.61, 0.54, 0.71. For 2003-2014 they are: 0.39, 0.00, 0.48, 0.15, 0.03, 0.37, 0.41, N/A. The differences are again much greater for 1991-2002 with an average of the differences of 0.57, compared to 0.26 for 2003-2014.

As for men, the lowest difference is for the second cohort in 2003-2014. For men the difference is negative but for women the difference is 0.00. However, again this may indicate the 1975-1984 cohort taking some positive lifestyle measures in their 30s, compared to the earlier cohort.

7.6.3 Interpretation of the cohort data for 1991-2002 and 2003-2014

The suggested interpretation is that the aging profile in modern conditions is given by the pattern of the aging model curves. In other words, if conditions were constant (for conditions at some point during the HSE time period), then each cohort would follow one of the curves. The cohorts are at different starting levels due to different experiences earlier in life. In addition, the changes in the conditions during 1991-2014 is tending to move the cohorts upwards onto higher curves. So, there is a combination over the time of the HSE of following the aging pattern of the curves and also some shifting upwards of the levels. This is just a hypothesis but it does seem to give a plausible explanation of the patterns observed. The suggestion is that the upper curve represents the aging pattern in the latest current conditions. If this is really the case and the current conditions (i.e., the many obesity factors) stay the same in the future, then over time all the cohorts and therefore the whole population would end up on the upper curve. The aging curves can be used for generating alternative future scenarios and this is done in Chapters 11 and 12. More recent HSE data for 2015-2018 is also compared with the aging curves in Chapter 14 in Section 14.7.1.

7.7 Implications and research questions

The general pattern identified for BMI and age is one of mean BMI increasing with age, and with particularly high rates of increase at the lower ages. This raises a number of questions as follows:

- Is this type of pattern mainly a result of the aging process and therefore difficult to alter?
- Given that age 18 has the lowest BMI values, can the pattern be changed so that BMI increases more slowly with age?
- At the moment standard BMI categories and guidelines are used for all adults. Would it be better to think of and try and identify a healthy lifetime trajectory of BMI, perhaps accepting that BMI will increase with age? This might imply having different guidelines for different ages such as a smaller BMI value for the upper limit of the healthy category for the youngest ages.
- How is the risk of health conditions related to the lifetime pattern of BMI? It is likely that the risk is some function of the amount of time spent at different BMI levels. Of course, this may vary for different health conditions. The patterns identified here and in the next couple of chapters might be useful for producing an estimate of the past lifetime trajectory of someone's BMI given their current BMI and age if past data for the person is not available.

These questions are not considered further here as being outside the scope of the work. However, they may be interesting issues for other researchers to investigate.

This chapter has focussed on mean BMI. The following three chapters look further at how BMI varies with age by deriving BMI distributions for different ages.

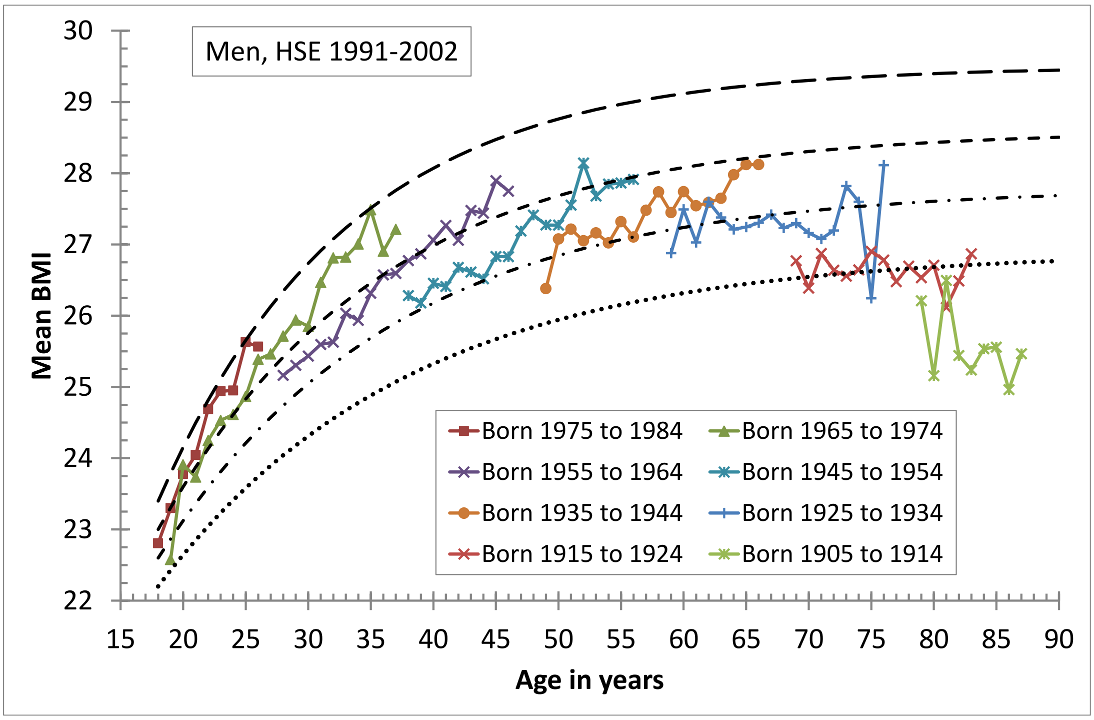


Figure 7.10 Mean BMI by age for 10 year birth cohorts for men for HSE 1991-2002.

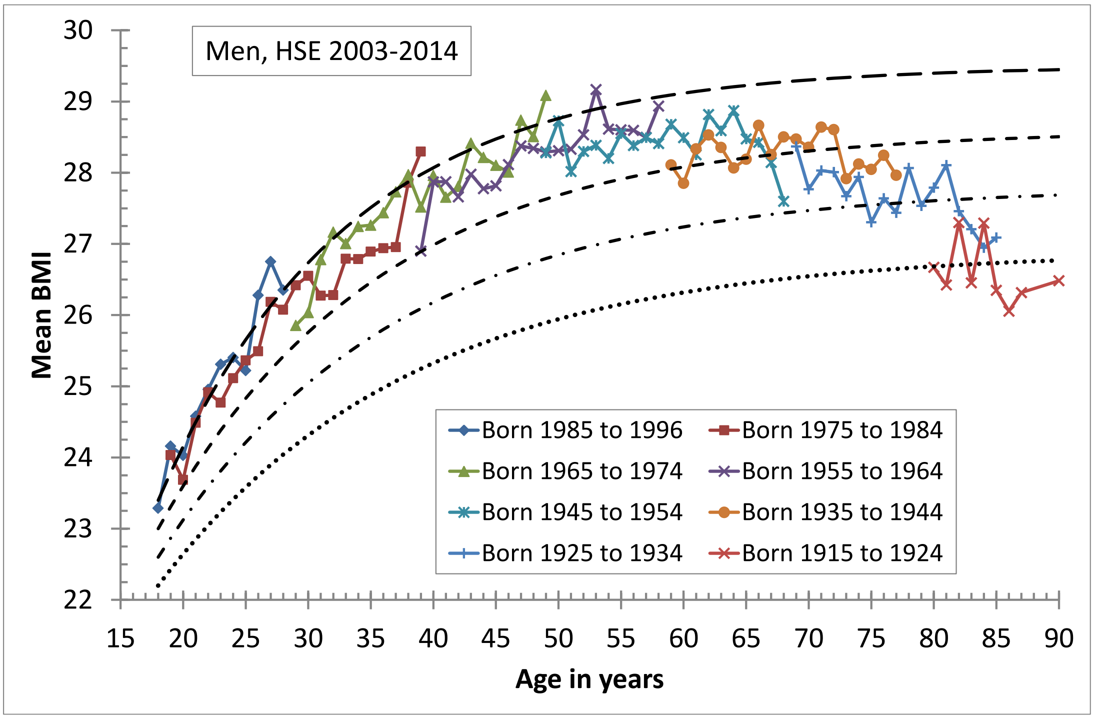


Figure 7.11 Mean BMI by age for 10 year birth cohorts for men for HSE 2003-2014.

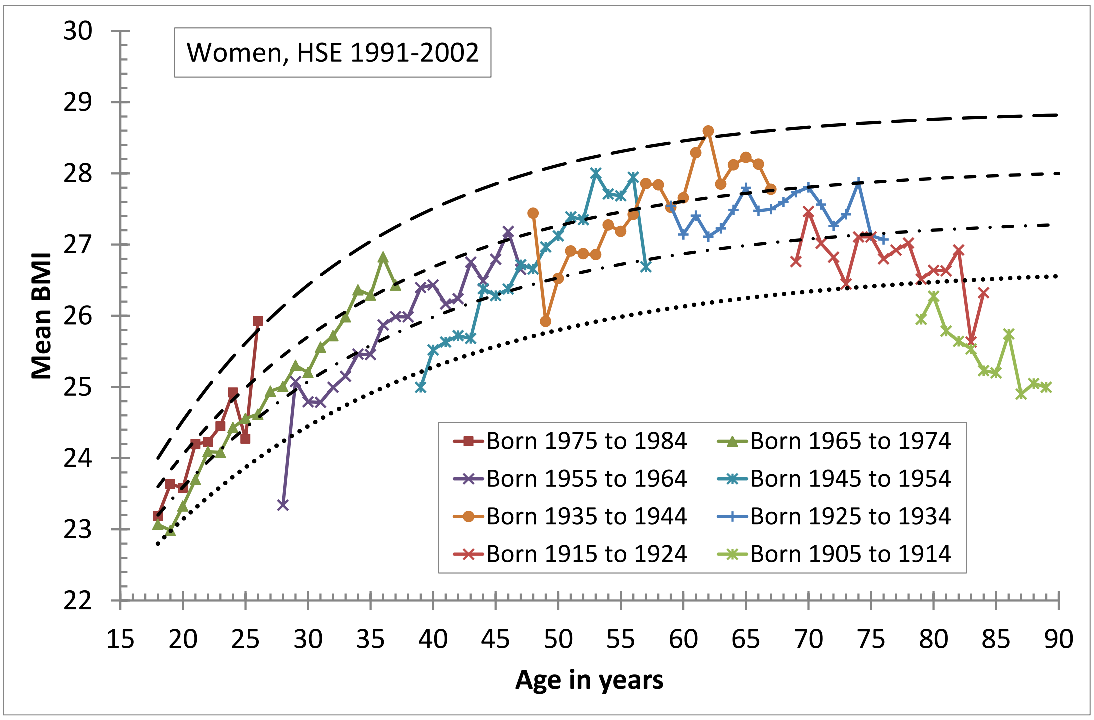


Figure 7.12 Mean BMI by age for 10 year birth cohorts for women for HSE 1991-2002.

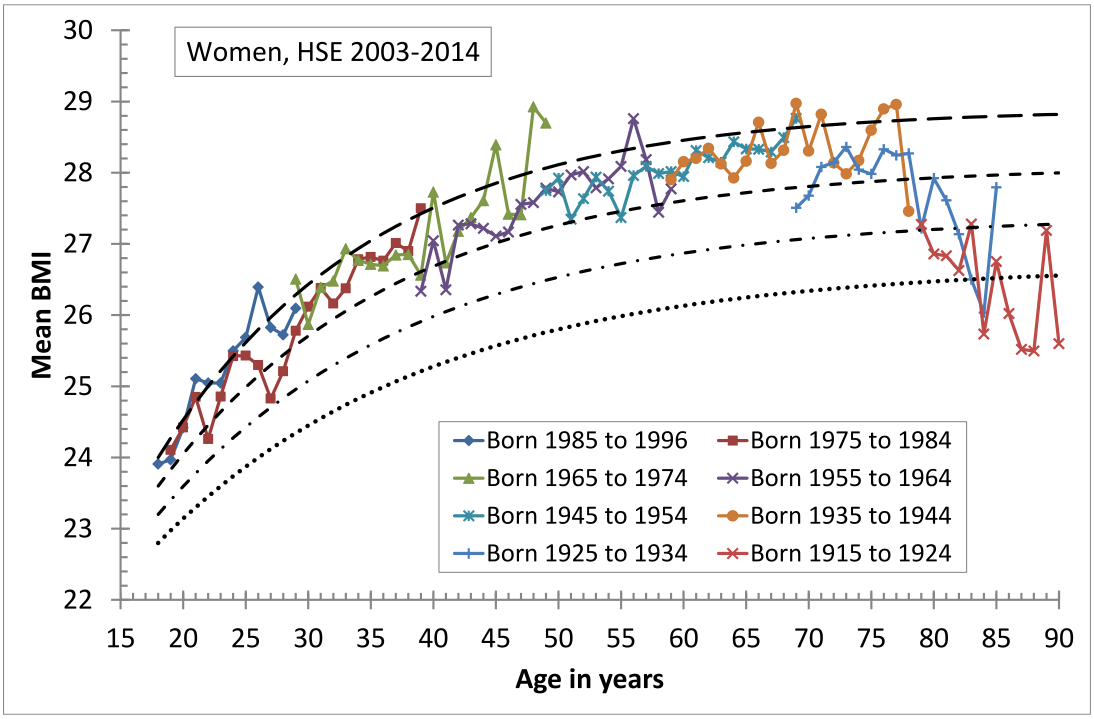


Figure 7.13 Mean BMI by age for 10 year birth cohorts for women for HSE 2003-2014.

Chapter 8: Age BMI distributions for men

Key points from this chapter:

- The main focus of this chapter is to look at the BMI distributions for different ages for men. Distributions were produced for age groups of about 10 years, and for single years of age. They were compared over time as to how they change for different HSE time intervals, and by age as to how they change by age for a particular HSE time interval.
- The changes in the BMI distributions over time are approximately a linear scaling. This is consistent with what was found for the population as a whole in Chapter 5. As discussed in Chapters 5 and 6, this implies that those with a higher BMI have been affected more by the changes that have occurred in the obesity factors. The finding here is that this applies for most ages.
- The changes in the BMI distributions with age have a different pattern to the changes over time. The changes from age 18 to 24 are mainly a linear scaling. However, above age 24 BMI increases with age mainly by the distribution shifting up the x-axis with little change in shape. Hence, above this age the tendency to increase in BMI with age is about the same for all starting BMI levels.
- The BMI distribution for age 18 has much healthier statistics than for older adults. For the 1999-2014 data there are 67.4% in the healthy BMI category with a mean BMI of 23.24. The values are even better for 1991-1998 with 77.6% in the healthy category and a mean BMI of 22.64. BMI increases rapidly from age 18 with each year of age. This indicates that targeting health messages and interventions at those of age 18 could perhaps be an effective and important strategy.

8.1 Overview

The results in Chapter 7 show strong and interesting relationships between mean BMI and age. The cohort analysis indicates that the relationship within each cohort is for BMI to increase with age, with the fastest rate of increase being at the youngest ages. The levels of the cohorts are also different, with earlier cohorts at lower levels. A hypothesised overall aging model was produced representing the suggested aging profile for current conditions.

Given the importance of the relationship between BMI and age, further analysis was done to look at the full BMI distributions for different ages. The results for the data for men are in this chapter. The results for the data for women are in Chapter 9. The distributions for men and women are compared with each other in Chapter 10.

One part of the work done was to categorise the HSE data into groups of ages, mostly with a range of 10 years (e.g., age 25-34). This is described in Section 8.2. A similar approach was taken to that in Chapters 2 and 3 to produce population distributions for BMI for the age groups, involving identifying the valid cases, calculating the relative frequency in 0.5 BMI intervals, and then smoothing the values. In this case, eight years of HSE data were used, rather than the four years of Chapters 2 and 3 so as to give reasonable sample sizes. This results in BMI distributions for the age groups for the three time intervals of 1991-1998, 1999-2006, and 2007-2014.

Having obtained the age group BMI distributions, the changes in the distributions were looked at in two aspects. One was to examine how the distributions change with time, by comparing the distributions for a particular age group over the three time intervals. This is similar to what is done in Chapter 5, where the changes over time for the complete population are shown to be approximately a linear scaling. Therefore, one question is whether this also applies to each of the age groups. This is considered in Section 8.3.

The other aspect was to look at how the distributions change with age. For this, the distributions for the different age groups were compared with each other for the most recent HSE time interval of 2007-2014. Again, one consideration is whether the changes are a linear scaling pattern. This is covered in Section 8.4.

Using age groups gives good sample sizes. However, it combines together a range of ages that have slightly different BMI levels. Therefore, some further analysis was done looking at the BMI distributions for single years of age. This is described in Section 8.5. The ages analysed are 18 to 54 in gaps of 6 years, and then each year of age from 18 to 24 so as to focus on the youngest ages when BMI increases at its fastest rate. To give larger samples only two time intervals of data were used, being 1991-1998 and 1999-2014.

The same data was used for the analysis here as in Chapter 7. This is the actual HSE case data using the interview weights, but with the case weights adjusted so that the average case weight for each age in each HSE year equals 1.

In an initial analysis the mean BMI was calculated for each age for each of the HSE time intervals of 1991-1998, 1999-2006, and 2007-2014. This is similar data to Figure 7.1 in Section 7.3. The results are in Appendix 8.I. As discussed in Chapter 7 this is limited in only looking at the relationship with age. There is a confounding effect of the birth year, with earlier cohorts being at lower levels. What Appendix 8.I shows is mean BMI moving towards the aging model from Chapter 7, in line with the discussion in Chapter 7. From 1991-1998 to 1999-2006 the increase in mean BMI is fairly similar across all ages with a mean increase of 0.81. From 1999-2006 to 2007-2014 there is very little change in BMI up to about age 45, with an increase in mean BMI of around 0.6 at ages above that.

8.2 BMI distributions by age group

The first analysis done was to calculate and compare the BMI distributions for age groups. The same general approach for producing estimated population distributions was used as described in Chapter 3. The weighted data was grouped into 0.5 BMI intervals and the relative frequencies smoothed using the Savitzky-Golay method (Section 3.5). The sample sizes here are smaller as they are split into age groups and so the 13 point rather than the 11 point smoothing was used in order to give slightly stronger smoothing (Section 3.5.2). The smoothing weights using this method for the 13 points are -11, 0, 9, 16, 21, 24, 25, 24, 21, 16, 9, 0, -11, divided by the total of 143. Descriptive statistics for the smoothed distributions were also calculated. Appendix 8.II gives some examples of the smoothing. The smoothed distributions are the estimates from the data of the population distributions.

Distributions were produced for the age groups of 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+. These are common age categories and were chosen to give a reasonable amount of data whilst still comparing different ages. From Chapter 7, BMI tends to increase most rapidly for the younger ages and so having a shorter first category helps to reduce the mixing of a wide spread of different mean BMI values. The distributions were produced using the HSE data for 1991-1998, 1999-2006, and 2007-2014. This is eight year intervals rather than the four years used for the analysis of the complete population BMI in Chapter 3, but this was done to give reasonable sample sizes. The number of cases in the samples vary from 2077 to 8004. The smallest sample sizes are for the 18-24 and the 75+ age groups.

All the resulting distributions are shown in Figures 8.1-8.3. Each chart shows the distributions for all the age groups for one of the HSE time intervals. This is the most concise way to show the distributions. Various other charts and descriptive statistics including BMI category percentages are given in Appendix 8.III. The distributions were compared both by how they change with time and with age and the results are discussed in the next two sections.

Distributions were also calculated using the case weights from Chapter 3. From a visual comparison of the charts, this produces very similar distributions indicating that the choice of weights has a minor effect for this analysis. Distributions were also calculated for four year HSE time intervals but the eight year versions presented here are preferred as having more data and therefore giving smoother distributions.

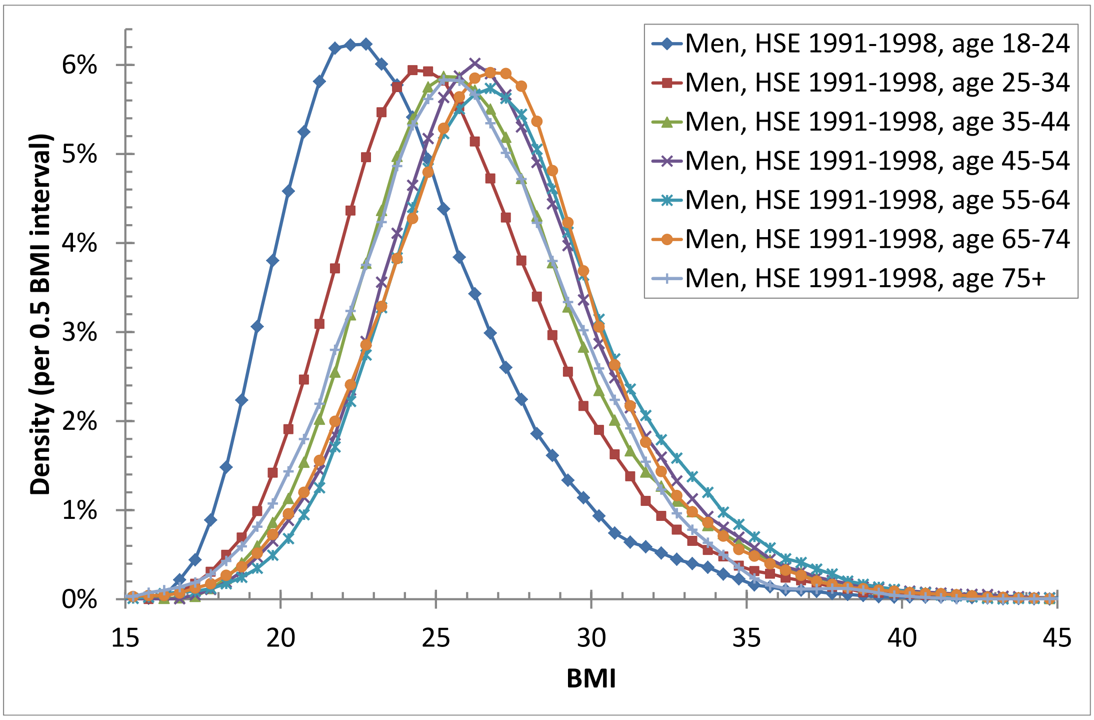


Figure 8.1 BMI distributions for different age groups for men for 1991-1998.

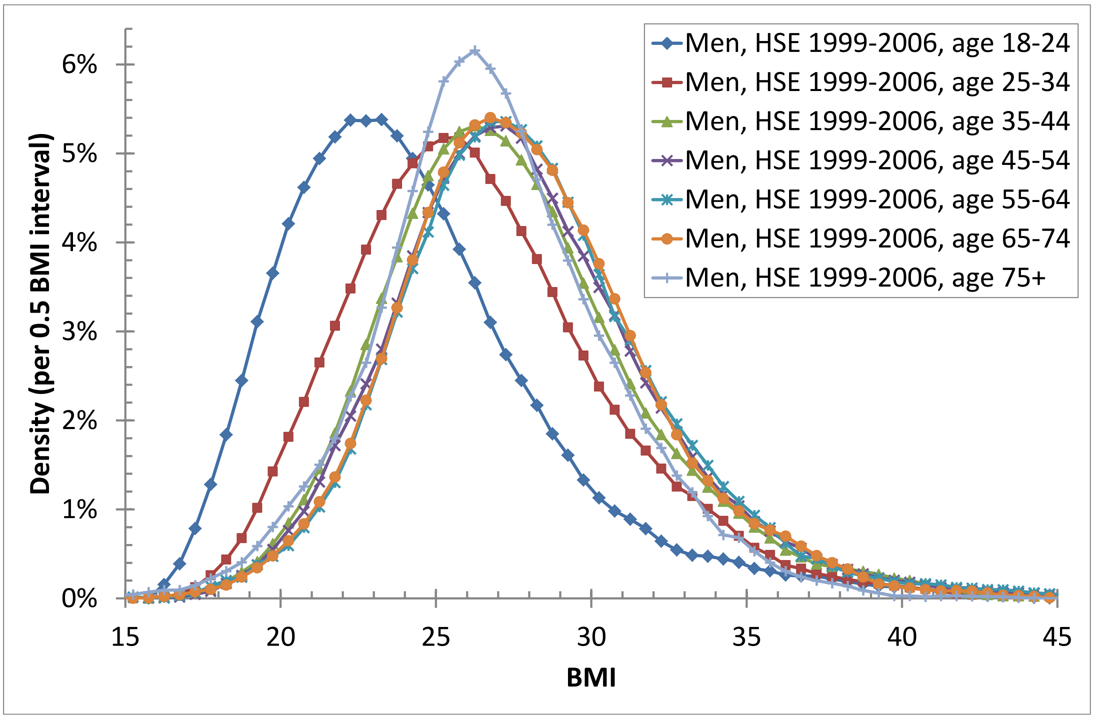


Figure 8.2 BMI distributions for different age groups for men for 1999-2006.

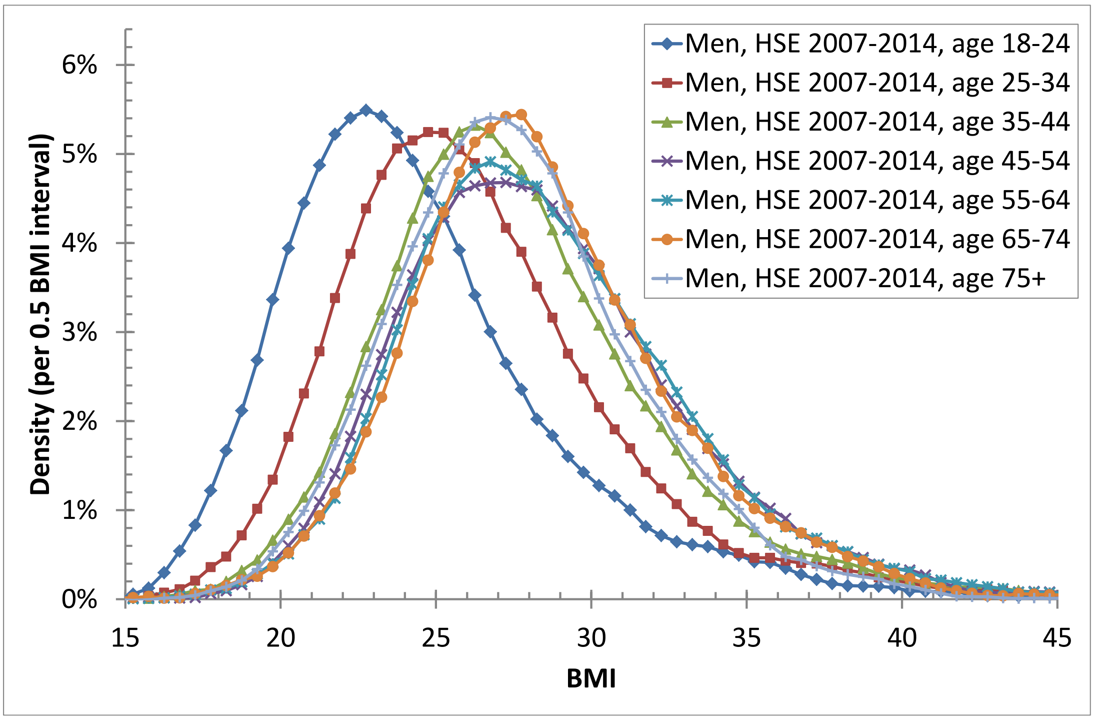


Figure 8.3 BMI distributions for different age groups for men for 2007-2014.

8.3 Variation in the age group BMI distributions by time

This section looks at how the age group distributions given in Section 8.2 have changed over the time span of the HSE. The results compare the distributions for each age group over time. This includes calculating the changes in the percentiles and modelling the changes using a scaling transformation.

8.3.1 Comparison of the distributions over time

The comparison of interest in this section is how the distribution for each age group has changed over the three HSE time intervals. Charts for example age groups 35-44 and 55-64 are shown in Figures 8.4 and 8.5. Charts for all the age groups are in Appendix 8.III. The patterns are fairly consistent across the age groups, although there is naturally some variation with the fairly small sample sizes of the data.

Looking at the charts in Figures 8.4 and 8.5 and in Appendix 8.III, BMI increases over time in a similar way to the whole population in Chapter 4, in the sense that it looks like a scaling transformation with the start of the left tail remaining in place and the rest of the distribution becoming more stretched out to the right with a lower peak and longer right tail. All age groups have a considerable change in the distribution from 1991-1998 to 1999-2006. From 1999-2006 to 2007-2014 the age groups of 35-44 and below each have little change with only the older age groups having noticeable increases in BMI. This is consistent with the analysis of mean BMI in Appendix 8.I. For age 25-34 there is actually a small decrease in BMI from 1999-2006 to 2007-2014 with the peak moving slightly left. An improvement in recent years in this age group is also identified in the previous chapter in Sections 7.4.1 and 7.6.1, and also in Appendix 8.II.

Tables and charts of the descriptive statistics for the distributions are provided in Appendix 8.III. The BMI category patterns are consistent with the general changes already described. One notable aspect is that, although the changes are generally small from 1999-2006 to 2007-2014, the severely obese category percentage increases considerably for all ages (Figure 8.III.19 in Appendix 8.III).

One difference for age 18-24 compared to the other age groups is that the left tail becomes more stretched out over time, as well as the right tail, and so the percentage of underweight men in this age group increases. The values for the three time intervals are 2.96%, 4.33%, and 4.69%, respectively. This report focusses on obesity and being overweight rather than being underweight. However, the increase in underweight men in this age group over time may also indicate a health issue and further research could look at this in more detail.

8.3.2 Percentile changes over time and scaling models

The results in Chapter 5 are that the changes over time for the population as a whole can be modelled well as a linear scaling. The analysis in this chapter has split up the data by age group. One question to consider is whether the changes over time for each age group are also approximately linear scaling transformations. It could be that the changes are a different pattern within each age group but when they are combined together in the overall population it results in a scaling pattern.

As discussed in Section 8.3.1, the changes in the age groups over time appear similar to a linear scaling transformation. To examine this further the increases in the percentiles were calculated. A linear scaling pattern above a starting BMI value will show as a straight line for the BMI increases going from 0 at the starting BMI value. The age group distributions give relative frequencies for each 0.5 BMI interval and linear interpolation was applied to calculate the percentile values, as explained in Section 4.6.1. Percentile values were calculated for each 10% value along with 1%, 5%, 95%, 99%.

Figure 8.6 shows the BMI increases for these percentiles, plotted against the original starting BMI value. These are for each age group from the distributions for the first time interval (1991-1998) to the distributions for the last time interval (2007-2014). Most of the series in Figure 8.6 are approximately a straight line with a positive slope, crossing or starting close to the x-axis , which supports a linear scaling as being a good model of the changes over time. The values are also fairly similar for the different series indicating that the changes over time are reasonably consistent across the age groups.

The two highest age groups of 65-74 and 75+ have a little bit of a different pattern with a fairly constant increase up to a BMI of about 27 and then larger increases above that. This would imply the distribution shifting up the x-axis to some extent, as well as some element of a scaling. The age groups of 45-54 and 55-64 have increases that start slightly above the x-axis and so the transformation may have a small element of a shift, whilst mainly being a scaling.

Appendix 8.IV has charts of the percentile changes between each of the three successive time intervals. As shown throughout this report, the BMI increases are much greater in the earlier years of the HSE than in the more recent time period. This is demonstrated again in the percentile values. In Figure 8.IV.2 in Appendix 8.IV, from 1991-1998 to 1999-2006 the increases for the age groups up to 55-64 show a strong and similar linear relationship, and they each look like a linear scaling. In Figure 8.IV.3, from 1999-2006 to 2007-2014 the tendency is for there to be little or no increase for the lower percentiles, with a small linear increase at higher values (above a BMI of 25 for many of the age groups). This still looks like a linear scaling but with smaller increases and a higher starting point. The younger age groups have only small changes with age 25-34 having small reductions in BMI for most percentiles. The combination of the different linear patterns in Figures 8.IV.2 and 8.IV.3 explains the slightly curved or two-piece linear pattern of some of the data in Figure 8.6.

Overall, the percentile analysis indicates that a linear scaling should be able to model the changes in the BMI distributions quite well. The linear scaling model was therefore applied for each age group for the changes from 1991-1998 to 2007-2014. This took the same approach as before (Section 5.3) of fitting the scaling starting value and the scaling slope to minimise the sum of squared differences between the fitted and target distribution. The values used are given in Table 8.1 and a chart of the BMI increases for the scaling functions is shown in Figure 8.7. The three older age groups and the all age category are plotted with dashed lines simply to help to distinguish more easily between the series.

The scaling functions BMI increases are quite similar for the different age groups and, as would be expected, are similar to the percentile values. A chart showing the model increases and the percentile values is Figure 8.IV.4 in Appendix 8.IV. The scaling models tend to match the percentiles most closely at BMI values of about 25 to 30, since this is where most of the data is for most of the age groups in 1991-1998 (Figure 8.1). The scaling model for the age group of 75+ has a very low starting value of 8.25 which means that the transformation is really a combination of a shift (translation) along the x-axis and a linear scaling.

Charts of the scaling models for age groups 35-44 and 55-64 are in Figures 8.8 and 8.9. From Table 8.1 age 35-44 has the lowest fitness function value and so Figure 8.8 is the age group with the best match of the scaling models. Age 55-64 has a larger value for the fitness value and so this scaling model is not as close a match. Charts for all the age groups are given in Appendix 8.V. They all do pretty well in modelling the changes and matching the 2007-2014 distribution for the particular age group.

Table 8.1 Values for the age group scaling models for men from 1991-1998 to 2007-2014.

Age group	18-24	25-34	35-44	45-54	55-64	65-74	75+	All
Start point	19.75	21.75	18.75	21.75	21.00	18.25	8.25	18.25
Scaling factor	1.142	1.124	1.128	1.229	1.160	1.118	1.073	1.142
Fitness value ($\times 10^{-5}$)	6.97	1.88	1.72	5.36	6.05	12.18	2.63	1.88

8.3.3 Implications of the age group BMI distribution changes with time

Chapter 7 showed the effect of age and cohort on mean BMI. The changes in time for the age groups from Section 8.3.2 can also be considered in terms of birth cohorts.

For example, consider the two distributions for age group 35-44 for 1991-1998 and 2007-2014. As in Chapter 7, the birth year can be calculated (approximately) as HSE year – age (see Section 7.2). So, the two distributions have age 35 for birth years 1956-1963 and 1972-1979, respectively. They have age 36 for 1955-1962 and 1971-1978. And so on, through to age 44 being 1947-1954 and 1963-1970. Each age is an eight year cohort and the cohorts differ by 16 years, since the HSE years are intervals of eight and differ from each other by 16. Therefore, each age is a fairly narrow cohort similar to the cohorts used in Chapter 7. Across the whole age group the range of cohort birth years obviously varies more. For 1991-1998 the birth years vary from 1947 to 1963, and for 2007-2014 from 1963 to 1979. Comparing them with the cohorts in Chapter 7, 1991-1998 is mainly cohorts 1945-1954 and 1955-1964 with the younger ages being mostly data from the later cohort and the older ages being from the earlier one. HSE 2007-2014 is mostly 1965-1974 with some data from 1975-1984 for the younger ages and some from 1955-1964 for the older ages.

The same sort of comparison applies to all the age groups. The changes in the age group distributions here therefore relate to the more recent cohorts moving up in levels in the charts in Chapter 7. Those of a given age who were born more recently have a higher mean BMI. This is due to living in a later time period.

The results here are that the increase in BMI is approximately a linear scaling transformation. The relationship between BMI and the increase in BMI over time is approximately linear. The implication is that for each age group the effect of living in a later time period in increasing BMI is greater for those with a higher BMI than for those with a lower BMI.

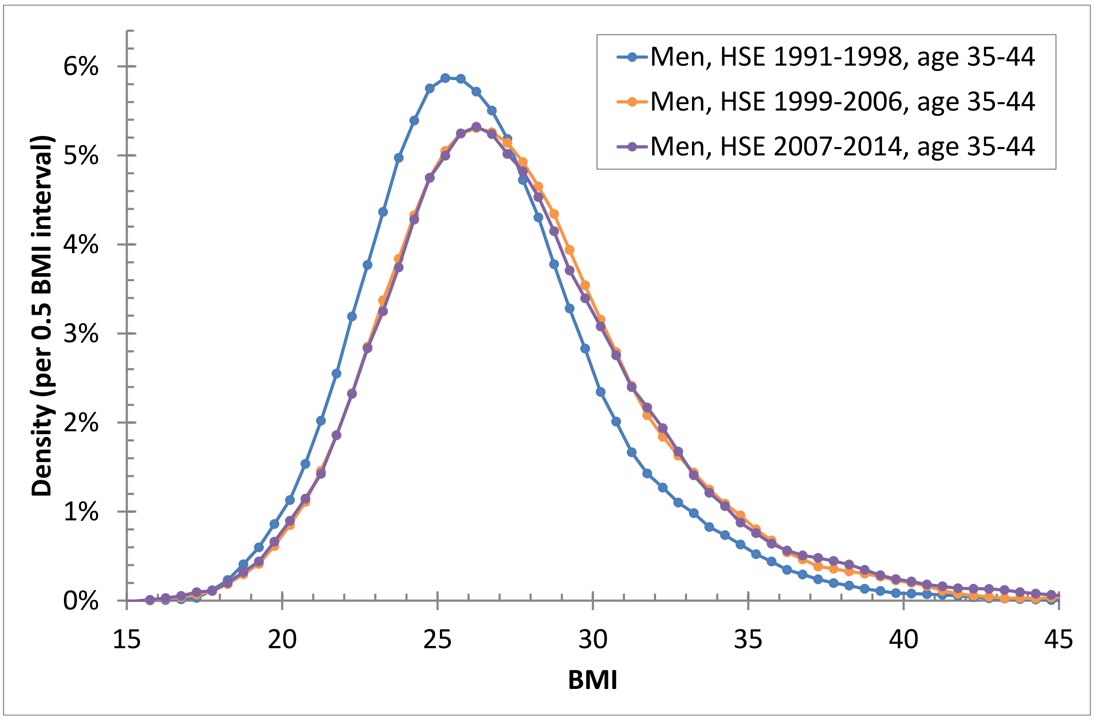


Figure 8.4 BMI distributions for men for age group 35-44.

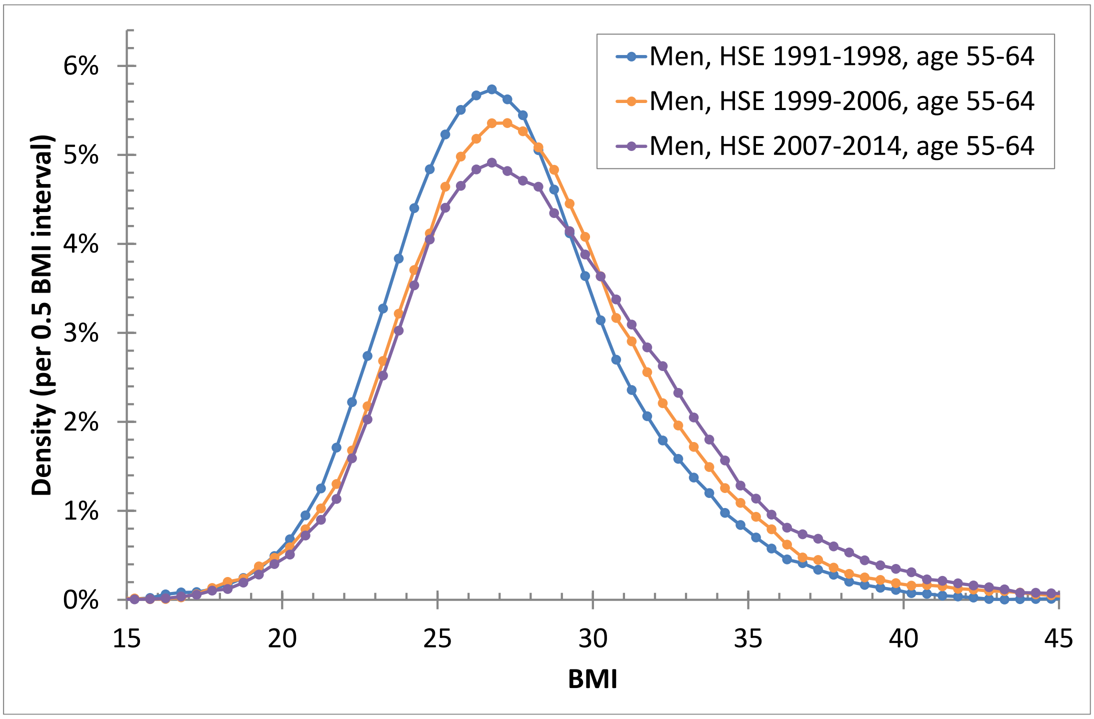


Figure 8.5 BMI distributions for men for age group 55-64.

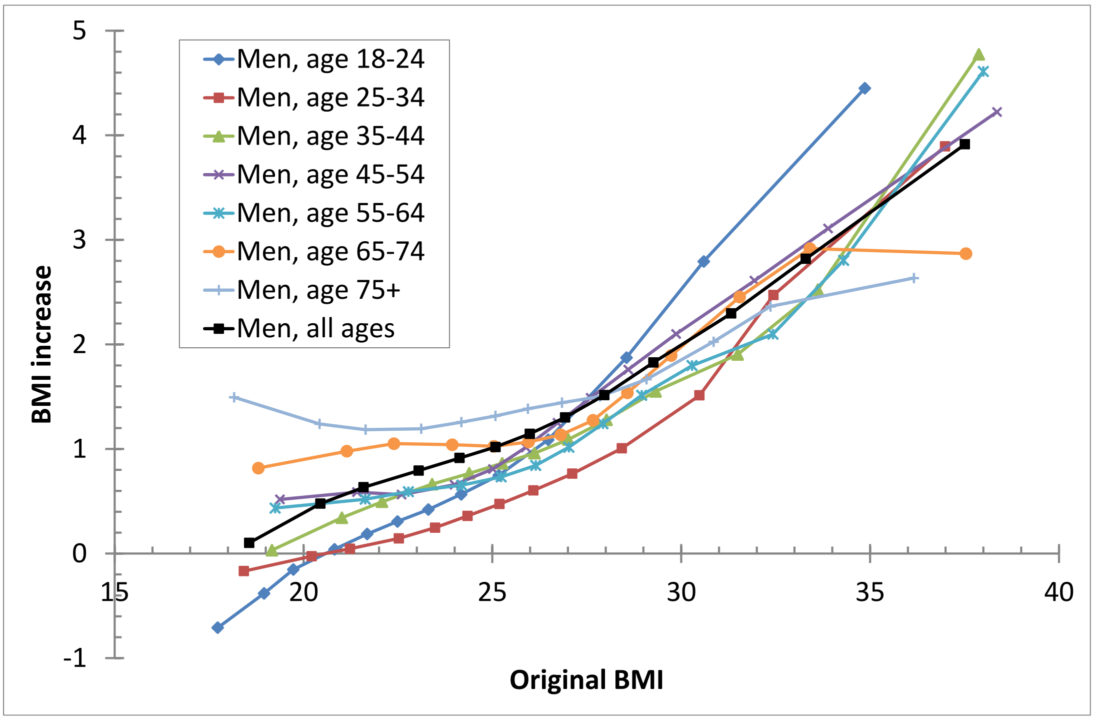


Figure 8.6 BMI increases for the age group percentiles for men from 1991-1998 to 2007-2014.

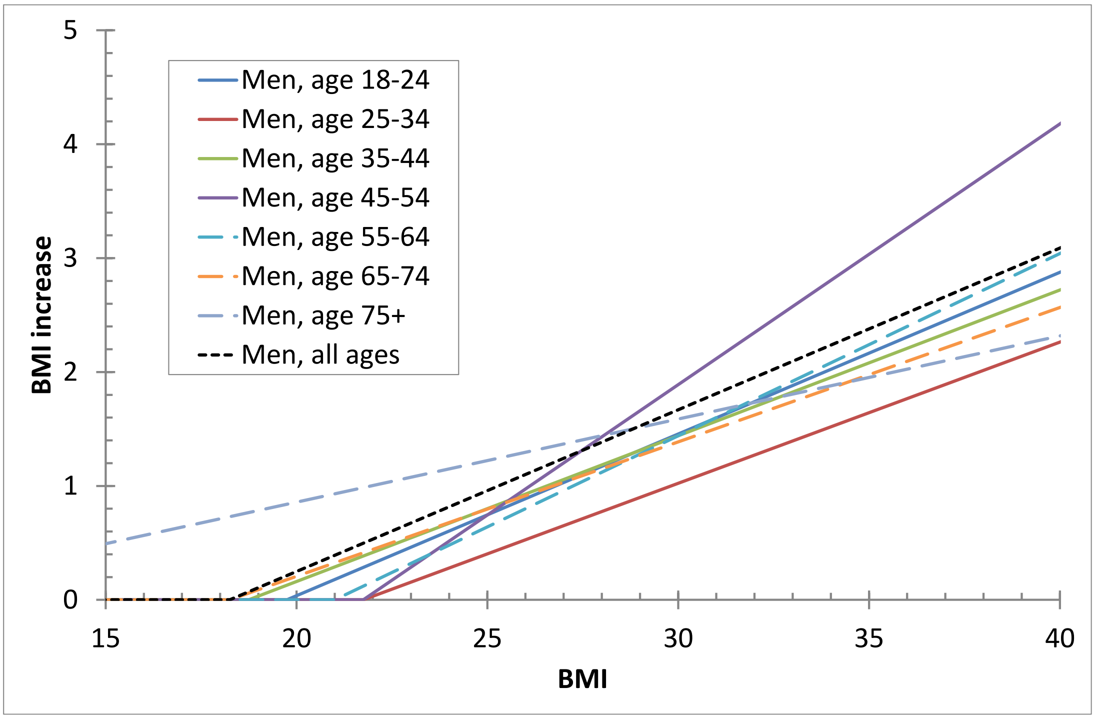


Figure 8.7 BMI increases for the age group scaling models for men from 1991-1998 to 2007-2014.

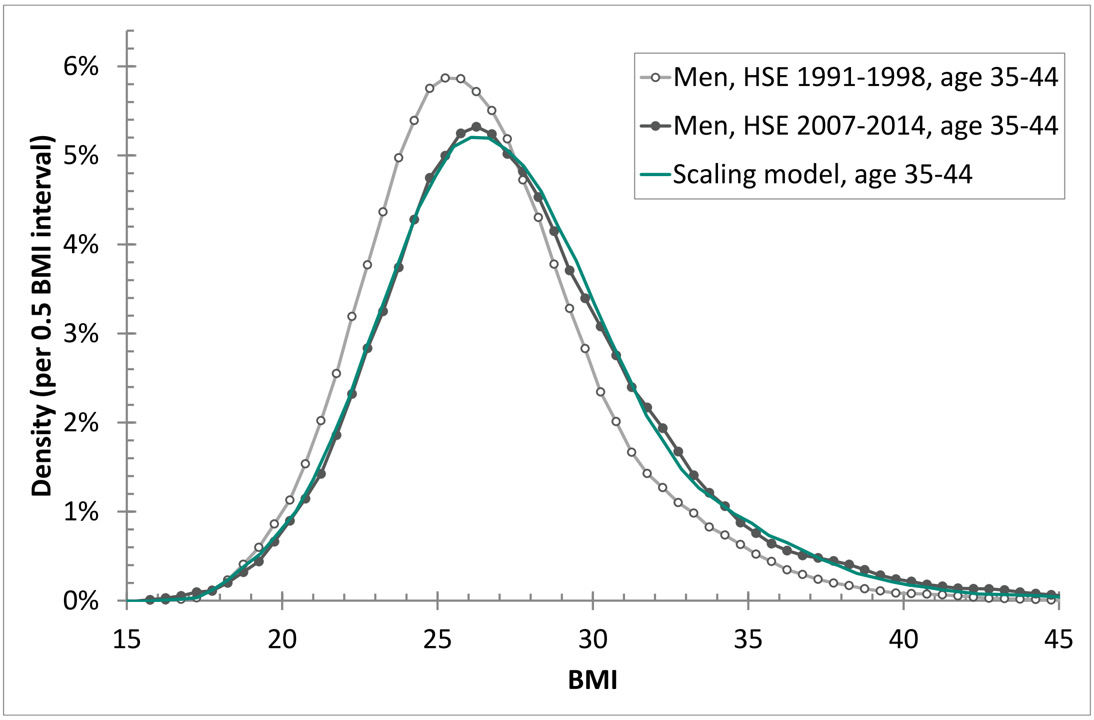


Figure 8.8 Scaling model for men for age group 35-44 from 1991-1998 to 2007-2014.

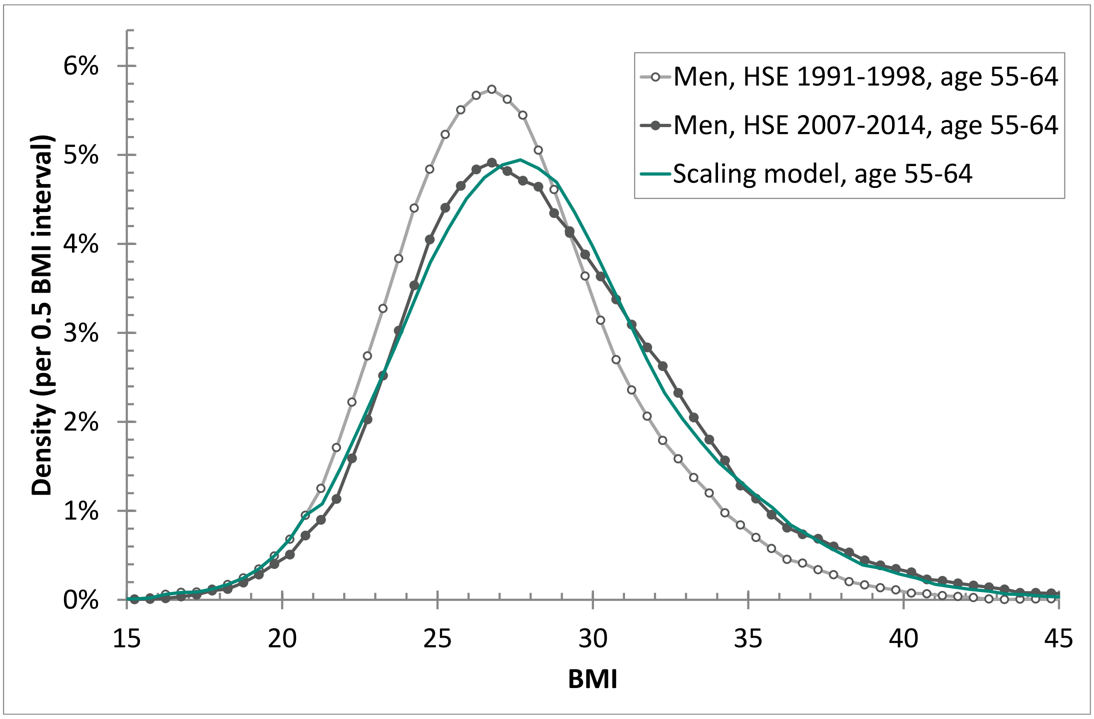


Figure 8.9 Scaling model for men for age group 55-64 from 1991-1998 to 2007-2014.

8.4 Variation in the age group BMI distributions by age

8.4.1 Comparisons of the age group BMI distributions by age

Section 8.3 shows that the age group distributions change over time approximately as a linear scaling transformation. This section looks at how the distributions change with age for a given point in time. This can be done for any of the HSE time intervals. For 1991-1998 this means comparing the distributions shown in Figure 8.1. Similarly, for 1999-2006 it is Figure 8.2, and for 2007-2014 it is Figure 8.3. In each case, the visual impression is that the shapes of all the age group distributions are much the same and that the changes are mainly just a shift (or translation) along the x-axis. For example, in each chart the peaks of the distributions are mostly of a similar height, which is one characteristic of such a shift. The statistics in Appendix 8.III also indicate the similarity. For each of the time intervals, the standard deviation varies little with age up to 55-64 (Figure 8.III.9). It is about 3.8 for 1991-1998, 4.3 for 1999-2006, and 4.6 for 2007-2014. It is a bit lower for the two higher age groups for the last two time intervals.

Focussing on the most recent time interval of 2007-2014 in Figure 8.3, from 18-24 to 25-34 to 35-44 the distribution shifts considerably to the right along the x-axis. The height of the peak of these distributions varies little with 18-24 being very slightly higher than the other two. The next two age group distributions for 45-54 and 55-64 have a slightly lower peak and are a bit broader. The oldest two distributions of 65-74 and 75+ then have a similar peak height to the youngest three distributions.

To analyse the changes further for 2007-2014, the percentiles were calculated. The changes in the percentile values compared to the previous (younger) distribution were plotted against the values in the previous distribution. The results for all the percentiles (each 10% percentile along with 1%, 5%, 95%, 99%) are shown in Figure 8.10. For example, the age 25-34 series shows the increase in each percentile value for the age 25-34 distribution compared to the age 18-24 distribution, plotted against the age 18-24 value.

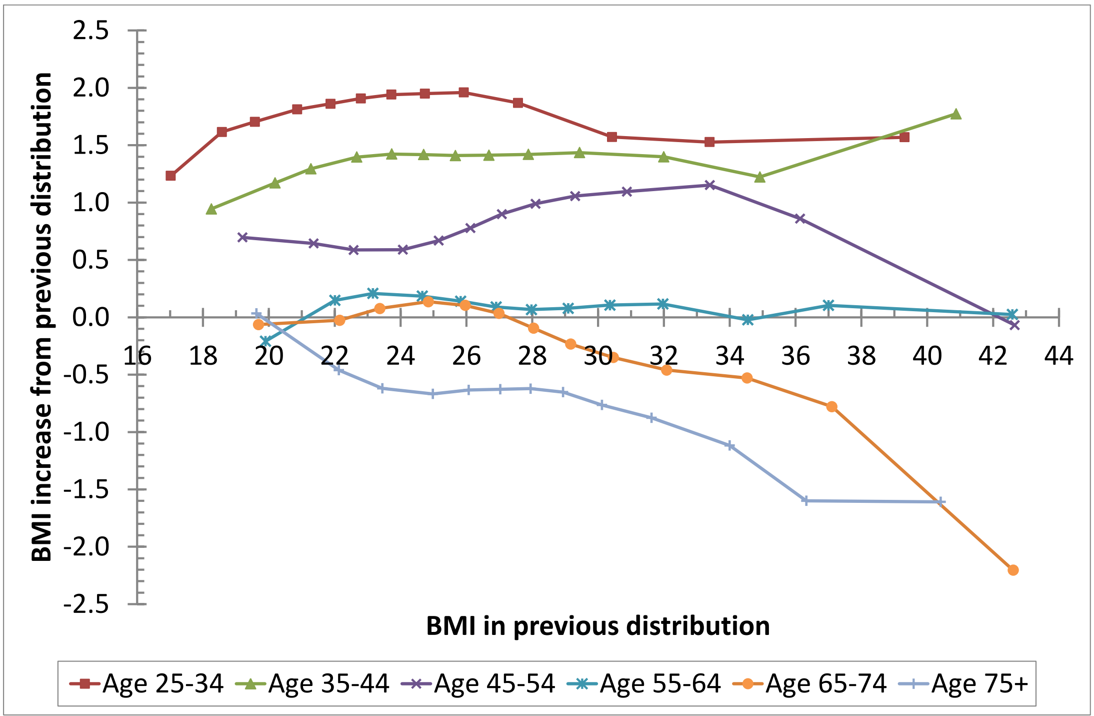


Figure 8.10 Percentile changes for the age group distributions for men for 2007-2014.

The lines in Figure 8.10 are approximately horizontal and so the increases are roughly constant, with the increases getting smaller with age (negative for older ages). This is consistent with the visual impression that the changes in the distributions are mainly a shift along the x-axis. In other words, the tendency for BMI to increase with age does not depend much on the BMI of the person. Men with a low

BMI tend to have an increase in BMI with age (on average) by a similar amount to those with a high BMI. This is very different to the linear scaling patterns observed for the changes over time.

To investigate further the interpretation of the percentile patterns, the change in the distributions was modelled as a shift for the three biggest changes from 18-24 to 25-34, 25-34 to 35-44, and 35-44 to 45-54. The shift value used was the change in mean (values of 1.804, 1.391, 0.823, respectively). This was able to model the changes well for the first two cases. The charts of these two models are in Appendix 8.V.

Linear scaling transformations were also tried and the optimum values for 18-24 to 25-34, and 25-34 to 35-44 each have a large negative starting value which means that the transformation for the range of values in the BMI distribution is mainly a shift, with only a small element of a linear scaling. Again, this supports the change with age being mainly being a shift. The distribution for 45-54 has a lower peak than for 35-44 and a linear scaling is actually a better model than a shift in this case, although the BMI changes are fairly small.

The peak of the distribution for age 65-74 is higher than for 45-54 and 55-64 and is similar to that of age 35-44. Transformations were calculated between the 35-44 and 65-74 distributions and showed that the difference is mainly a shift.

The magnitude of the differences in the distributions reflect the aging profile identified in Chapter 7 and in Appendix 8.II from the analysis of the mean BMI values. The rate of increase in BMI gets less with age and so the differences in Figure 8.10 get less with each age group. From Chapter 7 some of the older age groups are at lower levels and so BMI reduces, hence the negative values in Figure 8.10.

Since the HSE time interval is fixed, the different ages are from different cohorts. However, the hypothesis from Chapter 7 is that the younger age groups are close to the aging profile of current conditions. If that is the case, the changes from one age group to the next for the younger age groups are mainly the current aging effect. For older age groups it is a combination of the aging effect and the older cohorts being at lower levels due to different conditions experienced earlier in life.

8.4.2 Implications of the age group BMI distribution changes with age

The main conclusion from this analysis is that the BMI distribution for men changes with age in a different way to the changes with time. Changes with age tend to be mainly a constant increase of a shift or translation along the x-axis compared to the linear scaling increases seen with the changes with time. Therefore, the tendency to increase BMI with age has little association with the BMI level. Hence, the indication is that to a large extent the effect of aging on BMI is independent of the BMI level.

8.5 Single year of age BMI distributions

Some analysis was done looking at data for single years of age. Single years are, of course, still an age category but are of range one year rather than the seven years for 18-24 or the ten years for 25-34, 35-44, etc. From Chapter 7, BMI increases with age and the increases are the most rapid at the youngest ages for the ages of 18+ considered here. The age groups in Sections 8.2, 8.3, and 8.4 therefore combine varying BMI levels. Consequently, it was considered useful to look at some of the BMI distributions for just a single year of age. The disadvantage, of course, is that the sample sizes are much smaller.

The work throughout this report shows that BMI changes much more in the 1990s than in the years since then. For example, Figure 8.I.1 in Appendix 8.I shows that mean BMI changes little from 1999-2006 to 2007-2014 for ages up to about 45. Figures 8.VII.9 and 8.VII.10 in Appendix 8.VII show the similarity in mean and healthy category percentages for ages 18 to 24 between 1999-2006 and 2007-2014. Therefore, it was decided to combine these two time intervals to get bigger samples. Hence the single year distributions were produced for the HSE data for 1991-1998 and 1999-2014. The process was the same as set out in Section 8.2, grouping the data in 0.5 BMI intervals and smoothing it with the 13 point Savitzky-Golay method.

Section 8.5.1 describes analysis looking at the single year distributions in 6 year gaps covering ages 18-54. This range of ages was chosen to focus on the ages where BMI increases the most and where the BMI level in recent years is close to the proposed aging model in Chapter 7. From Chapter 7 and Figure 8.I.1, older ages than this are at lower levels due to being older cohorts and their BMI is lower presumably due to different experiences earlier in life. Ages with 6 years gaps (18, 24, 30, etc.) were used as this helps see the changes with age more clearly and smooths out some of the irregularities from one year of age to the next.

In the adult ages considered in this report, age 18 is the healthiest in terms of the BMI statistics with the lowest mean BMI and the highest proportion in the healthy range. BMI increases rapidly from age 18 and so Section 8.5.2 looks at the changes with each year of age from 18 to 24.

8.5.1 Ages 18-54 in 6 year gaps

The distributions for single years over an age range from 18 to 54 were calculated and analysed. This was done using both 3 year and 6 year gaps, and the patterns are similar. The results with 6 year gaps are presented here. The distributions for these ages for women are presented in Chapter 9 in Section 9.5.1, and the distributions for men and women are compared in Chapter 10.

The smoothed distributions for 1999-2014 are shown in Figure 8.11. Sample sizes vary from 561 to 846 for 1991-1998, and from 712 to 1113 for 1999-2014. Even though the sample sizes are small, the distributions produced are fairly smooth albeit with some irregularities. The full analysis and statistics for 1991-1998 and 1999-2014 are given in Appendix 8.VI.

Figure 8.11, along with the tables and charts of statistics in Appendix 8.VI, show the particularly big changes from age 18 to 24 to 30 to 36. Considering some of the statistics for 1999-2014, the mean values are: age 18: 23.24, age 24: 25.14, age 30: 26.36, age 36: 27.26. The healthy category percentage reduces from 67.4% at age 18 to just 31.3% at age 36 with the values for ages 18, 24, 30, 36 being: 67.4%, 52.4%, 40.7%, 31.3%. At age 54 the healthy percentage is only 23.4% and the mean is 28.21. The size of the changes with age are similar for 1991-1998 and 1999-2014, with BMI being higher for all ages in 1999-2014.

One focus with this analysis was to compare the patterns with those of the broader age groups in Sections 8.2-8.4 to check that they are similar. From the results in Appendix 8.VI this is generally the case. In particular, the percentile patterns are similar with the changes with time looking mainly like a linear scaling (Figure 8.VI.3), and the changes with age looking mainly like a shift along the x-axis (Figure 8.VI.4).

Scaling models were applied to model the changes over time for each age from 1991-1998 to 1999-2014. As would be expected from the percentiles in Figure 8.VI.3 the scaling transformations are able to model the changes in the distributions quite well and the BMI increases for the models are similar to the percentiles. The irregularities in the distributions limit how good the match can be, and the fitness function values vary from 2.76 × 10^{-5} to 13.54 × 10^{-5}. By comparison, the fitness values for the models in Chapter 5 vary from about 3 to 12 × 10^{-6} and so these values are higher by a factor of about 10. The detailed results are not included given the wide range of other results already presented in the report and in the appendices.

One of the main benefits in looking at the single year ages is that it shows the extent of the difference between age 18 and the other ages. For example, in Figure 8.11, there is a noticeable difference in the shapes with the age 18 distribution having a much higher and narrower peak than age 24 and older ages. The age 18 distribution has a lower standard deviation and a higher skewness than the distributions for the other ages. The percentiles also indicate some element of a change in shape from age 18 to 24 (Figure 8.VI.4).

Transformations were applied to the changes with age for the distributions in Figure 8.11 for 1999-2014. With the difference in shape from age 18 to 24, a linear scaling transformation works much better than a shift for the increase in BMI between these ages. However, for the differences with age above that, a shift along the x-axis (i.e., a translation) is a good model and, in most cases, the optimal linear scaling function has a large negative starting value so that it is also mainly a shift.

The quartiles for each age, the quartile increases, and the total BMI increase for the percentiles from age 24 are shown in Appendix 8.VI in Figures 8.VI.18-20. The patterns again indicate a shift transformation above age 24. A shift model was therefore applied to the total change from age 24 to each of the higher ages and works well. When a scaling model was tried this results in negative starting values, again supporting a shift as the main component of the transformation. The model for age 24 to 54 is shown in Figure 8.12 (and Figure 8.VI.21). The shift amount used is the difference in means of 3.070, and the fitness value for the model is 13.81 × 10^{-5}.

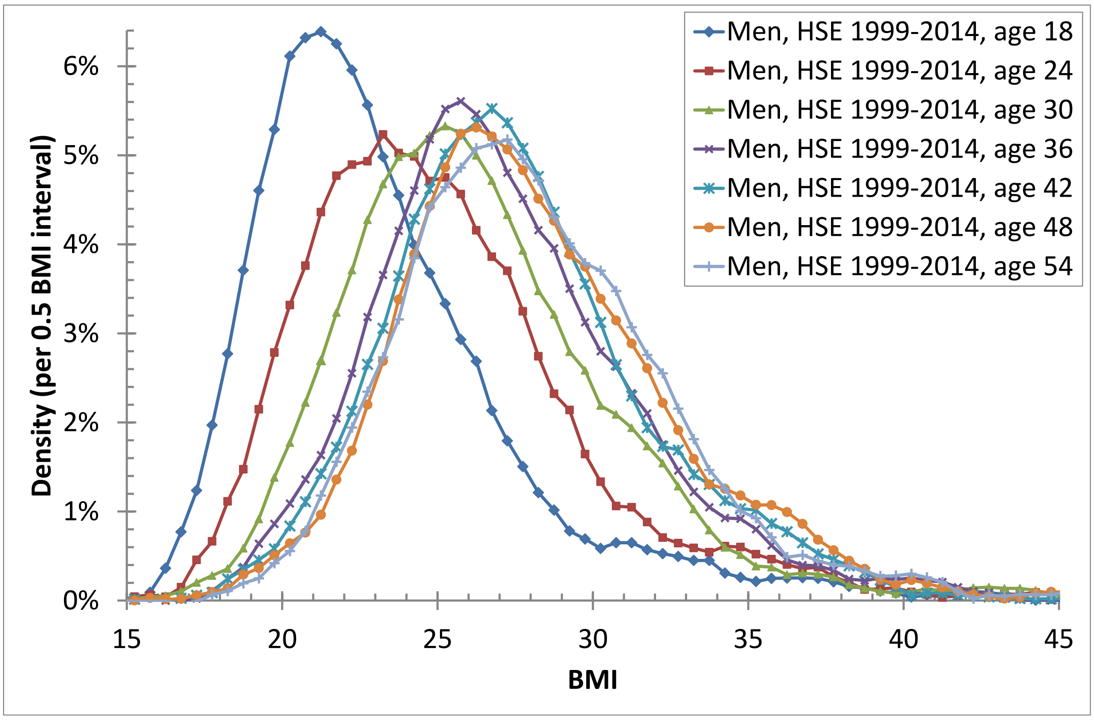


Figure 8.11 Distributions for ages in 6 years gaps from 18 to 54 for men for 1999-2014.

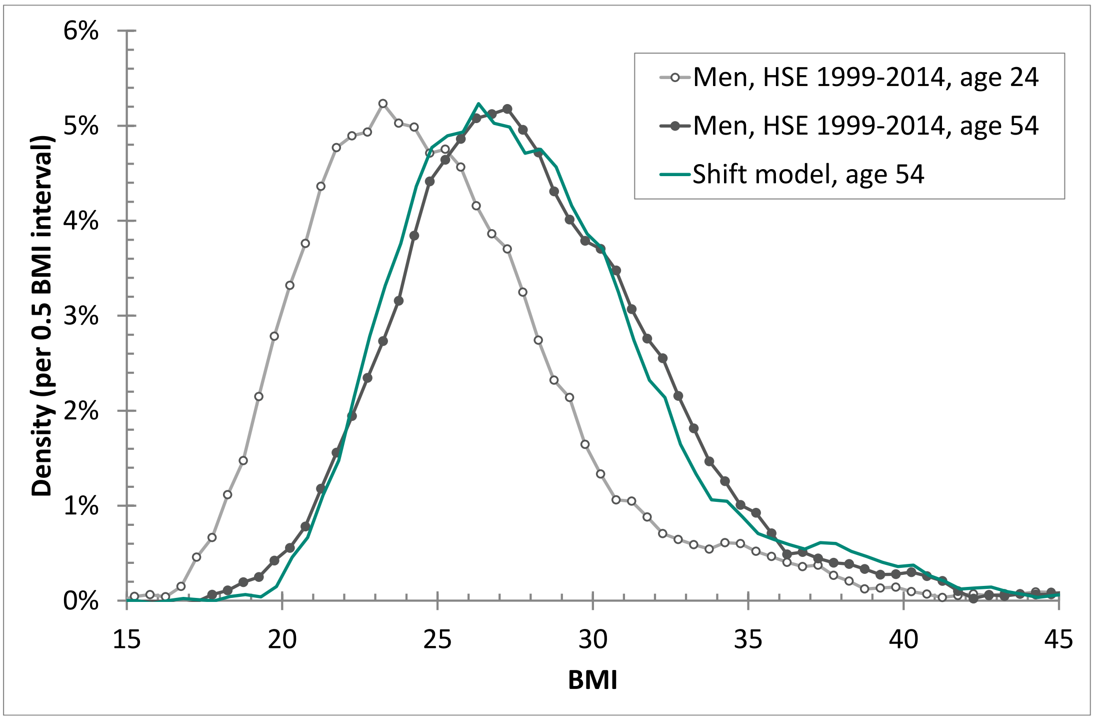


Figure 8.12 Shift model for the change in the BMI distribution from age 24 to 54.

8.5.2 Ages 18-24 for each year of age

Age 18 has the lowest BMI and the highest proportion in the healthy category. The previous analysis in this chapter and in Chapter 7 shows that BMI increases rapidly as age increases from age 18. In particular, Section 8.5.1 demonstrates the big change between the distributions for age 18 and age 24. It is therefore of interest to look in more detail at how BMI changes from age 18 with each year of age.

Statistics and smoothed BMI distributions were produced for each year of age from 18 to 24. The HSE time intervals of 1991-1998 and 1999-2014 were used, as in Section 8.5.1. Figure 8.13 plots the mean and median values. This shows strong trends of both an increase in BMI with age, and an increase in BMI over time from 1991-1998 to 1999-2014. The increases in mean and median BMI from age 18 to 24 are 1.96 and 2.18 for 1991-1998, and 1.90 and 2.12 for 1999-2014. The average increase for the seven ages between 1991-1998 and 1999-2014 is 0.61 for the mean and 0.37 for the median. Detailed results for both time intervals are in Appendix 8.VII.

The main interest in this section is in the current effect of age on BMI and so the focus is on the 1999-2014 results. Figure 8.14 shows the smoothed distributions for 1999-2014. The smoothed distributions naturally have some irregularities with the small sample sizes but there is a strong pattern of BMI increasing with each year of age. There is a visual difference in the distributions for age 18 and the older ages. The difference in shape for age 18 compared to age 24 and older ages is also noted in Section 8.5.1. Figure 8.14 shows that the distribution for age 18 has a noticeably different shape even to age 19, with a higher narrower peak. The shapes for the older ages are quite similar to each other, although age 24 has a lower peak than the others.

The percentiles are plotted in Appendix 8.VII. These have similar patterns to those discussed before in this chapter with the changes over time having a pattern like a linear scaling and the changes with age looking mainly like a shift along the x-axis. The change from age 18 to age 19 has a pattern indicating some element of a change in shape as well as a shift.

From the statistics in Appendix 8.VII, the distribution for age 18 is pretty healthy in reference to the standard categories: for 1999-2014 the mean BMI is 23.2 with 67.4% in the healthy BMI category. Only 7.4% are in one of the obese categories and there is actually a similar percentage who are underweight at 7.1%. By age 24 the mean BMI has increased to 25.1 with now only 52.4% in the healthy category, and with 11.9% in one of the obese categories. Only 2.5% are in the underweight category at this age.

The statistics are even better for 1991-1998. The age 18 statistics for this time interval are a mean BMI of 22.6 and 77.6% in the healthy category. Just 4.1% are in an obese category with 5.3% in the underweight category. For age 24 the mean BMI is 24.6 with 57.7% in the healthy category. There are 8.8% in an obese category and 1.3% in the underweight category.

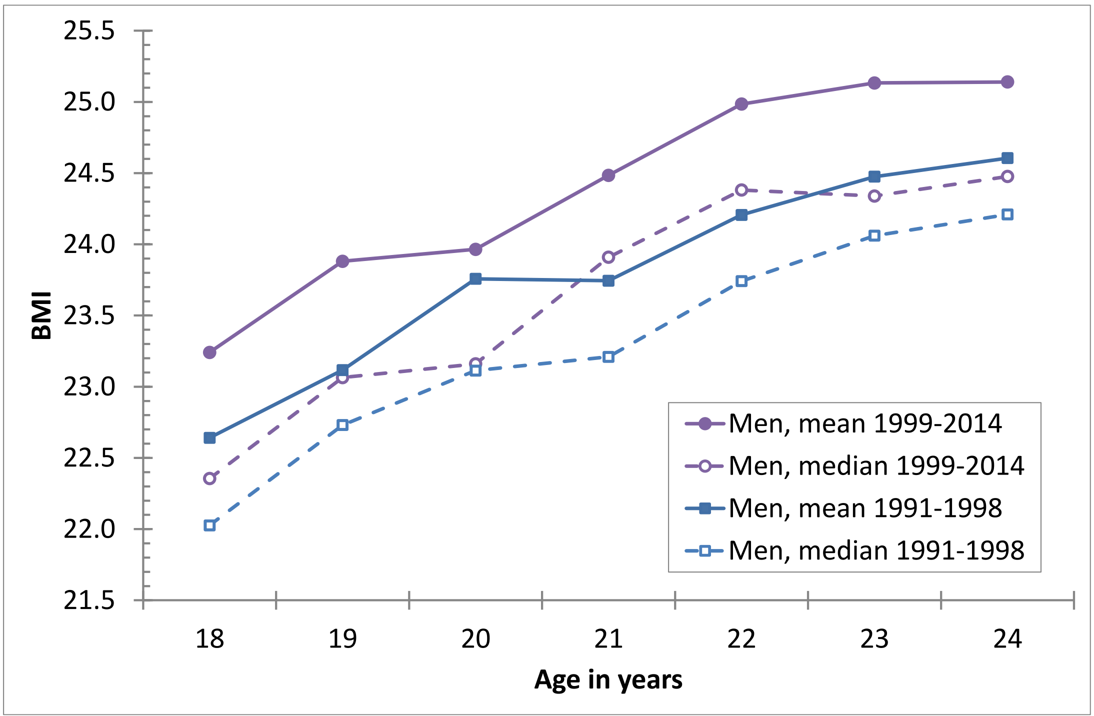


Figure 8.13 Mean and median for each year of age for men from 18 to 24.

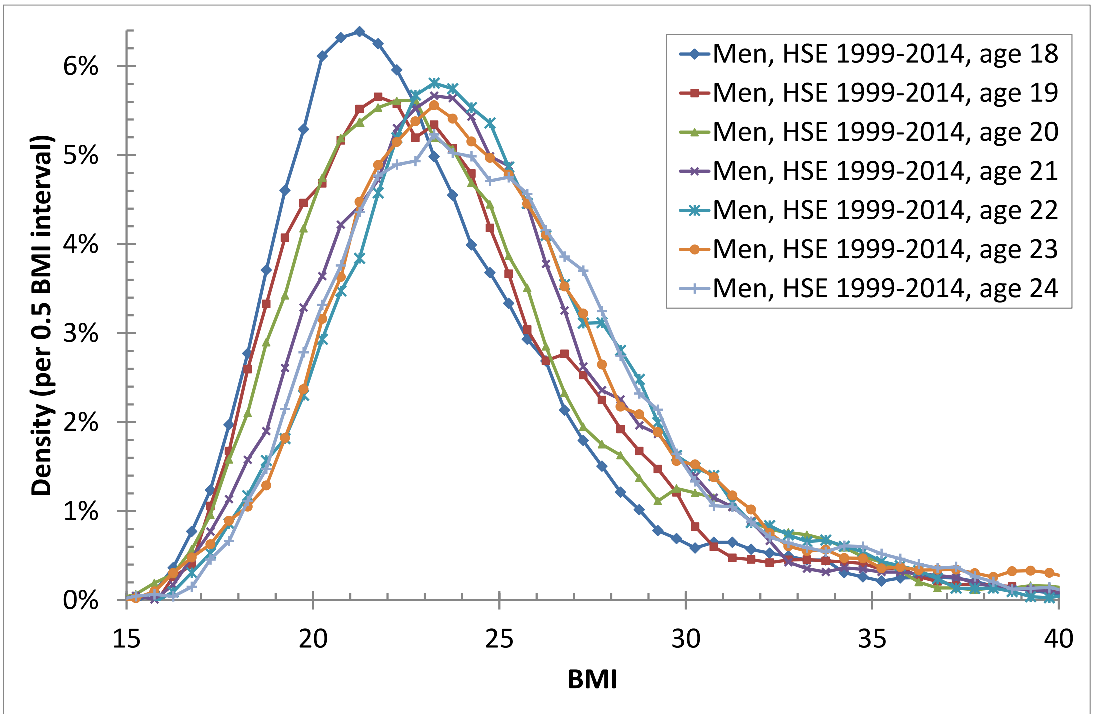


Figure 8.14 Smoothed distributions for each year of age for men from 18 to 24 for 1999-2014.

8.5.3 Implications of the single year of age BMI distributions

Age 18 has the lowest mean BMI and the highest healthy category percentage for the adult ages considered in this report. The increase in BMI from age 18 to 24 is considerable. I am not sure to what extent the increase in BMI over this age range is simply a result of the natural development and maturing of the body, and to what extent it is due to lifestyle effects and changes in circumstances and behaviour. This is an age range over which lifestyles can change a lot with factors such as moving from school to university or employment, and leaving the parental home. There can be changes in eating and drinking behaviours such as starting to do their own cooking or starting to consume alcoholic drinks.

The analysis here may indicate an important opportunity to target interventions at those of age 18 to try and reduce the BMI increases with age: for example, perhaps through information and education on healthy eating, nutrition, and cooking skills. This age group may have had some information on this during childhood at home or at school. However, health messages could particularly try and reach this group, and school or further education institutions could potentially develop beneficial initiatives. Developing skills and healthy eating habits at this age could be very beneficial throughout life. It might be easier for people to maintain their BMI or limit the increases from age 18 than it would be to reduce their BMI when they are older.

There is also a significant proportion in the underweight category at age 18, although this reduces considerably by age 24. Again, this could be due to natural body development. There are more underweight at each age from 18 to 24 in 1999-2014 compared to 1991-1998. This may indicate a general public health issue, although it is not the main focus of this report. This issue could be included in any health initiatives aimed at 18 year olds.

Chapter 9: Detailed analysis by age for women

Key points from this chapter:

- This chapter covers the same analysis as in Chapter 8, but using the data for women. The main focus is to look at the BMI distributions for different ages for women. Distributions were produced for age groups of about 10 years, and for single years of age. They were compared over time as to how they change for different HSE time intervals, and by age as to how they change as age increases for a particular HSE time interval.
- As was the case for the populations for men and for women as a whole (Chapter 5) and for the age distributions for men (Chapter 8), the changes in the BMI distributions over time are approximately a linear scaling transformation. This implies that the changes in obesity factors over time have tended to have a greater effect on those with higher BMI than on those with a lower BMI.
- The changes in the distributions with age are mainly a linear scaling transformation, but with some element of a shift along the x-axis. This implies a general tendency for all levels of BMI to increase in BMI with age, but with greater increases for those with higher BMI.
- BMI increases with age and so the best statistics for the adult data considered in this report are for age 18. For the 1999-2014 data, the age 18 distribution has 63.7% in the healthy BMI category with a mean of 23.81. For 1991-1998, the healthy category percentage is 69.7% with a mean of 23.02. As also discussed in Chapter 8, a focus on age 18 for health messages and interventions could therefore be a good strategy.
- The age 18 distributions are noticeably very similar for men and women, but the distributions are quite different for older ages, and this is examined further in Chapter 10.

9.1 Overview

This chapter considers the BMI age effects in the data for women. It looks in detail at the BMI distributions for different ages and how they change by both age and over time.

The same approach was taken as described in Chapter 8. BMI distributions were derived for age groups of mostly 10 years for the HSE data for 1991-1998, 1999-2006, and 2007-2014. The distributions obtained are presented in Section 9.2. The patterns of the changes over time from 1991-1998 to 2007-2014 were examined using percentiles and transformations. The results for this are in Section 9.3. The changes with age from one age group to the next were investigated in a similar way and this is set out in Section 9.4.

Distributions were also produced for single years of age, and this work is described in Section 9.5. The samples sizes are smaller than for the age groups, but these are narrower age categories and so in this respect give finer detail of how BMI varies with age. To make the sample sizes a bit larger the last two HSE intervals for the age groups were combined, and so the 1999-2014 data was used as the most recent time interval. Distributions for the ages of 18 to 54 in gaps of 6 years were derived first. Then, the youngest ages were investigated by looking at each year of age from 18 to 24. Again, the distributions were compared as to how they change over time and how they change with age.

The same data was used for the analysis here as in Chapter 7. This is the actual HSE case data using the interview weights, but with the case weights adjusted so that the average case weight for each age in each HSE year equals 1.

Chapter 7 looks at how mean BMI varies with age by cohort. BMI increases with age within the cohorts, particularly for younger ages. There is also a pattern of more recent cohorts being at a higher level than older cohorts. In an initial analysis, the mean BMI for each year of age was calculated for the data for women using time intervals of eight years. The results are in Appendix 9.I. The chart of mean BMI along with the proposed aging model from Section 7.3 is given in the appendix in Figure 9.I.1. This is limited in understanding the full relationship with age as it does not take account of cohort effects, which are analysed in Chapter 7. However, as discussed in Chapter 7, the hypothesis is that the cohorts over time are moving towards the aging model curve. Hence, the mean BMI values get closer to the aging model in each successive time interval with the increases getting less over time. The increase is generally about 0.8 from 1991-1998 to 1999-2006 and 0.3 from 1999-2006 to 2007-2014, and similar across all the ages (Figure 9.I.2). The chart of curves for four year intervals is also given in Figure 9.I.3 in Appendix 9.I.

9.2 BMI distributions by age group

The smoothed BMI distributions were produced for each age group for women. The same method was used as described in Section 8.2, with the relative frequencies calculated in 0.5 BMI intervals and then smoothed using the 13 point Savitzky-Golay method. Appendix 9.II gives some examples of the smoothing. The resulting distributions are the estimates of the population BMI distributions.

The age group distributions for 1991-1998, 1999-2006, 2007-2014 are shown in Figures 9.1-9.3. As age increases there are fairly steady changes of both the curve moving to the right and the curve flattening out. So, the peak gets lower and distribution becomes less skewed. This is a bit different to the pattern for men where the age group distributions are more similar in shape with the main change being the shifting of the distribution along the x-axis. The changes with age are examined further in Section 9.4.

Charts comparing the distributions for each age group for the three time periods are in Appendix 9.III. In each case the distribution tends to change much more from 1991-1998 to 1999-2006 than from 1999-2006 to 2007-2014. The changes over time are discussed in Section 9.3.

The descriptive statistics and category percentages along with some charts of the statistics are also in Appendix 9.III. One notable aspect is the consistent decrease in skewness as age increases reflecting the changing shape of the distribution with age. The age 75+ distribution and statistics are considerably better than 64-74 which, from Chapter 7, is due to this group being older birth cohorts having a mean BMI aging curve at a lower level.

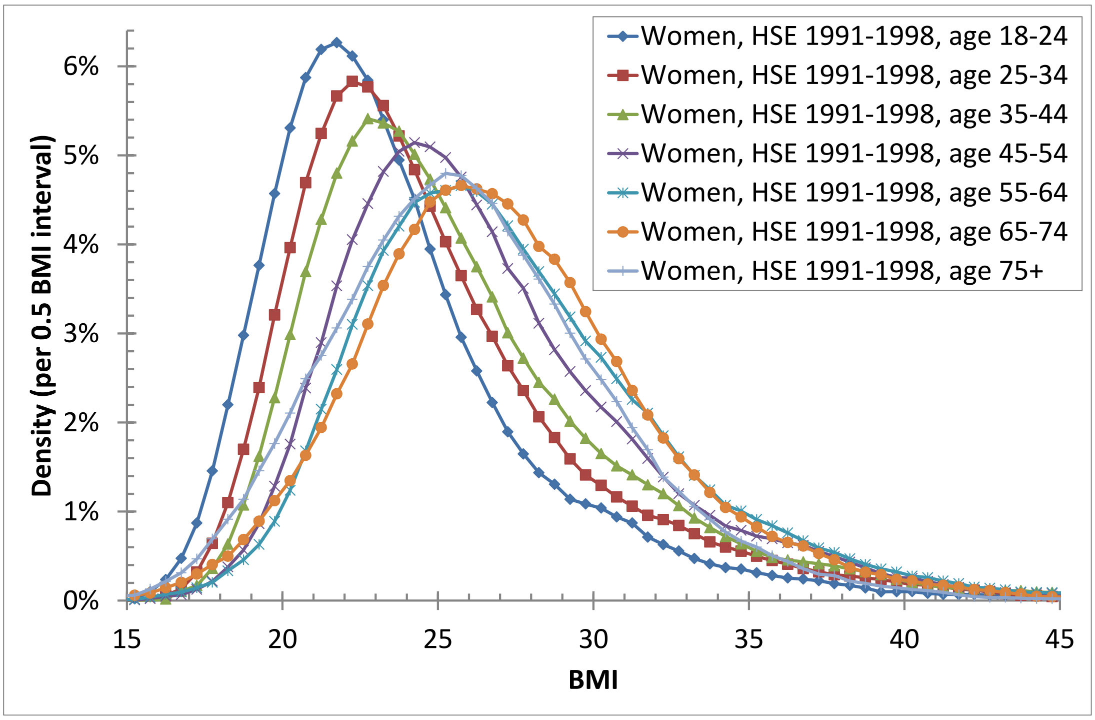


Figure 9.1 BMI distributions for different age groups for women for 1991-1998.

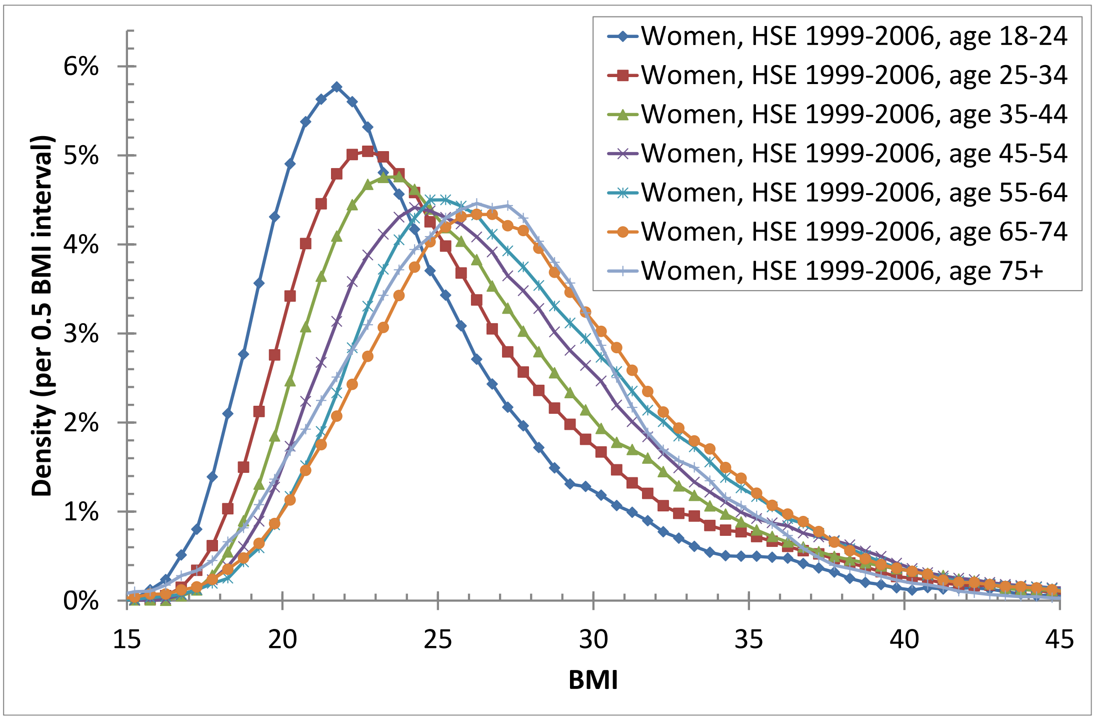


Figure 9.2 BMI distributions for different age groups for women for 1999-2006.

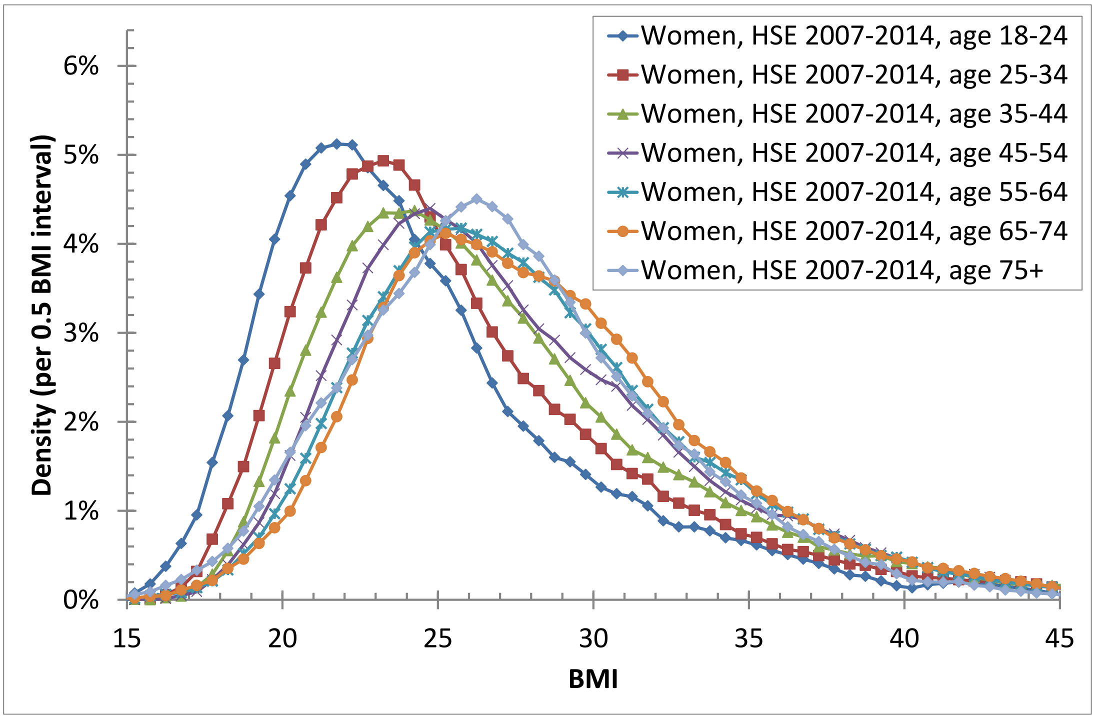


Figure 9.3 BMI distributions for different age groups for women for 2007-2014.

9.3 Variation in the age group BMI distributions by time

The changes over time for both men and women for the population as a whole are approximately a linear scaling (Chapter 5). This is also the case for the age groups for men (Section 8.3). The analysis here looks at whether this applies as well for each age group for women using the distributions in Section 9.2.

9.3.1 Comparison of the distributions over time

The charts comparing each age group for the three time intervals are in Appendix 9.III (Figures 9.III.1-9.III.7). As examples, the charts for age groups 35-44 and 55-64 are given in Figures 9.4 and 9.5. There are some variations and irregularities but the main patterns are pretty consistent across the age groups. Over time the peak of the distribution gets lower, the right tail extends, and the left tails stays much the same. The differences are greater from 1991-1998 to 1999-2006 than from 1999-2006 to 2007-2014.

Tables and charts of the descriptive statistics for the distributions are provided in Appendix 9.III. The BMI category patterns are consistent with the general changes already described. Some of the main changes over time for each age group are that the mean increases, the standard deviation increases as the right tail extends further, the healthy category percentage decreases, and the obese category percentages increase. The changes are generally small from 1999-2006 to 2007-2014, although in relative terms there is a considerable increase in the severely obese category percentage for most ages (Figure 9.III.19 in Appendix 9.III). One point that is different to the patterns for men is that the underweight percentage in age 18-24 does not change much over time whereas it increases considerably for men.

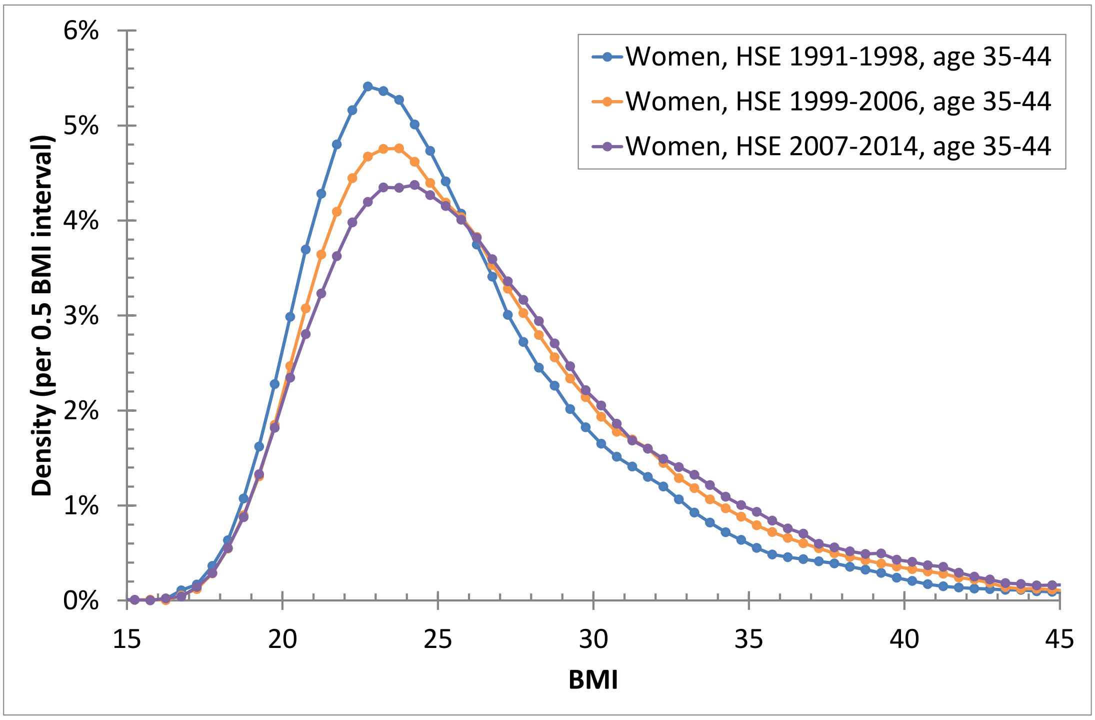


Figure 9.4 BMI distributions for women age 35-44.

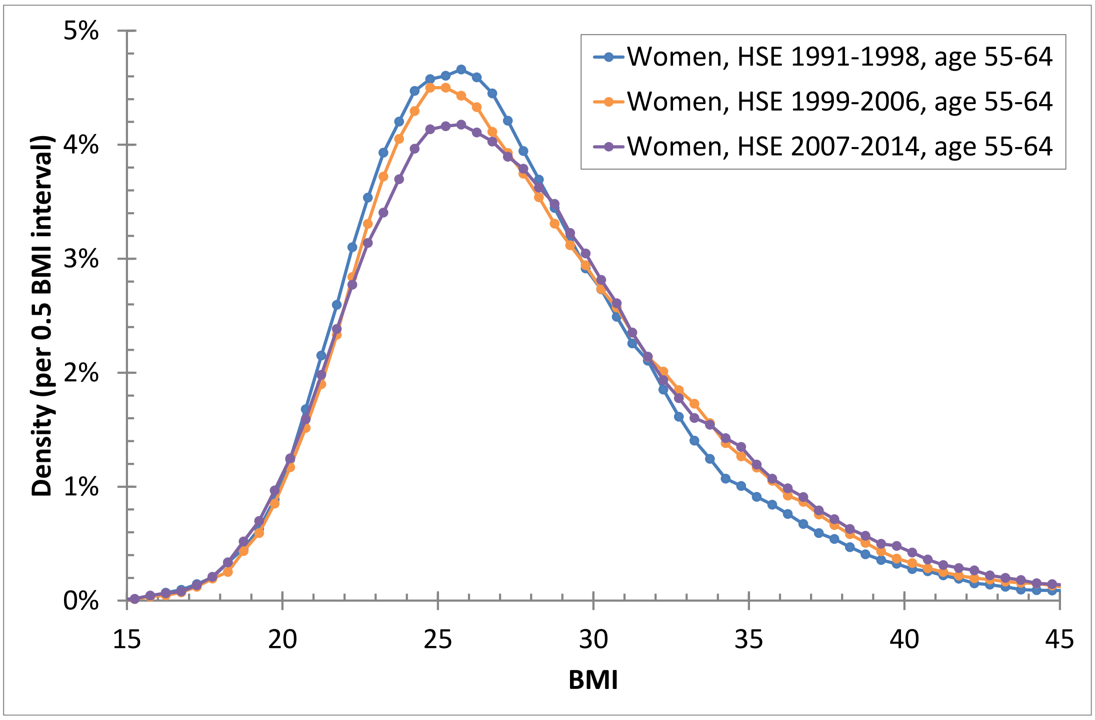


Figure 9.5 BMI distributions for women age 55-64.

9.3.2 Percentile changes over time and scaling models

To look at the pattern of the changes, percentiles were calculated and the increase from one time period to the next obtained. As for other data in this report, the percentile values were calculated for the 10% percentiles along with 1%, 5%, 95%, and 99%. Figure 9.6 shows the increases in the percentile values from 1991-1998 to 2007-2014 for all the percentiles. Most age groups show a similar approximately linear pattern that supports a linear scaling transformation as being a good model of the changes. A couple of the older groups have some slightly different features. Age 55-64 has smaller increases than the other groups. Age 65-74 has a linear type of pattern for higher BMI, but more of a constant increase at lower BMI levels. Appendix 9.IV has charts of the percentile changes between each time interval. As discussed elsewhere in this report, the BMI increases are much greater in the earlier years of the HSE than in more recent times. The percentile increases are consistent with this, being much larger from 1991-1998 to 1999-2006 than from 1999-2006 to 2007-2014.

Overall, the percentile analysis indicates that a linear scaling transformation should be able to model the changes over time in the BMI distributions pretty well, as is the case for the overall populations (Chapter 5) and for the age groups for men (Section 8.3.2). The linear scaling model was therefore applied to each age group for the changes from 1991-1998 to 2007-2014. This took the same approach as before (Sections 5.3 and 8.3.2) of fitting the scaling starting value and the scaling factor to minimise the sum of the squared differences between the fitted distribution and the target distribution.

A chart of the scaling functions is shown in Figure 9.7. The three older age groups and the all ages series are plotted on Figure 9.7 with dashed lines just to help to distinguish more easily between the series. The scaling functions are quite similar to the percentile values in Figure 9.6, as would be expected, and Figure 9.IV.4 in Appendix 9.IV shows the functions and the percentile values on the same chart. Charts of the scaling model distributions for age groups 35-44 and 55-64 are in Figures 9.8 and 9.9. The charts of all the scaled model distributions are in Appendix 9.V. The models do pretty well in modelling the changes. The scaling model for age group 65-74, where the percentile increases are further from being linear, does not fit quite as well as the other age groups.

The values used for the models and the fitness values are given in Table 9.1. From Table 9.1 age group 35-44 has the lowest fitness function value apart from the all ages data, and so Figure 9.8 is the age group with the best match of the scaling models. Age group 55-64 in Figure 9.9 also has a low fitness value. The fitness values overall are lower (better) than for the data for men and the scaling transformations all do well in modelling the changes.

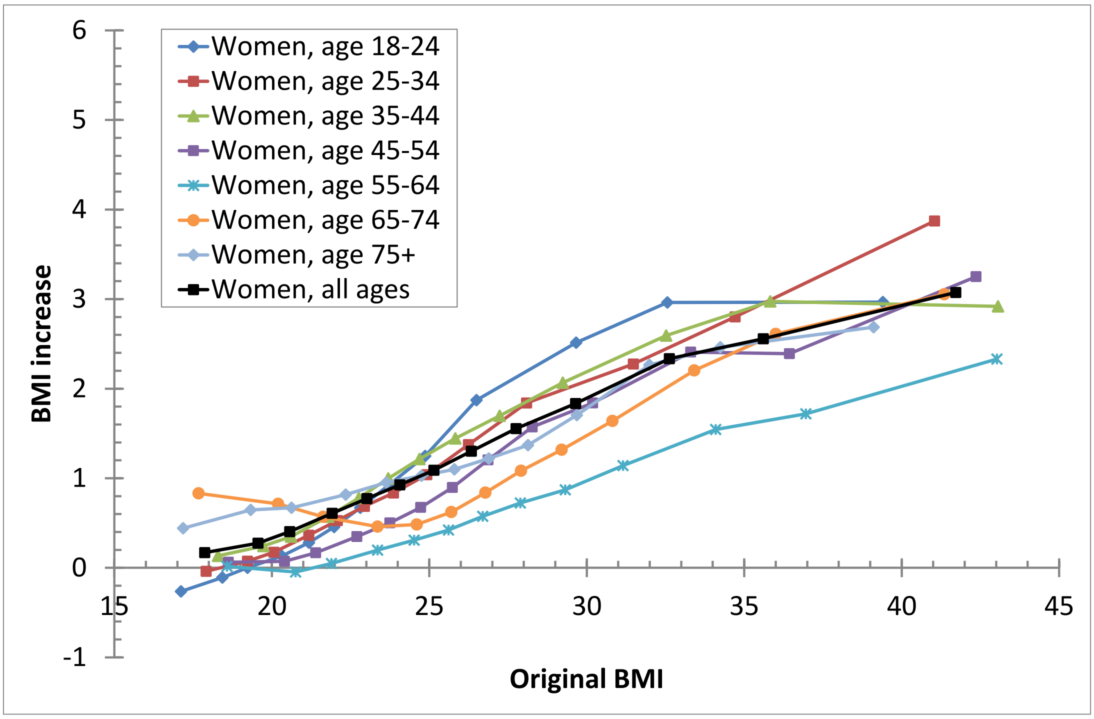


Figure 9.6 BMI increases for age group percentiles for women from 1991-1998 to 2007-2014.

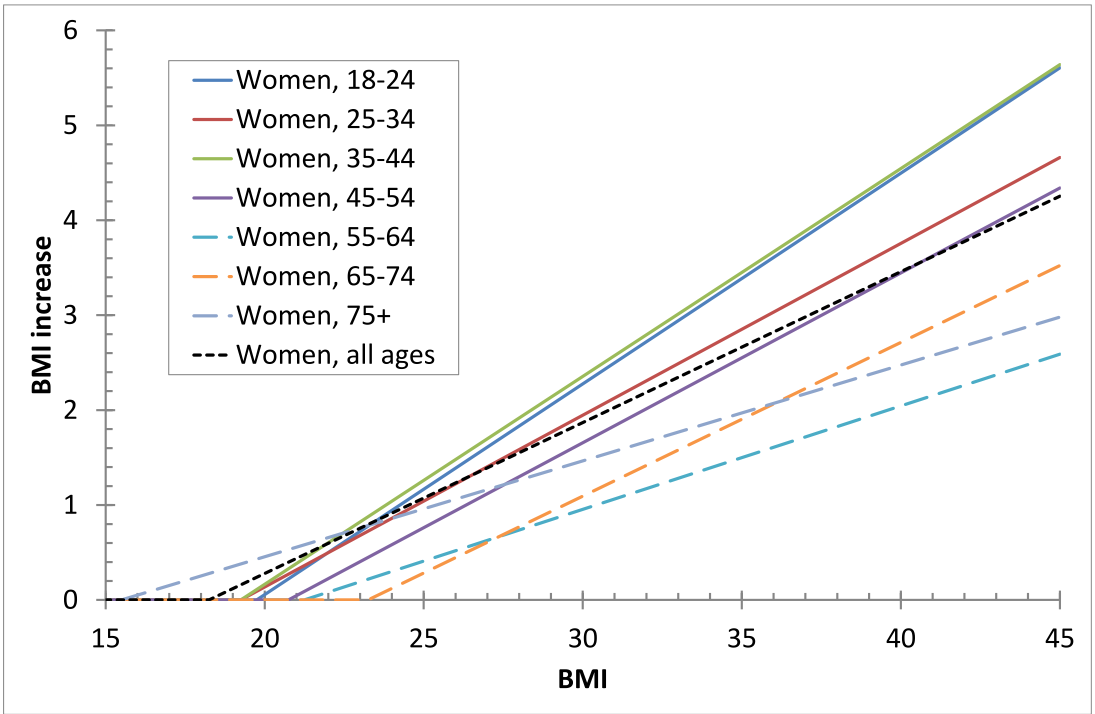


Figure 9.7 BMI increases for age group scaling models for women from 1991-1998 to 2007-2014.

Table 9.1 Values for the age group scaling models for women from 1991-1998 to 2007-2014.

Age group	18-24	25-34	35-44	45-54	55-64	65-74	75+	All
Start point	19.75	19.25	19.25	20.75	21.25	23.25	15.50	18.25
Scaling factor	1.222	1.181	1.219	1.179	1.109	1.162	1.101	1.159
Fitness value ($\times 10^{-5}$)	4.72	1.47	1.08	3.43	1.26	6.78	2.71	0.39

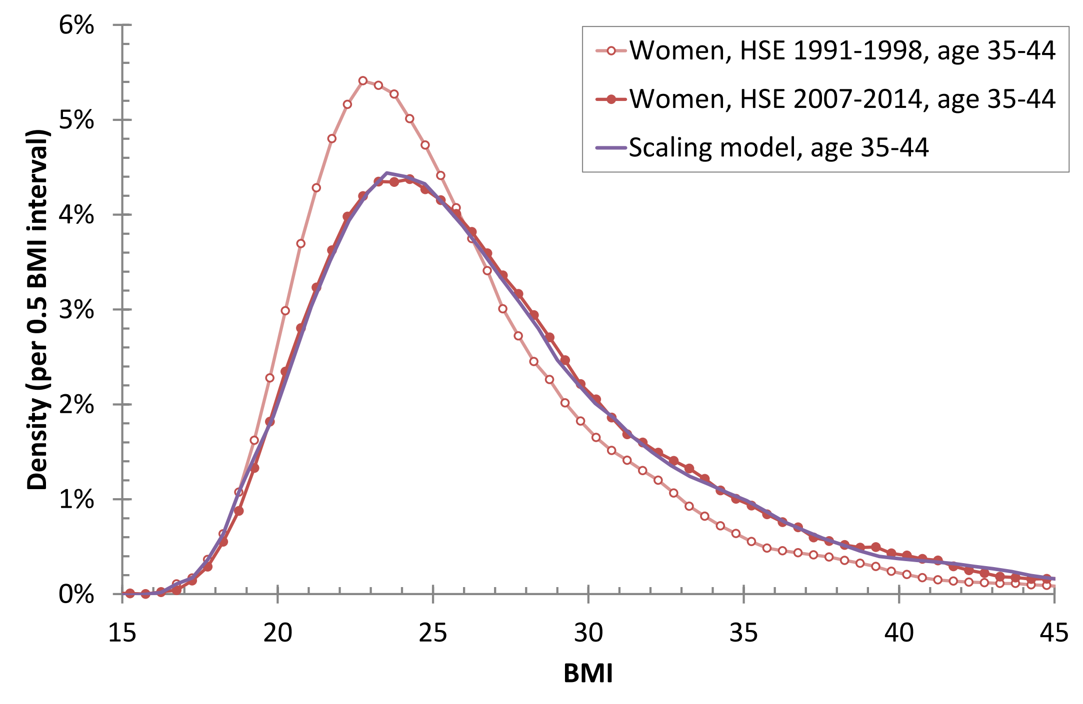


Figure 9.8 Scaling model for women for age group 35-44 from 1991-1998 to 2007-2014.

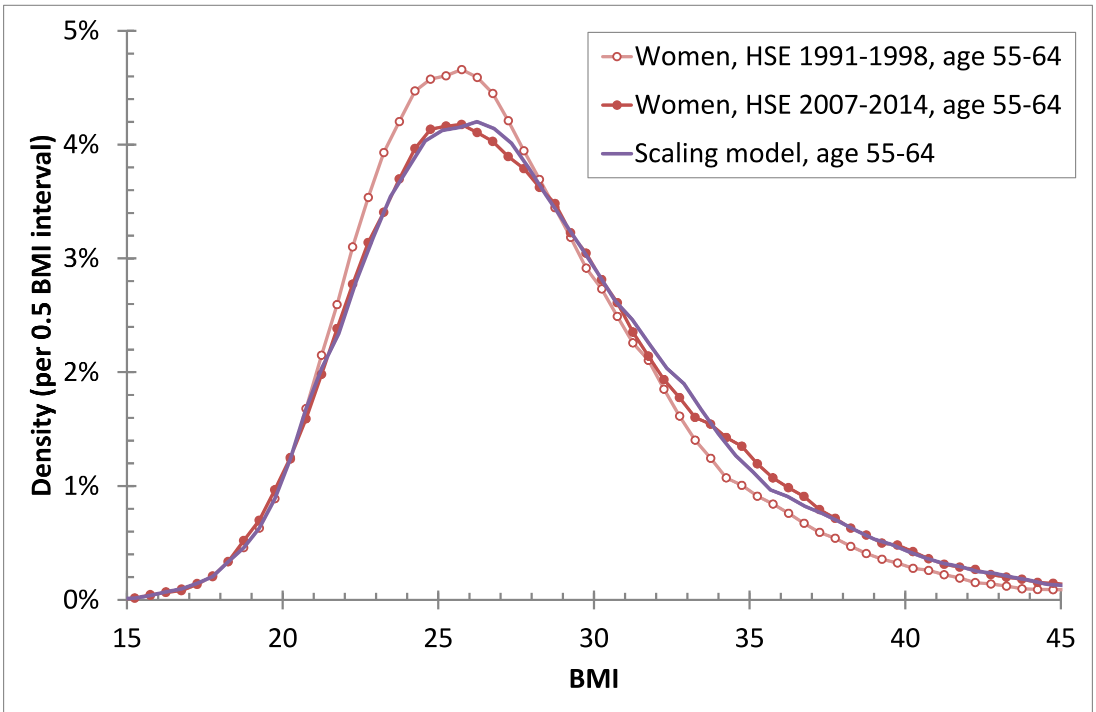


Figure 9.9 Scaling model for women for age group 55-64 from 1991-1998 to 2007-2014.

9.3.3 Implications of the age group BMI distribution changes with time

Section 8.3.3 discusses how the age groups relate to the birth cohorts in Chapter 7. The same comments apply here to the age groups for women. The age groups at later points in time relate to more recent cohorts. The results in Chapter 7 show that mean BMI is higher at the same age for more recent cohorts. From Chapter 7, within each cohort the relationship with age is similar with mean BMI increasing with age, but the more recent cohorts are at higher levels. This reflects mean BMI tending to increase over time for each age.

The analysis here extends the analysis of mean values by looking at how the distributions change over time for each age group. The results show that, as for the population as a whole, the changes over time for each age group are approximately a linear scaling transformation. This means that higher BMI values have higher increases in BMI. This is also the case for the age group distributions for men in Chapter 8. As discussed in Chapter 6, this is the result of living in a different time period and experiencing different conditions and having different lifestyles.

The implication is therefore that for each age group the effect of living in a later time period in increasing BMI is greater for those with a higher BMI than for those with a lower BMI.

9.4 Variation in the age group BMI distributions by age

9.4.1 Comparison of the age group BMI distributions by age

This section looks at the how the BMI distributions change with age for the data for women. Each time interval could be considered. For 1991-1998 these are the distributions in Figure 9.1. Similarly, for 1999-2006 it is Figure 9.2, and for 2007-2014 it is Figure 9.3. The visual impression is that part of the changes in the distributions with age is a shift along the x-axis since the left tail tends to move to the right. However, the peak of the distribution also tends to reduce with age and the right tail extends considerably, and so the changes also look to have some element of a scaling transformation. From Appendix 9.III, the standard deviation increases in each time period from age 18-24 to age 35-44, although it then reduces for older ages (Figure 9.III.9). The skewness tends to decrease with age (Figure 9.III.10). This indicates some change in shape, and is a little different to the distributions for men where the main change with age is a shift along the x-axis with not much change in shape(Section 8.4).

The main analysis done was for the most recent time interval of 2007-2014. The percentiles were calculated for each distribution and the changes from each distribution to the next age group distribution were computed. Figure 9.10 shows the values for 10% percentiles and 1%, 5%, 95%, 99%, for the age groups up to 65-74. The values for 75+ are not on this chart as they are all negative (since age 75+ has lower BMI values than age 65-74) and so it makes it clearer to see the values without this series on the chart.

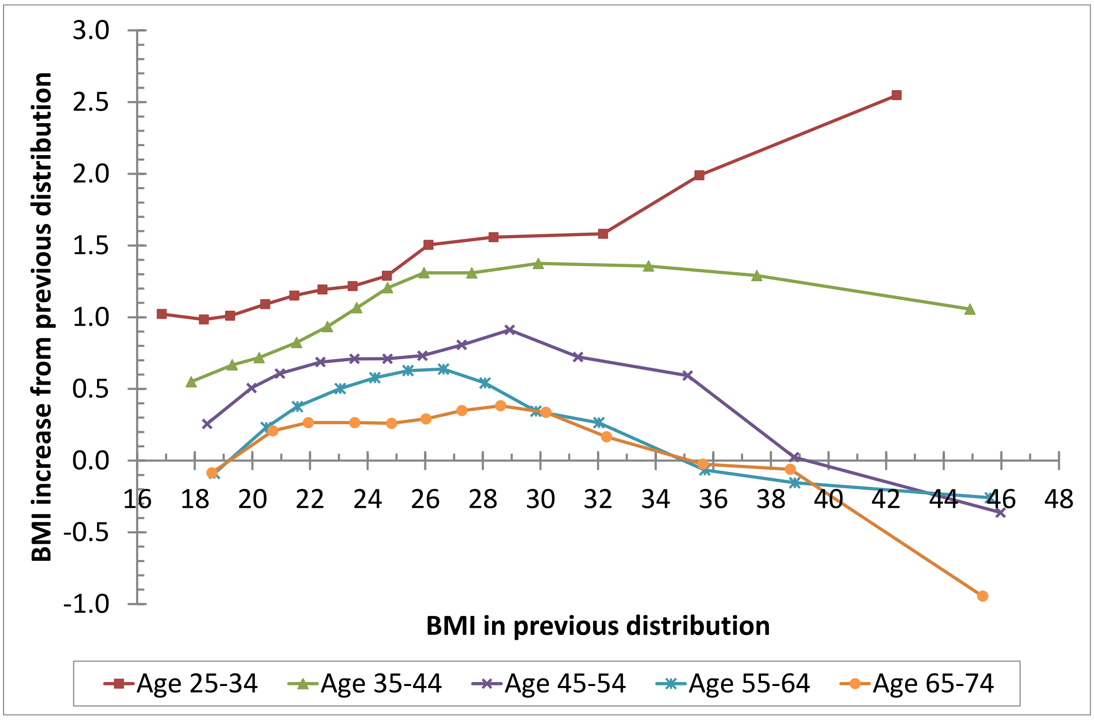


Figure 9.10 Percentile changes for the age group distributions for women for 2007-2014.

All of the series in Figure 9.10 show a positive gradient in the trend of values for the lower BMI values up to a BMI of about 28. However, the series for youngest ages for 25-34, 35-44, and 45-54, which have the largest changes, all start above 0. This supports the visual impression of the distributions in Figure 9.3, that the changes are a combination of a shift (translation) along the x-axis (where the percentile increases would be constant) and a linear scaling change in shape (where the percentile increases would be a straight line that starts from 0 at the BMI value where the scaling starts). The increases are greatest between the two youngest age groups and get lower with age.

The upwards trend for the series for 25-34 applies throughout all the points from the 5% value to the 99% value. The 1% value is very slightly higher than the 5% and 10% values. The 5% value has an increase of 0.98 at a BMI of 18.32 and the 99% value has an increase of 2.55 at a BMI value of 42.37.

The series for 35-44 has an upwards trend for the BMI increases for percentiles from 1% to 80% (up to the fourth from the right value in the series). The 1% value has an increase of 0.55 at a BMI value of 17.88 whereas the 80% value has an increase of 1.37 at a BMI value of 29.93. Visually, there is a considerable change in shape in the distributions for 18-24, 25-34, and 35-44 in Figure 9.3 with the peak in each case getting lower from one distribution to the next.

The changes in the distributions were modelled for the three biggest changes from 18-24 to 25-34, 25-34 to 35-44, and 35-44 to 45-54. This was done as both shift and scaling transformations. The scaling transformations fitted considerably better than a shift but with a low starting value for the scaling function. Therefore, within the range of the BMI values of the distributions these transformations are a combination of both a shift and a linear scaling. The details of the scaling models along with a chart of the BMI increases is in Appendix 9.V.

9.4.2 Implications of the age group BMI distribution changes with age

The way that the age group BMI distributions for women change with age is different to the way that they change with time. There is also some difference to the way that the age group BMI distributions change with age for men.

The changes in the BMI distribution with age for women are a mixture of a shift (translation) along the x-axis and a linear scaling change in shape. Hence, BMI tends to increase with age for all BMI levels but with higher BMI values tending to increase by more, particularly from age group 18-24 to 35-44. Hence, BMI tends to increase with age for all ages but with a tendency for greater increases for higher BMI values.

By contrast, the changes of the distributions with time are mainly a linear scaling transformation. The changes in the distributions with age for men are mainly a translation of the distribution shifting along the x-axis. For women the changes in the distributions with age are a mixture of both a translation and a scaling. This means that there is more change in shape in the BMI distributions with age for women than for men, and this comparison is considered further in Chapter 10.

9.5 Single year of age BMI distributions

The age groups considered so far in this chapter combine a reasonably wide spread of ages (seven years for 18-24, ten years for 25-34, 35-44, etc.). This gives good sample sizes but it is also interesting to look at a narrower range of a single year of age to see what those distributions look like and how they change. Section 9.5.1 gives the results for ages 18-54 in six year gaps. Then, Section 9.5.2 considers the change each year from 18-24 since that is the ages over which BMI increases most rapidly.

9.5.1 Ages 18-54 in 6 years gaps

The distributions for single years over the range of 18 to 54 were calculated and analysed. Following the same approach applied to the data for men (Section 8.5.1), this was done using 6 year gaps for data for 1991-1998 and 1999-2014. The ages from 18 to 54 cover most of the aging increase in BMI in the current data (as can be seen, for example, from the mean BMI values in Figure 9.I.1 in Appendix 9.I).

The smoothed distributions for 1999-2014 are shown in Figure 9.11. The full statistics for 1991-1998 and 1999-2014 are given in Appendix 9.VI. Sample sizes vary from 610 to 968 for 1991-1998, and from 855 to 1396 for 1999-2014. The sample sizes are quite small and so the distributions have some variations and irregularities, but are still reasonably smooth. The changes in the statistics with age are considerable. For 1999-2014, from age 18 to 36 to 54 the mean BMI goes from 23.8 to 26.7 to 27.8, and the healthy percentage goes from 63.7% to 44.6% to 34.9% (Table 9.VI.2).

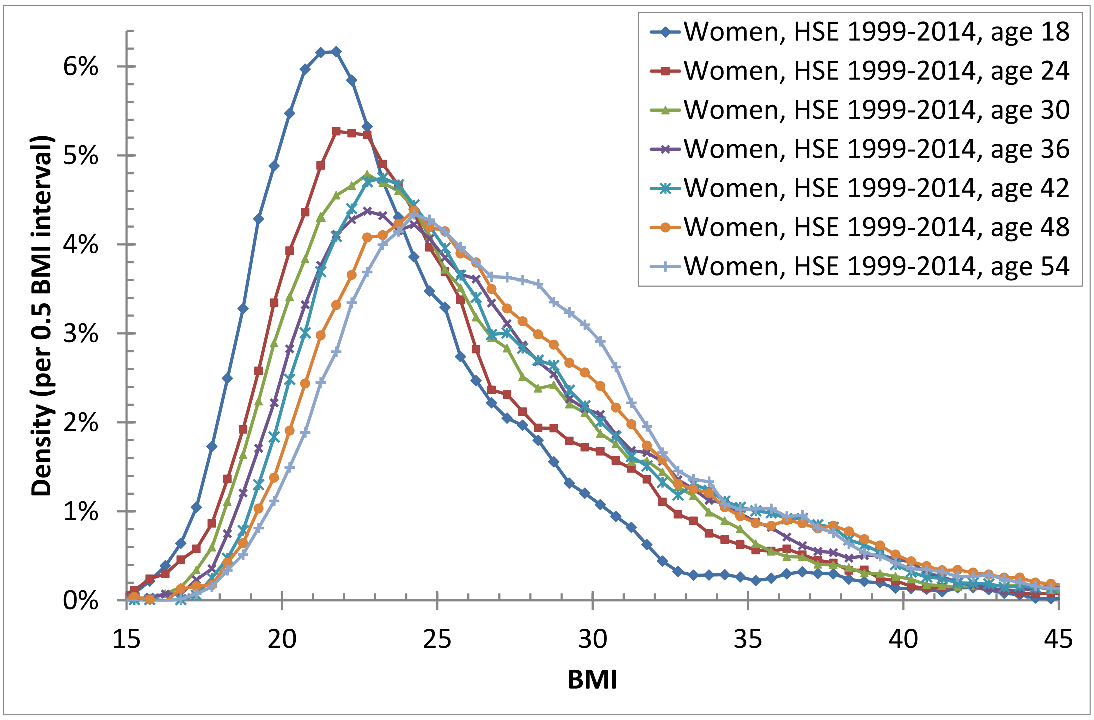


Figure 9.11 Distributions for ages in 6 years gaps from 18 to 54 for men for 1999-2014.

Percentiles were calculated by time and by age, in the same way as for the age groups. The results are given in Appendix 9.VI and the patterns are generally similar to those for the age groups in Sections 9.2, 9.3, and 9.4, although with more irregularities. The changes with time are in Figure 9.VI.3. The patterns are approximately a linearly increasing trend starting from the x-axis. The changes with age are in Figure 9.VI.4 and are more variable. From age 18 to 24 the pattern is roughly a linear increase but starting above 0. The other ages have much smaller increases, and some have a trend of a small positive gradient for the increases for the lower BMI values.

Scaling models were applied to model the changes over time for each age from 1991-1998 to 1999-2014. As would be expected from the percentiles the linear scaling transformations are able to model the changes in the distributions reasonably well and the BMI increases for the models are similar to the percentiles. The fitness values for the seven ages are ($\times 10^{-5}$): 12.51, 7.11, 5.00, 5.53, 25.08, 8.77, 17.64. The values are quite high for ages 42 and 54, which mainly reflects irregularities and detailed differences in the shapes of the distributions. The matches for the other distributions are all quite good. The detailed results of the scaling models are not included given the range of results already presented in the report and the appendices.

As was also the case for the data for men, looking at the single year ages highlights the extent of the difference between age 18 and the other ages. In Figure 9.11, the age 18 distribution has a higher peak than age 24, which in turn has a higher peak than the older ages. The peaks for ages 18 and 24 are also narrower than for the older ages. This is also reflected in the statistics given in Appendix 9.VI with the age 18 distribution having a lower standard deviation and a higher skewness than the distributions for the other ages. The percentiles also indicate a change in shape from age 18 to 24 (Figure 9.VI.4), with much higher increases for higher BMI values than for lower BMI values.

Transformations were applied to the changes with age for the distributions in Figure 9.11 for 1999-2014. For 18 to 24, 24 to 30, and 30 to 36, a linear scaling works much better than a shift. The fitness values are ($\times 10^{-5}$): 6.21, 7.47, 2.79; the scaling starting values are: 15.50, 15.50, 14.50; the scaling factors are: 1.191, 1.076, 1.072. So, the transformations are mainly a scaling but with low starting values and so there is some increase in BMI even for low values in the starting distribution.

The changes for the distributions for ages above 36 are smaller and more irregular and so the best transformation is less clear. For example, from Figure 9.11, age 36 has a lower peak than age 30. However, the peak then increases with age 42 being similar to age 30. Then ages 48 and 54 are lower again and similar to age 36, and the changes from 36 to 48 or 54 can be modelled best as a shift. The distribution for age 54 also has a bulge in the right tail. These irregularities probably reflect the small sample sizes. The overall changes from age 18 to 48 or 54 can be modelled well as a scaling, albeit with fairly low starting values of 14.00 and 12.25 respectively. The quartile values with age are considered in Appendix 9.VI in Figures 9.VI.18-19. These show higher values increasing more with age again indicating a change of shape that is to some extent like a linear scaling transformation.

The general indication is that the increase in BMI with age is mainly a linear scaling transformation, particularly at younger ages. The scaling starting values are quite low. The changes at older ages have some element of a shift (translation). Overall, there is a tendency for BMI to increase with age for all starting values of BMI, but with the increases being greater for higher BMI values.

9.5.2 Ages 18-24 for women

BMI increases most rapidly for the youngest ages and so smoothed distributions and statistics were produced for each year of age from 18 to 24 to examine the changes in these ages. The HSE time intervals of 1991-1998 and 1999-2014 were used, as in Section 9.5.1. Detailed results are in Appendix 9.VII.

Figure 9.12 shows the mean and median values. The mean and median tend to increase with each year of age. There are some irregularities, for example age 21 has high BMI for 1999-2014 with the mean and median then decreasing for age 22, and for 1991-1998 age 22 has slightly higher mean and median BMI than age 23. The sample sizes are small being between 575 and 785 for 1991-1998, and between 748 and 856 for 1999-2014. The indication is that the sampling variation is quite large relative to the changes from one year of age to the next, and hence the irregularities in the patterns. The overall changes from age 18 to 24 are big, however, with the mean and median BMI increasing by 1.39 and 1.06 for 1991-1998 and by 1.49 and 1.33 for 1999-2014. This is not quite as much as for men where the increases in mean and median are 1.96 and 2.18 for 1991-1998 and 1.90 and 2.12 for 1999-2014 (Section 8.5.2).

In Figure 9.12, the increases in mean and median BMI from 1991-1998 to 1999-2014 are considerable with the mean and median much higher for each age in 1999-2014. The average increase for the seven ages is 0.82 for the mean and 0.52 for the median. This is a bit more than for men where the increase values are 0.61 and 0.37 (Section 8.5.2).

The smoothed distributions for 1999-2014 are in Figure 9.13 and there is a trend of the distributions moving to the right with age, with a lower peak and a longer right tail. As noted above, the changes with age from 18 to 24 are not quite as large for women as for men The visual differences between the distributions are also not as clear: for men the distribution for age 18 looks very different to the older age distributions (Figure 8.14), but this is not the case to the same extent for women.

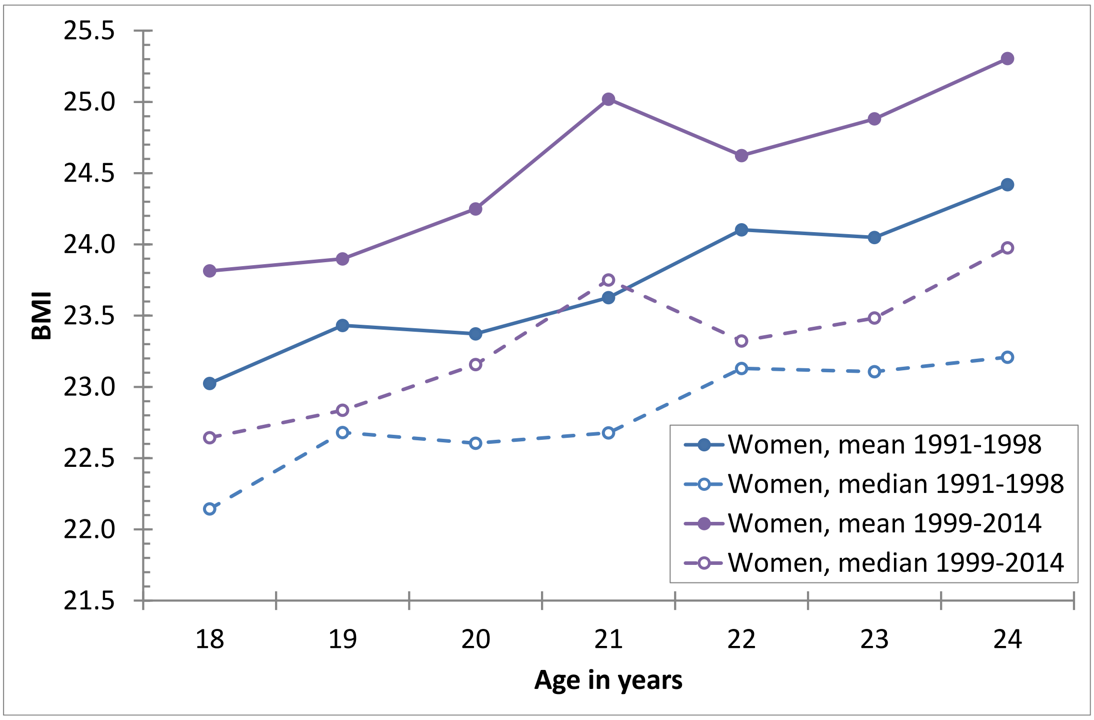


Figure 9.12 Mean and median for each year of age for women from 18 to 24.

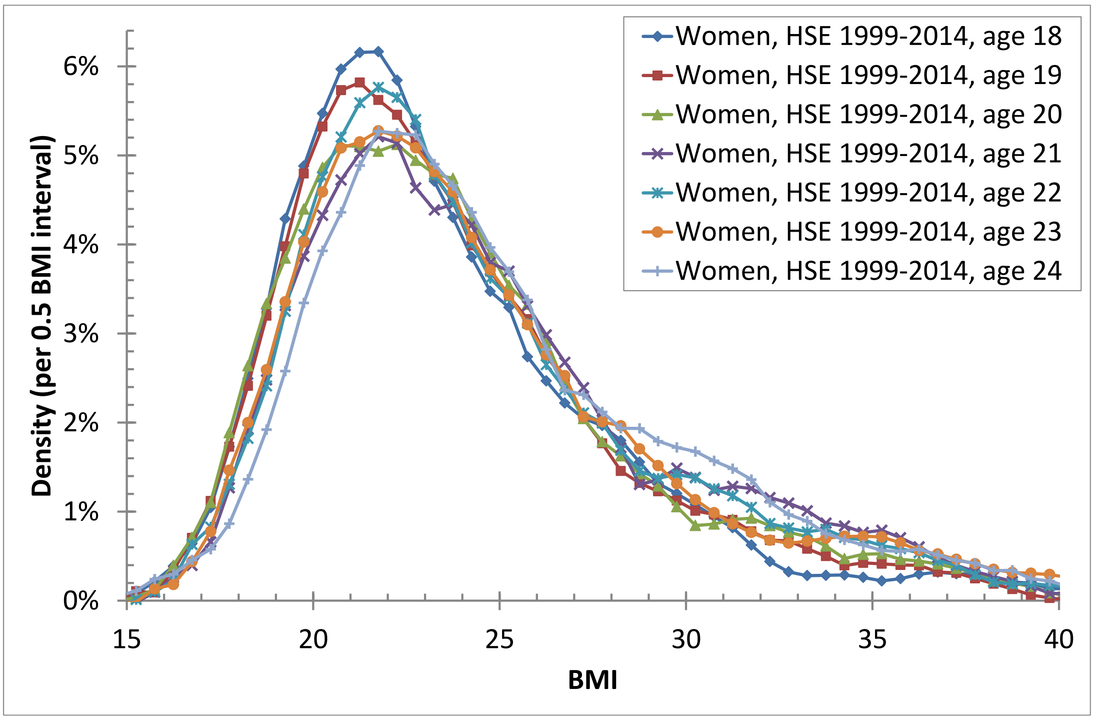


Figure 9.13 Smoothed distributions for each year of age for women from 18 to 24 for 1999-2014.

The percentiles are plotted in Appendix 9.VII. The changes with time in Figure 9.VII.3 are consistent with the other results in this chapter in having a pattern corresponding generally to a linear scaling, with the percentiles approximately linearly increasing from the x-axis.

The changes with age in Figure 9.VII.4 are more variable in pattern, reflecting the comments above about the irregularities (e.g., the mean BMI for age 22 for 1999-2014 being lower than for age 21). The percentiles for gaps of six years in Section 9.5.1 therefore give a clearer picture in that the differences are larger between the successive distributions and so the relative effect of the sampling variation is smaller.

The distribution for age 18 is the most healthy from the statistics in Appendix 9.VII. For 1999-2014 (Table 9.VII.2) the mean BMI is 23.8 with 63.7% in the healthy BMI category. Only 9.2% are in the obese categories, while there are 6.4% who are in the underweight category. By age 24 the mean BMI has increased to 25.3 with now only 54.7% in the healthy category and a big increase to 17.4% in the obese categories. The underweight percentage is 3.9% at this age.

The statistics are even better for 1991-1998 (Table 9.VII.1). The age 18 statistics are a mean BMI of 23.0 and 69.7% in the healthy category. Just 6.9% are in an obese category with 7.6% in the underweight category. For age 24 the mean BMI is 24.4 with 62.1% in the healthy category. There are 12.4% in an obese category and 3.6% in the underweight category.

9.5.3 Implications of the single year age BMI distributions

BMI tends to increase with each year of age from age 18 with high increases in the younger ages. As stated in Section 8.5.3, I am not sure to what extent the increase in BMI over the younger ages is simply a result of the natural development and maturing of the body, and to what extent it is due to lifestyle effects and changes in circumstances and behaviour.

As is the case for men, the BMI distributions for women at age 18 have a high percentage in the healthy category. In fact, the BMI distributions for men and women are noticeably very similar at this age and this aspect is explored further in Chapter 10.

Making a significant reduction in the extent of the increases in BMI from age 18 would, over time, considerably reduce the overall prevalence of obesity in the population. Therefore, the results here reinforce the comments in Section 8.5.3 that targeting interventions at those around age 18 may be an especially effective strategy. As suggested in Section 8.5.3, health messages could particularly try and reach this age group, and school or further education institutions could potentially develop beneficial initiatives.

The underweight category percentage is also high at age 18 and so interventions to address this could be considered, although this percentage does reduce as age increases.

Chapter 10: Comparison of the changes in BMI with age for men and women

Key points from this chapter:

- The BMI distributions at age 18 are very similar for men and women, with much the same shape and location. Consequently, the statistics are similar for age 18. As noted in the previous three chapters BMI increases with age and so age 18 has the lowest BMI values and the best statistics for the adult age range considered in this report. For the age 18 distributions for the 1999-2014 data there are 67.4% of men and 63.7% of women in the healthy category.
- BMI increases with age for men and for women but with different patterns. The changes imply a similar tendency for BMI to increase across all the different starting BMI values for men above age 24. For women there is a tendency for high BMI values to increase more than low BMI values.
- The distributions become quite different with increasing age. In the age examples considered here in the 30s, 40s and 50s, the distributions for men are more symmetrical than those for women, with a peak further to the right and a lower right tail. These different shapes mean that the healthy percentages are much lower for men. High BMI values are more prevalent for women.
- There may be implications for health interventions in that a broader strategy may be better for men and a more targeted strategy better for women.

10.1 Choice of analysis

The previous two chapters look in detail at BMI for different ages. Chapter 8 analyses the data by age for men in various ways and Chapter 9 has the same set of analysis for the data for women. Each of the results can be compared, and Chapter 9 makes some comments about where the patterns for women are similar or different to those for men.

One interesting aspect identified in Chapters 8 and 9 is that the BMI distributions for age 18 are reasonably healthy for both men and women. In fact, the distributions and statistics appear quite similar. As age increases from 18, BMI increases rapidly although at different rates for men and women. The aim in this chapter is to do a more detailed direct comparison between the distributions and statistics for men and women for these aspects. Specifically, comparing the BMI distributions for age 18 and then how they change with age. This was done using the BMI distributions for single years of age at six year intervals for the HSE data for 1999-2014 as set out in Sections 8.5.1 and 9.5.1 (as shown in Figures 8.11 and 9.11).

10.2 Comparison of the BMI distributions at age 18

The BMI distributions for men and women at age 18 for the HSE data for 1999-2014 from Sections 8.5.1 and 9.5.1 were plotted together to compare them. Figure 10.1 shows the distributions for age 18 between BMI values of 15 and 35 to focus on the main part of the distribution. A plot for BMI values of 15 to 45 is in the next section in Figure 10.2. The descriptive statistics are in Table 10.1.

The BMI distributions for age 18 are very similar. The shape of each distribution in Figure 10.1 is roughly the same. The small differences are that the distribution for women has a slightly lower and narrower peak, with a higher right tail. The peak of the distribution for women is located a little to the right of where it is for men. Regarding the statistics, these differences in shape result in the standard deviation and skewness being higher for the distribution for women. The overweight percentage and each of the obese category percentages are slightly higher for women with the healthy percentage being a little lower (63.7% compared to 67.4%). The mean BMI for women is 0.6 higher and the median is 0.3 higher.

These differences are all small and the main observation is that the distributions are very close to each other. Given this similar starting point at age 18 it is then interesting to see how things change with increasing age for men and women.

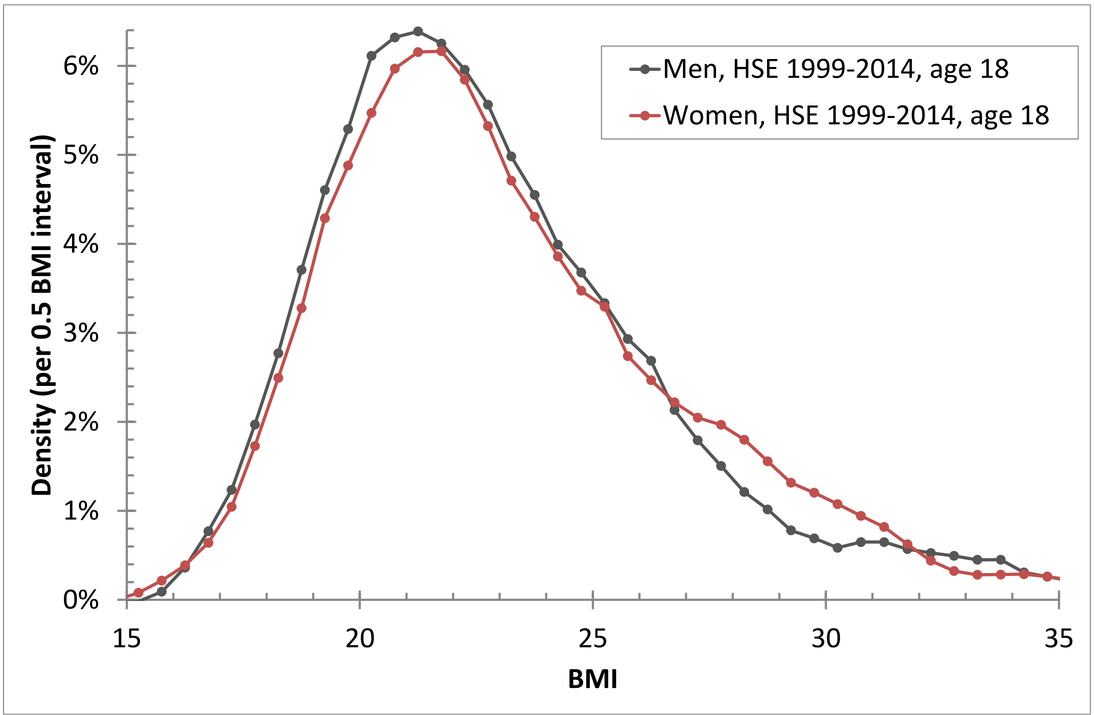


Figure 10.1 BMI distributions for age 18 for men and women for 1999-2014 from BMI 15 to 35.

Table 10.1 Descriptive statistics for the age 18 distributions for men and women for 1999-2014

Population category	Men age 18	Women age 18
Sample size of actual data	769	855
Mean	23.24	23.81
Median	22.35	22.64
Modal interval mid-point	21.25	21.75
Standard deviation	4.26	5.02
Skewness	1.47	2.12
Underweight (BMI < 18.5)	7.1%	6.4%
Healthy (18.5 ≤ BMI < 25.0)	67.4%	63.7%
Overweight (25.0 ≤ BMI < 30.0)	18.1%	20.6%
Obese I (30.0 ≤ BMI < 35.0)	4.9%	5.3%
Obese II (35.0 ≤ BMI < 40.0)	1.9%	2.5%
Severely obese (BMI ≥ 40.0)	0.6%	1.4%

10.3 Comparison of the BMI distributions from age 18 to 54

10.3.1 Comparison of the distributions and statistics

Charts were produced of the distributions for men and women for each single year of age in gaps of six years from 18 to 54. These are shown in Figures 10.2-10.8. Each chart is plotted on the same scales for the axes for direct comparison between all the charts. The sample sizes are fairly small, all being between 712 and 1396, and so there are some irregularities. The distributions for both ages 18 and 36 are plotted together in Figure 10.9, to show the contrasting ways in which the distributions for men and women change with age. The statistics are in Appendix 10.I. The category percentages are in Figures 10.10 and 10.11, and the mean and median values are plotted in Figure 10.12. Appendix 10.I has separate charts of each BMI category with the values for men and women shown on the same chart for comparison, as well as charts of the changes in the quartile values with age.

At age 18 the BMI distributions for men and women are very similar (Figure 10.2). As age increases the distributions for men and for women change in different ways. This has been considered before when discussing the distributions and the percentiles in Sections 8.5.1 and 9.5.1. The charts here highlight and show the clear differences in the changes for men and women. The distributions for men mainly shift right along the x-axis whilst keeping a similar shape. The distributions for women do not shift as much but change shape with a longer right tail. One aspect is that the left tail shifts with age a lot more for the distributions for men than for women.

The result is a considerable difference in the distributions for men and women as age increases. At older ages, the distributions for men have a higher peak that is situated at a much higher BMI value. The distributions for men are reasonably symmetrical whereas for women they are strongly positively skewed with a long right tail. The differences in the changes are highlighted in Figure 10.9 in showing the distributions for ages 18 and 36. At age 18 the distributions for men and women are very similar, whereas they are quite different at age 36.

For older ages, the curves for men and women cross just before the peak of the distribution for men and again in the right tail. The prevalence is higher for women for low and high BMI values, but is higher for men for middle values. For the age 36 distributions in Figure 10.9, compared to the distribution for men the distribution for women has frequencies that are 16.0% higher for BMI < 24.0, 20.1% lower for 24.0 ≤ BMI < 33.0, and 4.1% higher for BMI ≥ 33.0. The percentages for the categories for men and women respectively for age 36 are: underweight: 0.3%, 1.6%; healthy: 31.3%, 44.6%; overweight: 45.9%, 30.1%; obese I: 17.2%, 14.7%; obese II: 4.0%, 6.1%; severely obese: 1.3%, 3.0%. Hence, the healthy category percentage is much higher for women by 13.2%, being 44.6% compared to 31.3% for men. With the underweight percentage being quite small for both men and women, the total percentage being overweight or obese is naturally a similar level of difference and is 14.5% higher for men, with 68.3% for men and 53.8% for women. For the high BMI values in the obese II and severely obese categories, the total prevalence is 3.8% higher for women at 9.1% compared to 5.3%.

The charts for the category percentages in Figures 10.10 and 10.11, and for the mean and median in Figure 10.12 naturally reflect the comments above. The healthy category percentage reduces more with age for men than for women, and both the mean and median increase by more with age for men than for women. On the other hand, the obese II and severely obese percentages are higher for women for all ages and also increase by more with age (Figures 10.I.8 and 10.I.9 in Appendix 10.I).

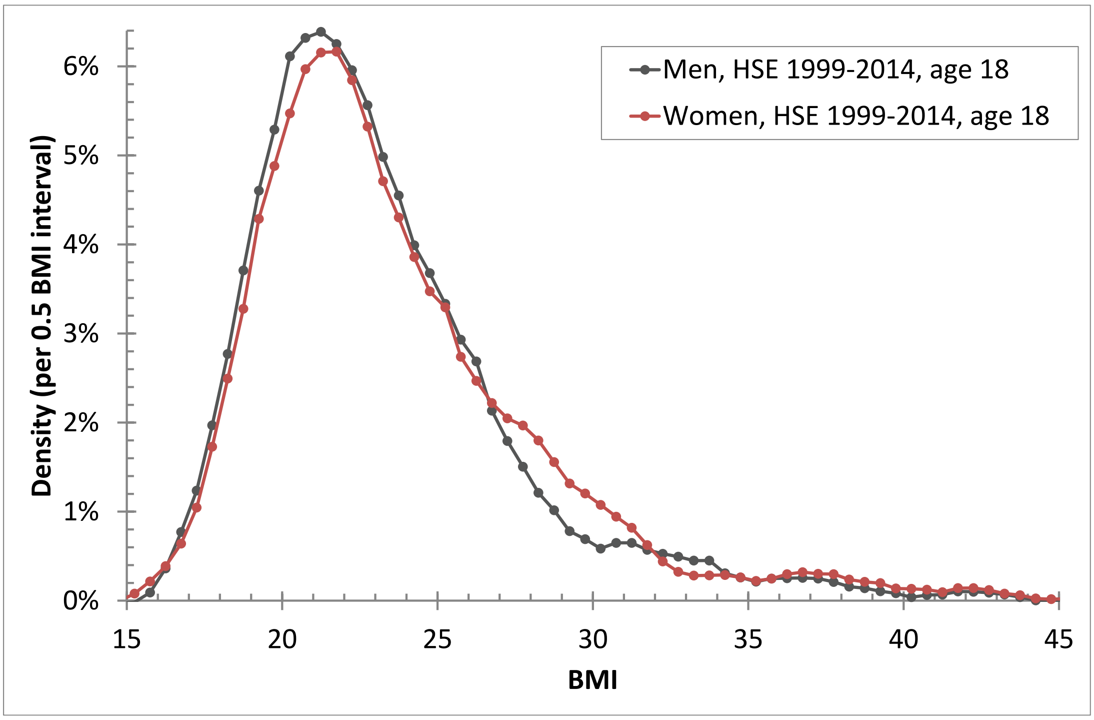


Figure 10.2 BMI distributions for age 18 for men and women for 1999-2014.

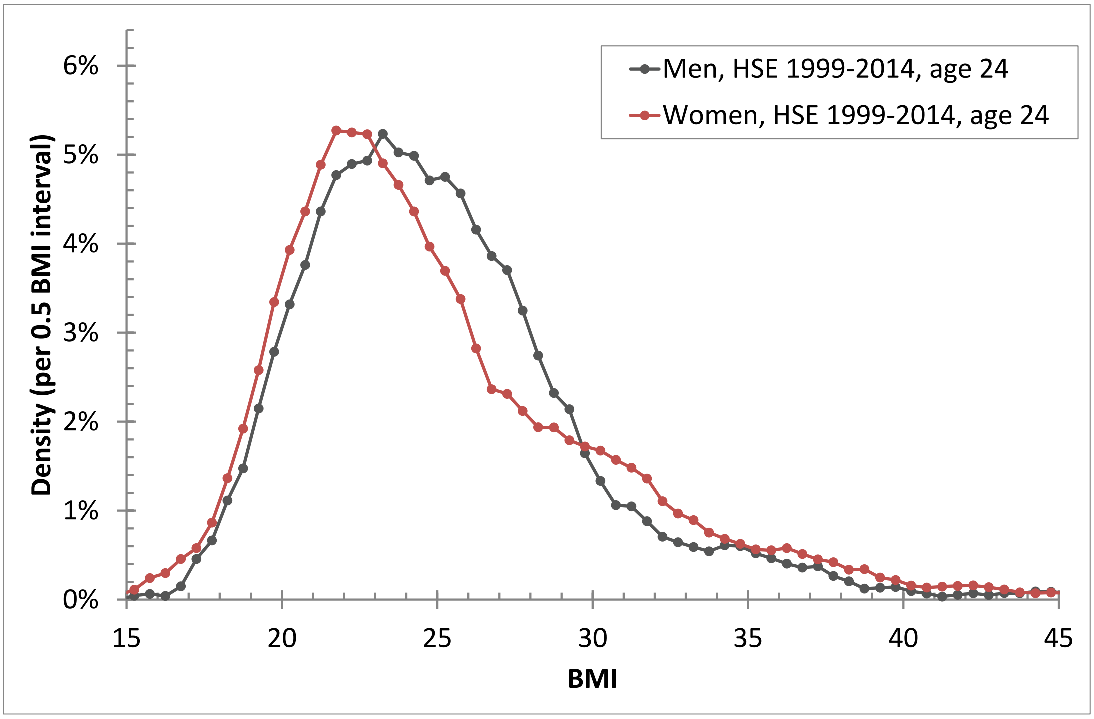


Figure 10.3 BMI distributions for age 24 for men and women for 1999-2014.

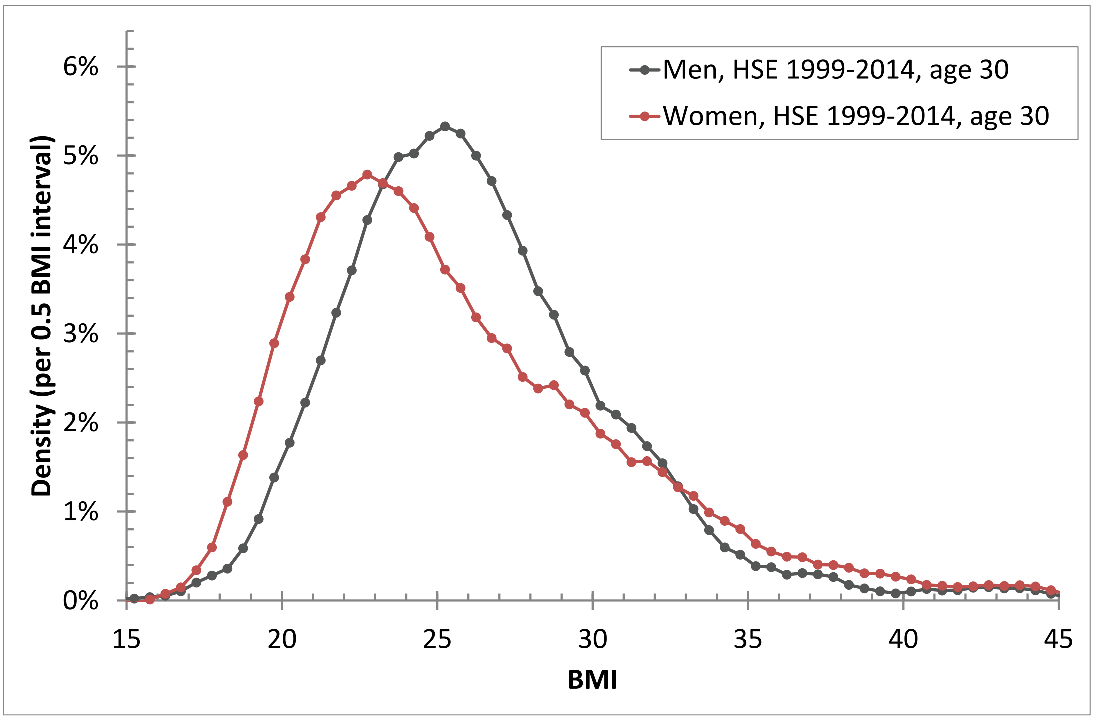


Figure 10.4 BMI distributions for age 30 for men and women for 1999-2014.

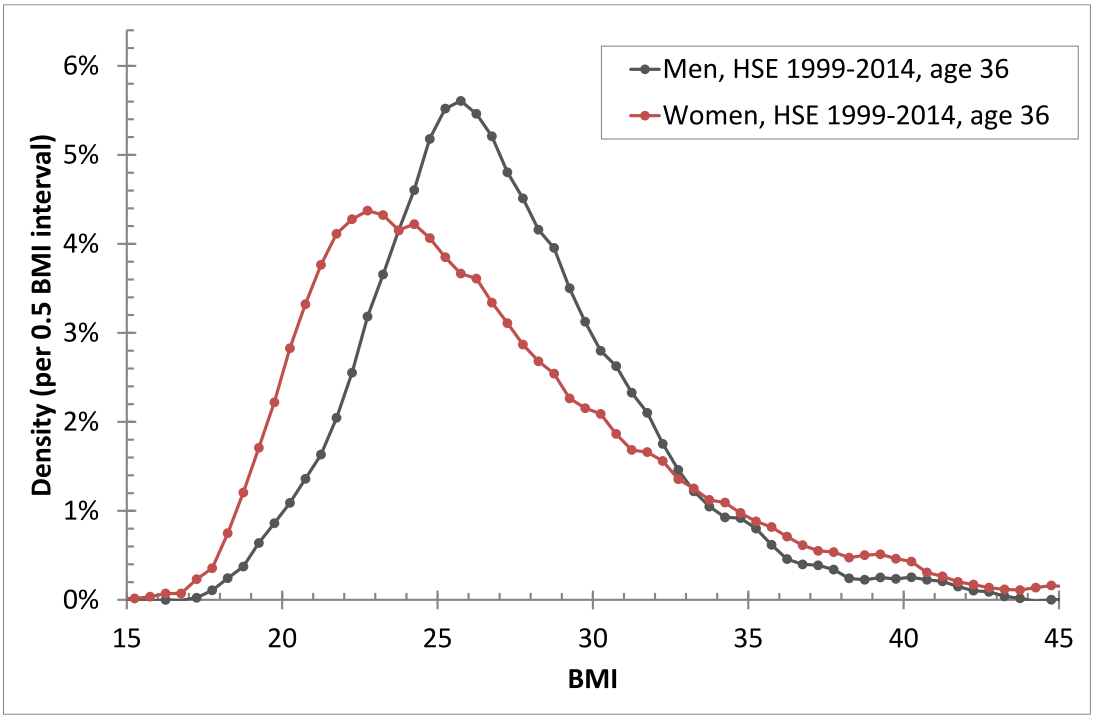


Figure 10.5 BMI distributions for age 36 for men and women for 1999-2014.

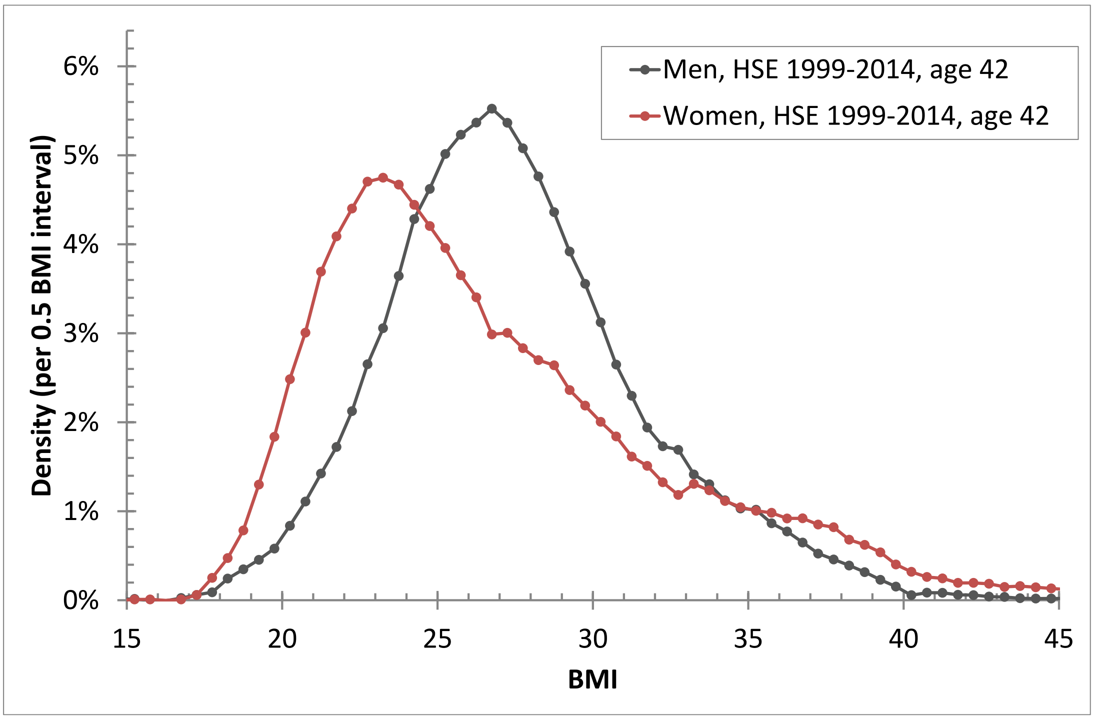


Figure 10.6 BMI distributions for age 42 for men and women for 1999-2014.

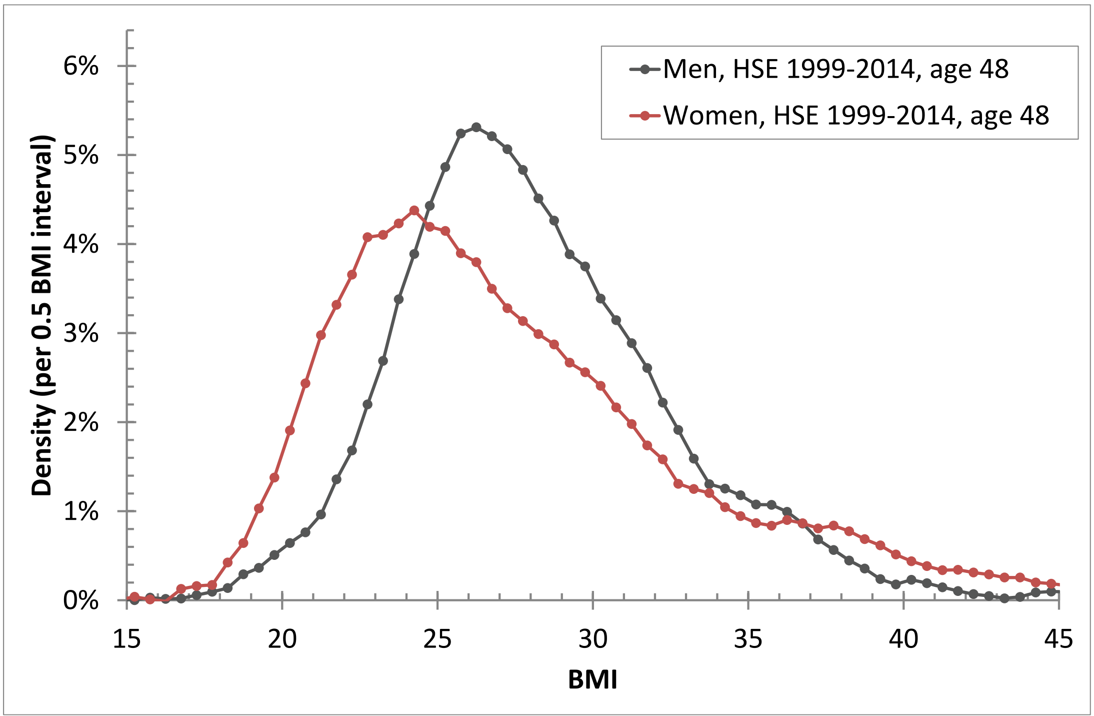


Figure 10.7 BMI distributions for age 48 for men and women for 1999-2014.

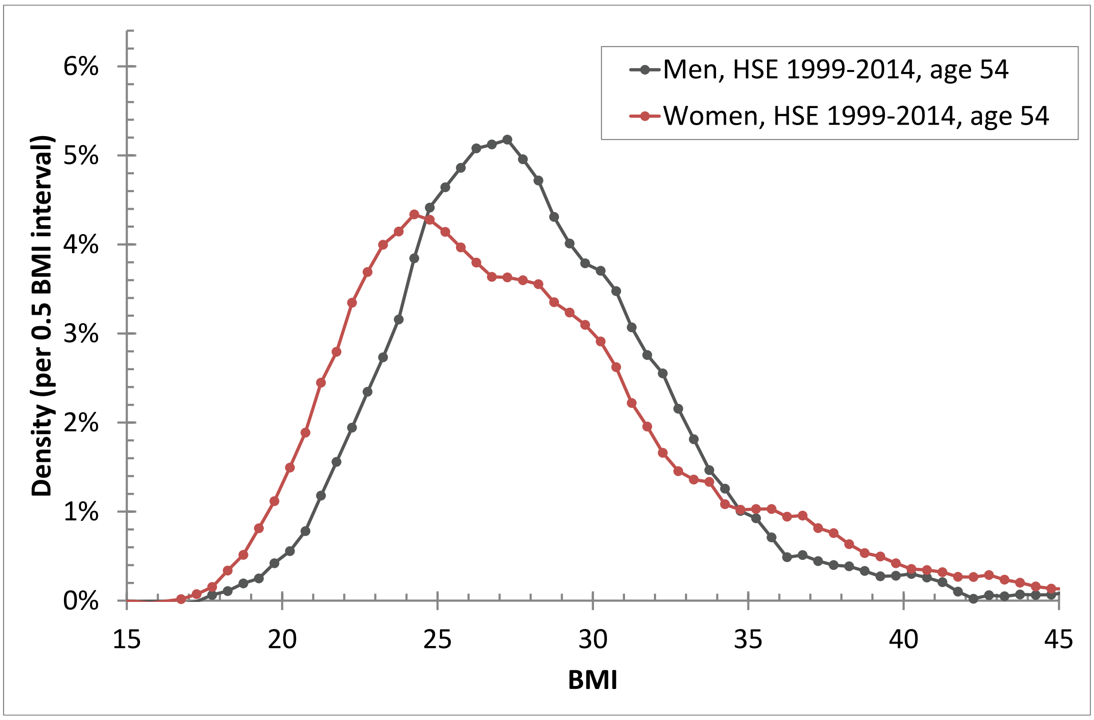


Figure 10.8 BMI distributions for age 54 for men and women for 1999-2014.

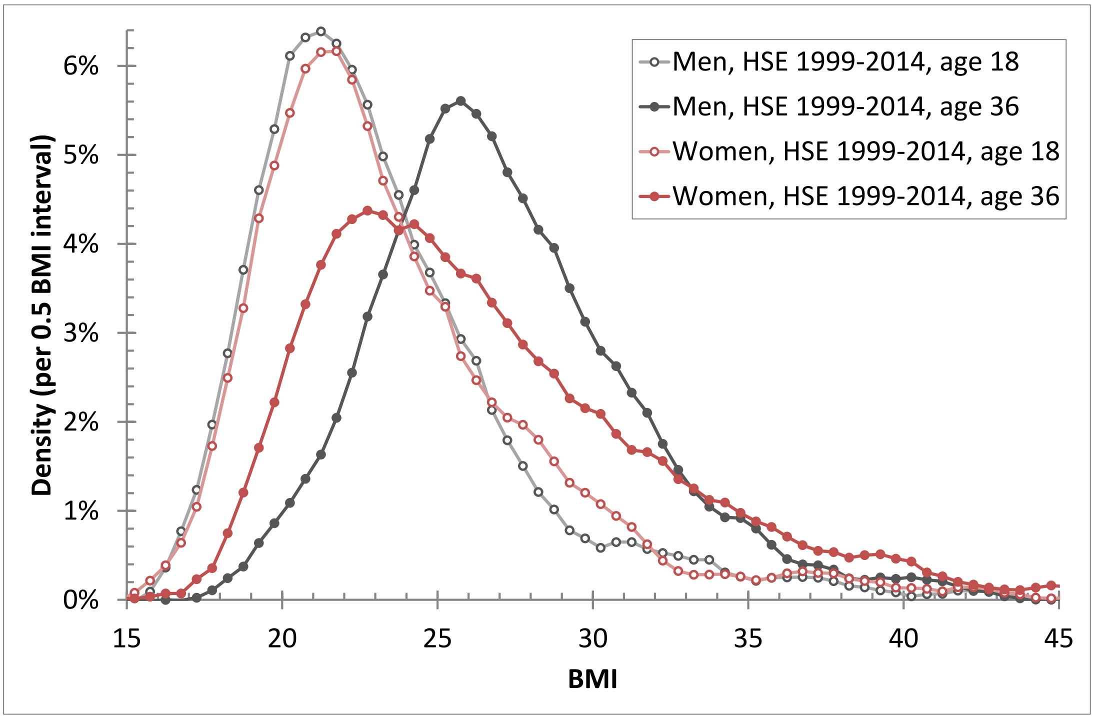


Figure 10.9 BMI distributions for age 18 and 36 for men and women for 1999-2014.

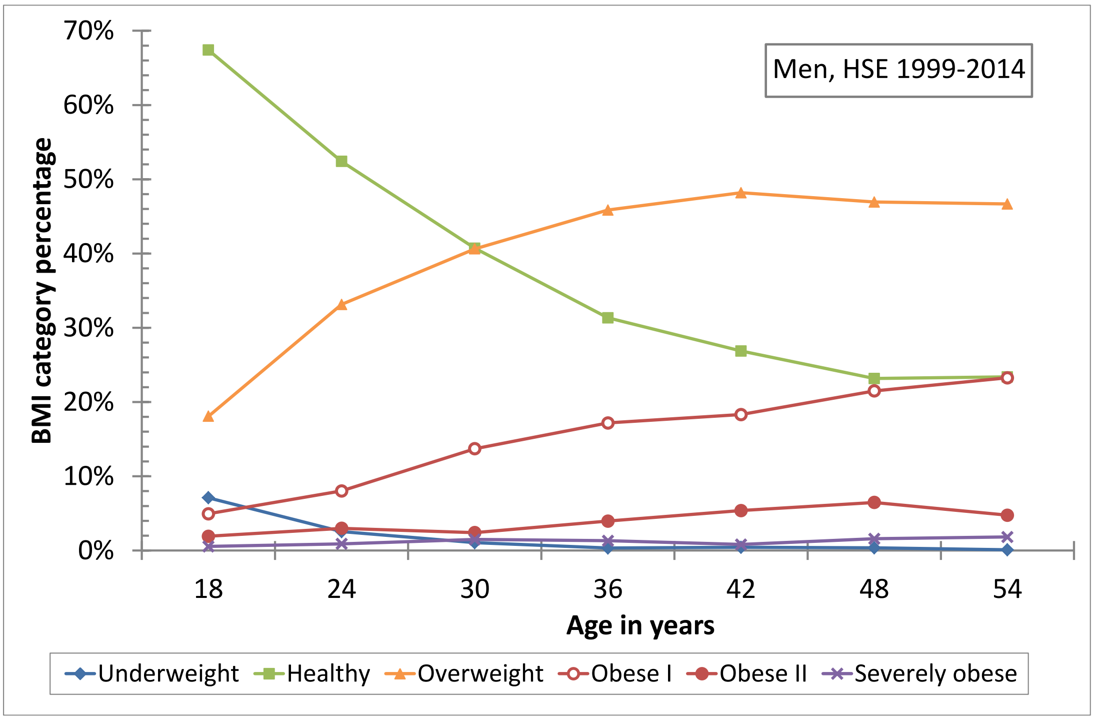


Figure 10.10 Category percentages for selected ages from 18 to 54 for men for 1999-2014.

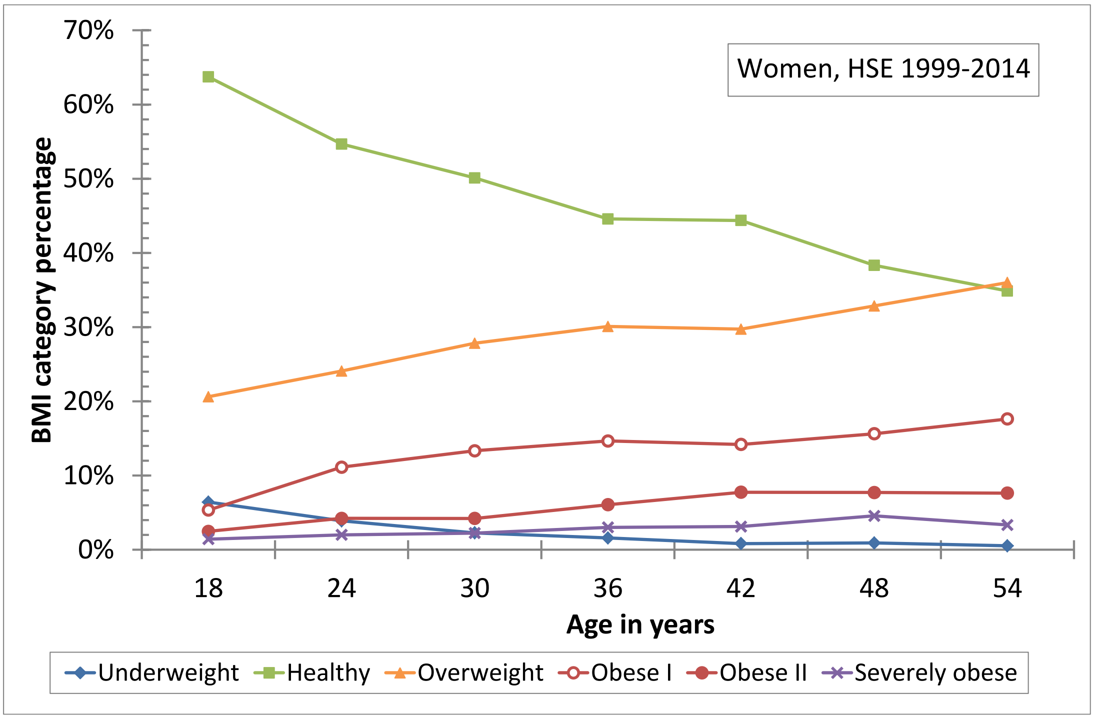


Figure 10.11 Category percentages for selected ages from 18 to 54 for women for 1999-2014.

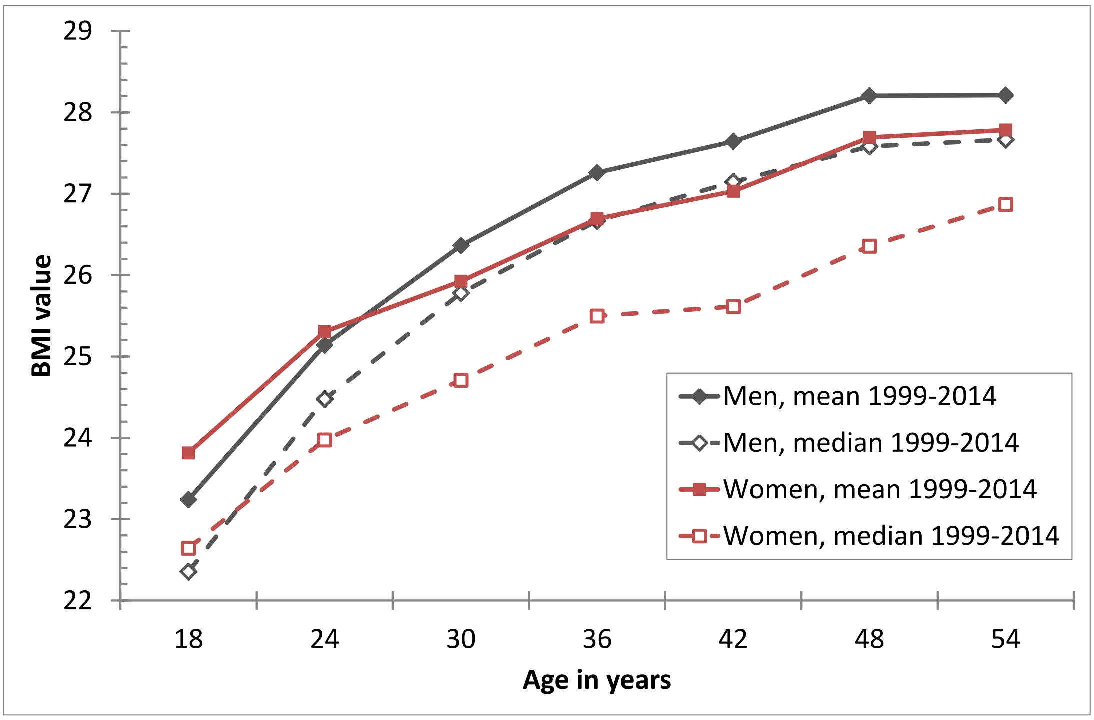


Figure 10.12 Mean and median for selected ages from 18 to 54 for 1999-2014.

10.3.2 Implications of the different changes with age

From the results presented in Section 10.3.1, aging at a population level affects the BMI of men and women in different ways. BMI tends to increase more evenly for men than for women. The shape of the BMI distribution does not change much for men above about age 24 and, as previously discussed in Section 8.4, this implies a similar tendency for BMI to increase for men irrespective of the starting BMI. For women, the changes are of BMI increasing more for those with a higher BMI than for those with a lower BMI.

The result of these differences is that at the older ages the healthy percentages are much lower for men than for women and the total prevalence of being overweight or obese for men is much greater (the total overweight and obese percentage is plotted in Figure 10.I.11 in Appendix 10.I). In that sense the general obesity issue is more widespread for men at the older ages. At the higher categories of obesity, though, the prevalence is greater for women and these are potentially the most serious cases for health risks.

As mentioned in Chapter 6, some of the literature indicates that variation in BMI within the population is considered to be a mixture of genetic and environmental effects. I am not sure how much the pattern of these different changes with age is due to human biology and innate physiology as the body matures and ages, and how much it is due to different variations in lifestyles and circumstances. It does imply, though, that different types of interventions may be needed for men and women.

A broader, more widespread, and consistent approach may be needed for men since BMI tends to increase with age to a similar extent for all starting BMI values and this ends up with a very high proportion being overweight or obese.

Perhaps a more targeted approach may be better for women since BMI tends to increase more with age for those with a higher starting BMI value. A greater percentage stay in the healthy category as age increases than for men. However, there is a greater prevalence than for men at older ages of very high BMI values.

Chapter 11: A modelling framework for constructing BMI scenarios

Key points from this chapter:

- A method is developed for modelling BMI scenarios. This involves modelling a BMI distribution for each year of age and combining these together into a distribution for the whole population.
- The model for each year of age is based on a lognormal distribution. For women this is the model used. For men a small adjustment is made to the lognormal using a cosine function to give a closer match of the shape with the actual data.
- The model uses parameter values for the mean and standard deviation for each year of age. It also needs a minimum value for each year of age. The version for men needs two further parameters for the cosine period and amplitude.
- The model was tested by trying to reproduce the past population distributions for 1993, 1997, 2001, 2005, 2009, and 2013 from Chapter 4. The mean and standard deviation values from the data were used for the model parameter values. The minimum values were based on the best values from matching the age distributions. The cosine parameters were fitted to the data.
- The model is able to match the past population distributions well. This gives confidence in using the model for generating future scenarios, which is done in Chapter 12.

11.1 Overview of modelling method

The analysis in previous chapters, particularly Chapter 7, looks at how BMI varies with age. Chapter 7 includes some suggested BMI aging profile curves. These can be used as the basis for constructing scenarios and calculating what the overall population BMI distribution would be given particular circumstances and assumptions. An aging profile curve gives a model of the mean BMI values for each age. If we also have standard deviation values and a distribution shape, then probability distributions can be produced for each age and combined together for the whole population. This chapter describes the analysis done for this approach, which found that the lognormal distribution (with certain adjustments) is suitable as the probability distribution for each age.

This gives a method for modelling the population BMI distributions as follows:

- Specify the mean and standard deviation for each year of age.
- Choose a probability distribution for modelling BMI for each year of age.
- Specify the parameter values required for the distribution using the mean and standard deviation, along with any other parameter values required. The main other parameter used here is the minimum value for the distribution.
- Model the distribution for each year of age using the probability distribution and the parameter values.
- Combine the modelled age distributions together to produce the overall population distribution. Other analysis such as age group distributions can also be produced if required.

The method was tested by modelling each of the population BMI distributions in Chapter 4 for 1993, 1997, 2001, 2005, 2009, 2013. As in Chapter 4, the populations are denoted by these years in this chapter for ease of notation. They are derived from HSE data for 1991-1994, 1995-1998, etc.

This chapter describes the work done for each element of the method. This is set out for the data for men in Sections 11.2-11.4 and for the data for women in Sections 11.5-11.7. The approach is essentially the same in both cases, although the age distributions for men use a cosine function adjustment which is not needed for the data for women.

11.2 Choosing the probability distribution for men

11.2.1 Lognormal probability distribution

The analysis described in Chapters 7-10 included producing many smoothed BMI distributions for age groups and for single years of age. From looking at the shapes of these distributions it was considered that a lognormal distribution could be a good model for BMI distributions for each year of age for both men and women. Therefore, as part of the analysis that was done, in each case a lognormal distribution was plotted and compared with the smoothed distribution.

A lognormal distribution means that the natural log of the data follows a normal distribution. The lognormal needs two parameters that together determine the scale and shape of the distribution. These can either be specified as the mean and standard deviation of the underlying normal, or as the mean and standard deviation of the lognormal, depending on the software used. There are relatively simple equations (given below) that convert between the two. Parameter values can be derived by maximum likelihood estimation or using moments. The ultimate aim here is to use the lognormal for modelling, where we specify the mean and standard deviation values for each year of age. Therefore, the moments approach was applied where the parameters use the mean and standard deviation for the actual distribution.

Therefore, when comparing a lognormal to an actual smoothed distribution for an age group or a single year of age, the mean and standard deviation from the actual smoothed distribution were used for the lognormal parameter values. This approach has the effect that the mean and standard deviation of the lognormal will match the values for the actual distribution. The analysis here was done in Excel and the lognormal formula in Excel requires the mean and standard deviation of the underlying normal as its parameters, using the equations below.

The other parameter that needs to be specified here is the minimum value. The nature of BMI is that there are very few values less than 10, for example. Therefore, having a positive minimum value is reasonable. Specifying a minimum value simply means that this value is added to the lognormal function value with the minimum being deducted from the mean used to specify the lognormal parameters. The minimum will affect the shape of the distribution, but the mean and standard deviation will still match the data. For modelling and comparing with a specific age group distribution, the minimum values were derived in Excel by using solver to choose the value that gives the best fitness value, as measured by minimising the sum of squared differences between the actual and the lognormal distribution. The choice of minimum values for modelling the historical distributions is described in Section 11.3.

The BMI distribution D for a particular year of age is the minimum value, a, plus the lognormal distribution function value, denoted $LogN$, with parameters for the mean μ and standard deviation σ of the underlying normal derived from the mean m and standard deviation s of the data. The resulting distribution D will also have a mean m and standard deviation s.

Writing this as equations, we have:

$$D \sim a + LogN(\mu, \sigma)$$
$$\mu = \ln\left((m-a)^2 / \sqrt{(s^2 + (m-a)^2)}\right)$$
$$\sigma^2 = \ln(1 + s^2/(m-a)^2)$$

11.2.2 Adjustment to the lognormal distribution for men

The full process for analysing the data and modelling the distributions was done for the data for men using the lognormal distribution. This was tested by generating the historical distributions from Chapter 4 for 1993, 1997, 2001, 2005, 2009, and 2013. The distributions that were generated matched quite well with the Chapter 4 distributions. However, there was a consistent pattern of difference where the modelled distribution using the lognormal had a slightly lower and wider peak than the actual distribution. The differences were quite small. As examples of the differences, for all six distributions the modelled distribution has healthy and obese I category percentages a little higher than the population with the difference on average being 1.26% and 0.84% respectively, with the overweight percentage being lower with the difference on average being 1.93%. This type of difference in shape was also generally apparent from the modelling of the various age distributions.

The difference between the actual and modelled distribution was plotted in each case. Plotting against a Z value (the number of standard deviations of the BMI value from the mean) the differences have a fairly consistent shape somewhat similar to a cosine curve. This implied that a small adjustment to the lognormal distribution using the cosine function could improve the fit. The function chosen was a cosine shape with a peak of a given amplitude A between $-\pi/2$ and $\pi/2$ radians, and two troughs of half the amplitude $A/2$ between $-3\pi/2$ and $-\pi/2$ radians, and between $\pi/2$ and $3\pi/2$ radians. There are two parameter values to choose: the amplitude A and the length of the cosine period measured in Z values. The BMI distributions are calculated in 0.5 BMI intervals and so the units used for A are the density per 0.5 BMI interval. The cosine period was specified by the Z value corresponding to $\pi/2$ radians and this is denoted P. Hence, the Z values for the cosine peak go from $-P$ to P.

To test the adjustment it was applied to the distributions for single year ages with six year gaps (18, 24, 30, etc.) from 18 to 72 for the time interval 1999-2014 (these distributions from age 18 to 54 are described in Section 8.5.2). The process was to first model the ages with the lognormal distribution, choosing the minimum value separately for each age that gives the best fit with the actual data. Then the cosine adjustment was added without changing the minimum values. The same values for the amplitude A and period P were used for all the age distributions. Excel solver was used to choose the values for A and P that give the best fit, and these were $A = 0.3427\%$ and $P = 0.6323$. The resulting cosine adjustment function is shown in Figure 11.1.

The adjustment function was applied by simply adding it to the lognormal distribution function. The function is symmetrical around the mean of the lognormal and so it does not alter the mean. The BMI distributions are calculated in 0.5 BMI intervals and so the adjustment made was to apply the total of the function value over each interval. This was obtained by integration, which is straightforward as the integral of the cosine function is the sine function. The adjustment function has positive and negative values with the total integral being 0 and hence the adjusted function frequency still sums to 100% without any other adjustment being needed.

The effect of adding the cosine function to the lognormal is to make the peak higher and narrower. This is because it adds some density around the mean and removes some away from the mean. Most of the BMI functions for men are fairly symmetrical with a small positive skew and so the mean is close to the peak value. Examples are given for ages 18, 36 and 54 in Figures 11.2-11.4. These each show the population distribution that is being modelled, the lognormal, and the lognormal plus the cosine. The cosine function adjustment is quite small but is effective in giving a better match with the target distribution. The biggest change is at the mean value where it adds the maximum amplitude. So, it adds about 0.34% to the distribution at this point, where the lognormal function value is typically about 5%. Hence, it typically increases the peak height by about 7% (= 0.34 / 5).

The benefit of adding the cosine function can be measured with the fitness function used here of the sum of the squared differences. Adding the function cannot make the overall fit worse as the amplitude could always be 0 and so would just be the original lognormal. However, since single values for A and P were chosen for all distributions the fitness for some could improve whilst others worsen.

The improvement obtained is substantial with the cosine adjustment reducing the total fitness value for all 10 distributions by 43%. The 9 distributions from age 18 to 66 all have large reductions in the fitness value of between 33% and 66%, with a reduction in the total fitness value for the 9 distributions of 53%. This, along with a visual inspection of the charts, shows that the adjustment is working well.

The highest age of 72 is the exception as its fitness value gets worse and more than doubles. This is the smallest sample size and the data has some quite large irregularities. It also has a wider peak than the other distributions and so the adjustment does not work particularly well for this specific case.

The improvement can also be seen in the BMI category statistics. The cosine function just affects the middle part of the distributions. It applies between BMI values of $m - 3Ps$ to $m + 3Ps$, where m and s are the mean and standard deviation of the distribution. For example, for age 36 in Figure 11.3, the cosine applies in the range of BMI from 19.0 to 35.5. Hence the changes are mainly in the healthy, overweight, and obese I percentages. These are all much closer to the actual values, with the mean absolute deviation for the 10 distributions for these three categories reducing by half.

The cosine function adjustment is effective in improving the ability of the lognormal distribution to model the BMI distributions for each age. The lognormal plus the cosine was therefore used as the distribution in the modelling methodology for men.

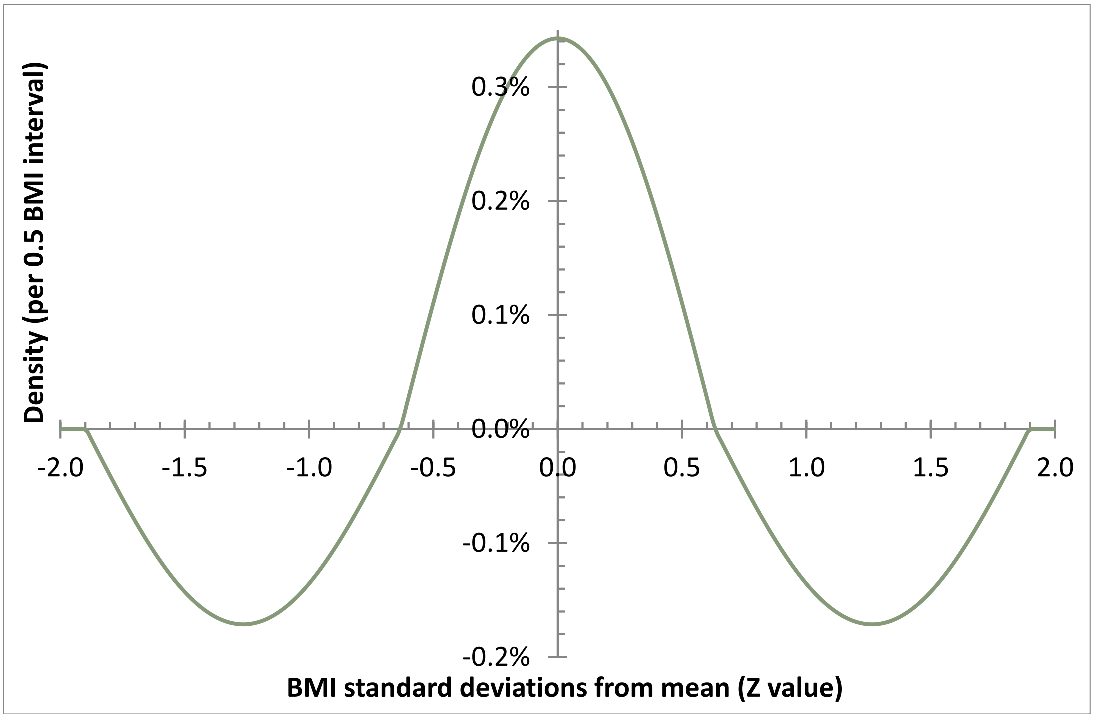


Figure 11.1 Cosine adjustment function used for the single year ages for men.

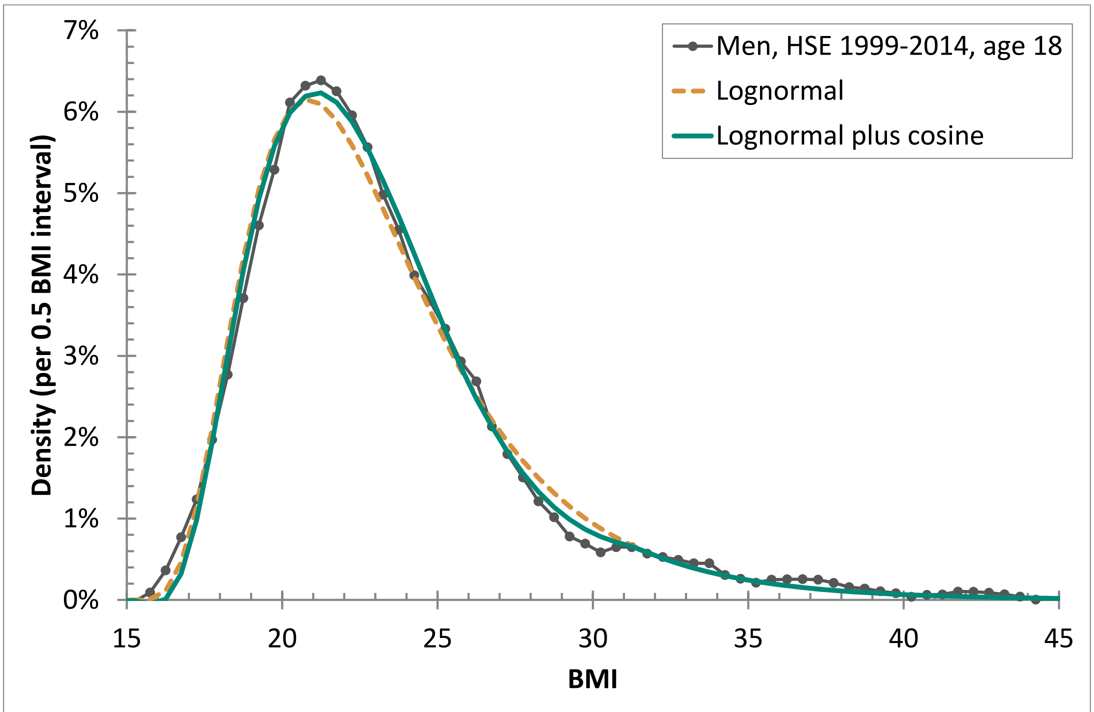


Figure 11.2 Lognormal distribution for age 18 for men for 1999-2014.

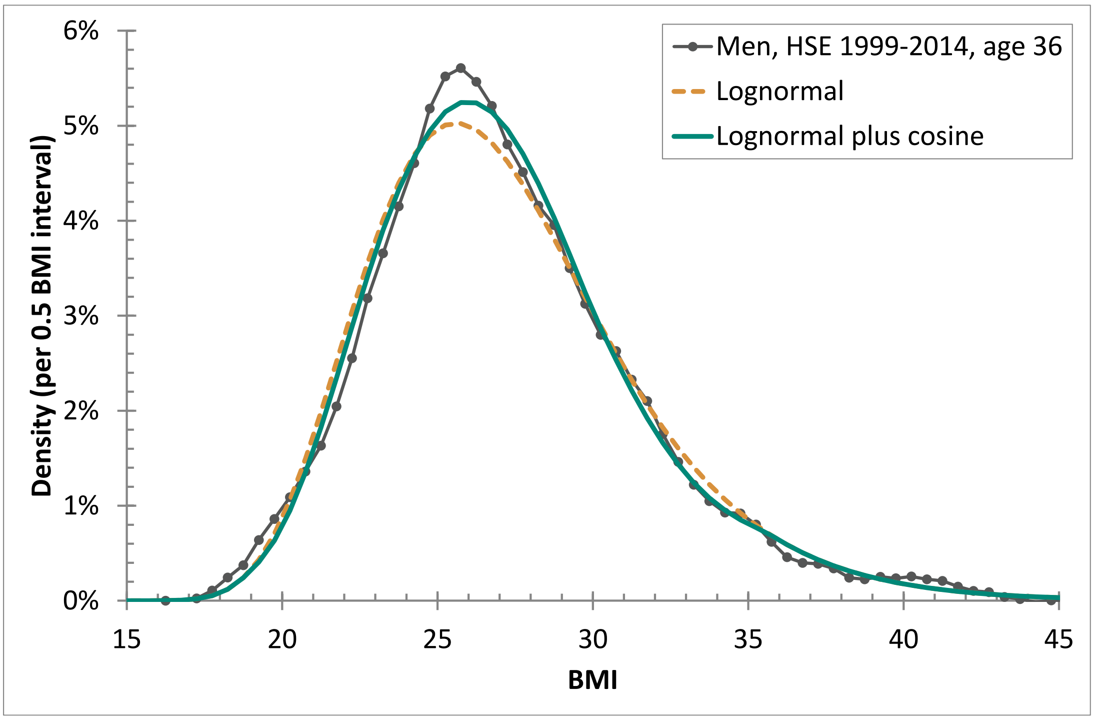


Figure 11.3 Lognormal distribution for age 36 for men for 1999-2014.

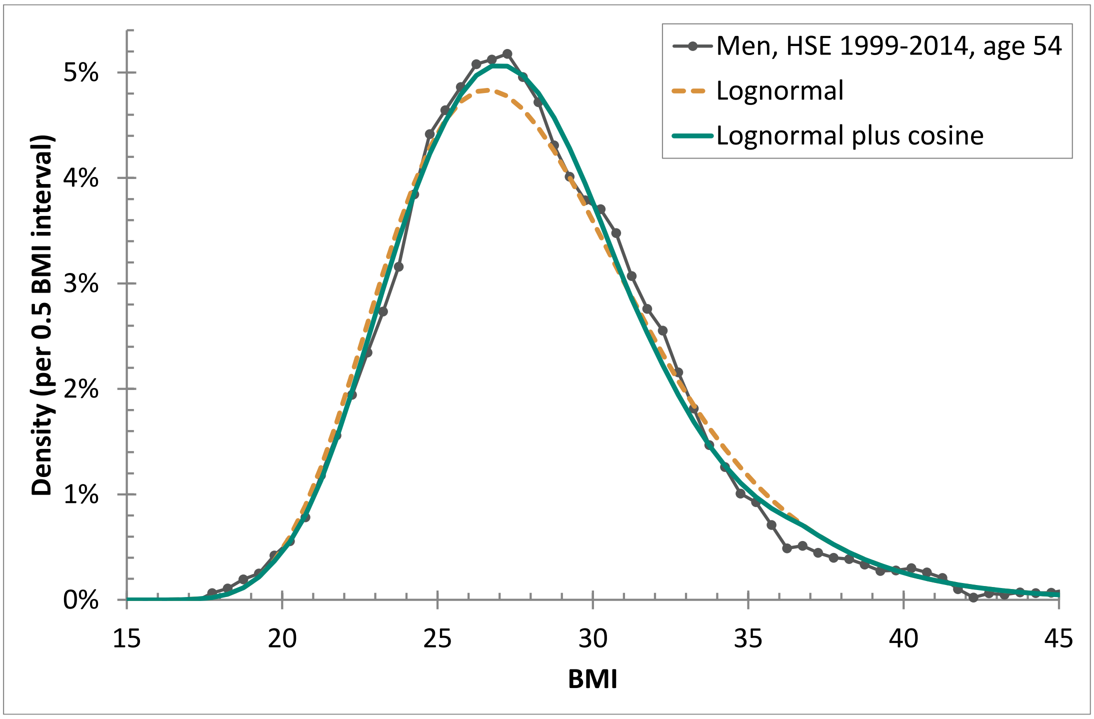


Figure 11.4 Lognormal distribution for age 54 for men for 1999-2014.

11.3 Choosing the minimum values for men

In order to model the historical distributions, minimum values need to be chosen. A lot of possible values are available from the analysis already done. This is because lognormal distributions were fitted each time a smoothed distribution for an age group was calculated from the data. In each case the mean and standard deviation values of the target distribution were used directly, and the minimum value was obtained using Excel solver by searching for the value that minimises the sum of squared differences between the distributions. The distributions analysed were:

- Each individual year of age from 18 to 90 for 1999-2014
- Each individual year of age from 18 to 90 for 1991-1998
- Age groups (Chapters 8 and 9) for 1991-1998, 1999-2006, 2007-2014

Some analysis was done more than once with alternative weightings although the effect of changing the weighting was always considered insignificant. For the individual years of age, the sample sizes are fairly small. Some older ages have very little data and so the results for these are not reliable. Some analysis was also done of age groups in four year time intervals but these are more variable than for eight years due to the smaller sample sizes (although gave similar results) and so the results were not used.

Generally, the different analyses gave quite consistent results for the mean, standard deviation, and minimum value. There is some variation from one value to the next for the individual ages, due to the small sample sizes.

For men, for ages up to about 60, most minimum values fitted for the individual ages are between BMI values of 10 and 14 for 1999-2014, and between 7 and 13 for 1991-1998. As age increases above that the minimum value tends to decrease with quite a few values of 0 for older ages (the minimum value was constrained in the solver search to be ≥ 0).

It was considered that it was best to use minimum values obtained from a good quantity of data so as to get reliable values. Therefore, the decision was taken to use the values for the age groups for eight year HSE time intervals. These minimum values were used when modelling both of the four year intervals corresponding to it. So, 2013 (i.e., 2011-2014) and 2009 (i.e., 2007-2010) both use the values from the age group analysis for 2007-2014. Similarly, 2005 and 2001 use the values from 1999-2006, 1997 and 1993 use the values from 1991-1998.

The exception was for age 18 since this distribution has a rather different more skewed shape to the other ages (for example, Figures 11.2-11.4 and Figure 8.14) and fits better with a higher minimum value. For this age, it was decided to use the minimum values from the individual distributions for 1999-2014 (a value of 14.63 applied to 2001, 2005, 2009, 2013), and for 1991-1998 (a value of 14.82 applied to 1993, 1997). The resulting minimum values used for modelling the historical distributions are shown in Table 11.1.

Table 11.1 Minimum values used for modelling the historical distributions for men.

Distribution year	1993	1997	2001	2005	2009	2013
Age 18	14.82	14.82	14.63	14.63	14.63	14.63
Age 19-24	12.49	12.49	12.01	12.01	12.90	12.90
Age 25-34	10.14	10.14	8.38	8.38	12.46	12.46
Age 35-44	10.46	10.46	10.45	10.45	12.76	12.76
Age 45-54	7.20	7.20	8.09	8.09	10.85	10.85
Age 55-64	6.92	6.92	7.71	7.71	12.33	12.33
Age 65-74	0.00	0.00	5.58	5.58	9.96	9.96
Age 75+	0.00	0.00	0.00	0.00	3.64	3.64

11.4 Modelling the historical distributions for men

11.4.1 Method used for the historical distributions for men

To test the modelling approach, it was applied to model the population BMI distributions for 1993, 1997, 2001, 2005, 2009, 2013 from Chapter 4. This is a validation test to see how well it can reproduce the historical distributions.

The method used was to generate values for each year of age and then combine them together to model the whole population. As described in Section 11.2, the lognormal plus cosine adjustment was used as the distribution for each year of age.

The lognormal part of this requires parameter values for the minimum, mean, and standard deviation. The lognormal is then modelled using the equations in Section 11.2.1.

The mean and standard deviation were calculated for each year of age from the actual HSE data and the case weightings described in Chapter 3. This is the data used to generate the population distributions in Chapter 4 with the Savitzky-Golay smoothing process. In this analysis the HSE data before smoothing was used since the smoothing is applied to the combined data for each 0.5 BMI interval and loses the links with the case values such as age. The HSE data was analysed to calculate the mean and standard deviation values for each year of age. For example, for the 2013 distribution the HSE data for 2011-2014 was used. The mean and standard deviation were calculated from this data for age 18, 19, 20, … , 89, 90 (which is 90+). These were then applied as the mean and standard deviation parameter values for the lognormal distribution.

The minimum value is also required and, as described in Section 11.3, the values for this were based on the age group distribution fitting and are given in Table 11.1.

The cosine element of the overall distribution has parameter values for the amplitude *A* and cosine period *P*, as described in section 11.2.2. Single values were used for the two parameters for all the ages from 18 to 90 for the particular distribution. These were obtained using Excel solver to minimise the sum of squared differences between the modelled distribution and the population distribution. The values used are given in Table 11.2. The cosine period values are about that same in each case at close to 0.6. The amplitude, though, tends to be larger for the older years indicating that the distributions are a bit further from a lognormal shape and so a bigger adjustment is needed.

Table 11.2 Cosine parameter values used for modelling the historical distributions for men.

Distribution year	1993	1997	2001	2005	2009	2013
Amplitude	0.5562%	0.4811%	0.4640%	0.3873%	0.2947%	0.3457%
Cosine period	0.5969	0.5762	0.5771	0.5539	0.6272	0.6235

The population distributions were then modelled by generating percentage frequency values for each year of age using the lognormal plus cosine distribution. As in previous analysis this was done by calculating the percentage frequencies in each 0.5 BMI interval. The values for each year of age were then combined together into a modelled population BMI distribution, using weightings of the population age profile percentages for each year of age that are also used for the BMI distributions in Chapter 4 (as explained in Section 3.3).

In the process for obtaining the population distributions in Chapters 3 and 4, all the data values across all ages were combined together to give percentage frequencies in each 0.5 BMI interval. There was then a further step of smoothing to produce the overall population distribution. Therefore, the shape of the distribution comes from the actual data, along with the smoothing. The Savitzky-Golay smoothing method used (Section 3.5) maintains the mean and standard deviation so that these are the same as the data distribution.

By using the actual mean values here, the mean of the modelled distribution will inevitably match the population distribution (apart from very minor differences due to grouping the data into 0.5 BMI intervals). The actual standard deviation values are also used. The standard deviation of the modelled distribution will be affected a little by the cosine adjustment but should still be very close to the population standard deviation.

The test in assessing the quality of the model is therefore whether the shape of the modelled distribution provides a good match with that of the population distribution. In the model, the shape is produced by the combination of all the lognormal plus cosine distributions for each year of age. In the population distributions in Chapter 4, the shape is produced by the whole data (the relative frequency in each 0.5 BMI interval) along with the smoothing process.

Given that the distributions for each year of age match the data quite well (e.g., Figures 11.2, 11.3, 11.4), then it can be expected that model should produce a realistic shape. However, it is important to compare the model and the actual distributions to test this. The results are in the next section.

11.4.2 Results for the historical distributions for men

The model was able to match the historical distributions very well for each of the six distributions. Figure 11.5 shows the distributions for 2013, whilst the charts for all the years are in Appendix 11.I. The fitness function that was used as a measure of the match between the distributions is the sum of the squared differences in the relative frequencies for each 0.5 BMI interval up to a BMI of 80 (i.e., basically the whole distribution). These are perhaps most useful in giving a relative measure of how well the models match the population distributions. The values for each of the six years from 1993 up to 2013 are, in units of 10^{-5}: 2.63, 0.74, 3.69, 0.97, 2.32, 1.56. The model has the lowest fitness value in 1997, as also evident by the very close correspondence of the distributions in Figure 11.I.2 in Appendix 11.I.

For the model to be effective in this situation it is important that it represents the changes over time well. The healthy category percentages are shown in Figure 11.6. The model values are close to the population values. They are slightly higher in each case, but by only 0.40% on average (compared with 1.26% for the model without the cosine adjustment). The descriptive statistics including all the BMI categories are in the tables in Appendix 11.I. A couple of categories have a consistent direction for the differences, with the underweight percentage values all being lower for the model than for the population for each year and the healthy category percentages all being higher. The other categories all have some values higher and some lower. However, the mean absolute differences for each category across the six years are all very small: underweight 0.27%, healthy 0.40%, overweight 0.09%, obese I 0.24%, obese II 0.10%, severely obese 0.10%. Hence, the statistics for the models and the distributions match well.

In addition, a comparison was made as to how well the model could represent age group distributions. This was done for the data for 2007-2014. This again used the case weightings from the analysis in Chapters 3 and 4. Age group population distributions were produced for the usual age groups of 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+. These are basically the same as those in Figure 8.5. The weightings here are different but this makes very little difference to the results. The lognormal model uses the actual mean and standard deviation values for each year of age from the HSE data for 2007-2014. It also uses a constant minimum value for all the ages in each age group, which are the optimum values from fitting a lognormal to the actual age group distributions (basically the same values as in Table 11.1 for 2009 and 2013, although not rounded here and using the 19-24 value also for age 18). The cosine factor values that were used are the average of the 2009 and 2013 values in Table 11.2. The age group distributions were produced as a weighted average of the values for each year of age weighted by the population age profile percentages. Charts and statistics of the actual and model distributions are given in Appendix 11.II. The way that the model is constructed ensures that the mean BMI values will be basically the same as the data for each age group (just very minor differences of 0.01). Comparison of the shapes of the distributions and the category statistics show a good match and so the model is again working well. For example, the mean absolute deviation across the seven age groups is less than 1% for every BMI category.

Overall, the strong match obtained with the population distributions gives good confidence in using the model for future scenarios, which is done in the next chapter.

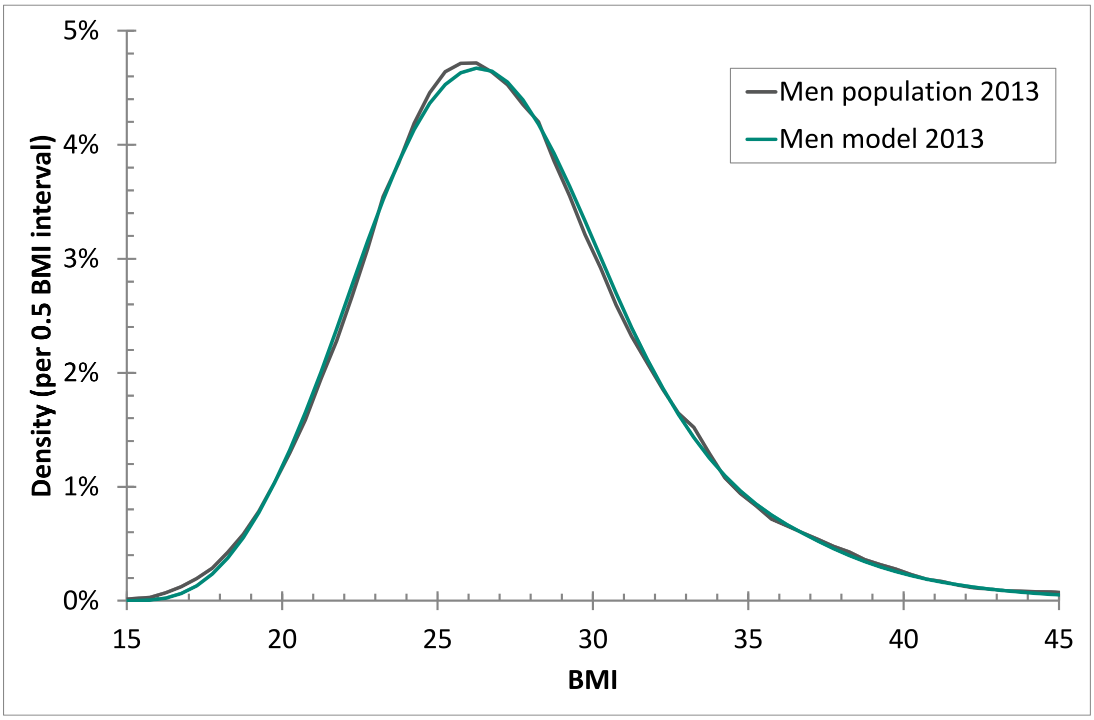


Figure 11.5 Population BMI distribution and modelled distribution for men for 2013.

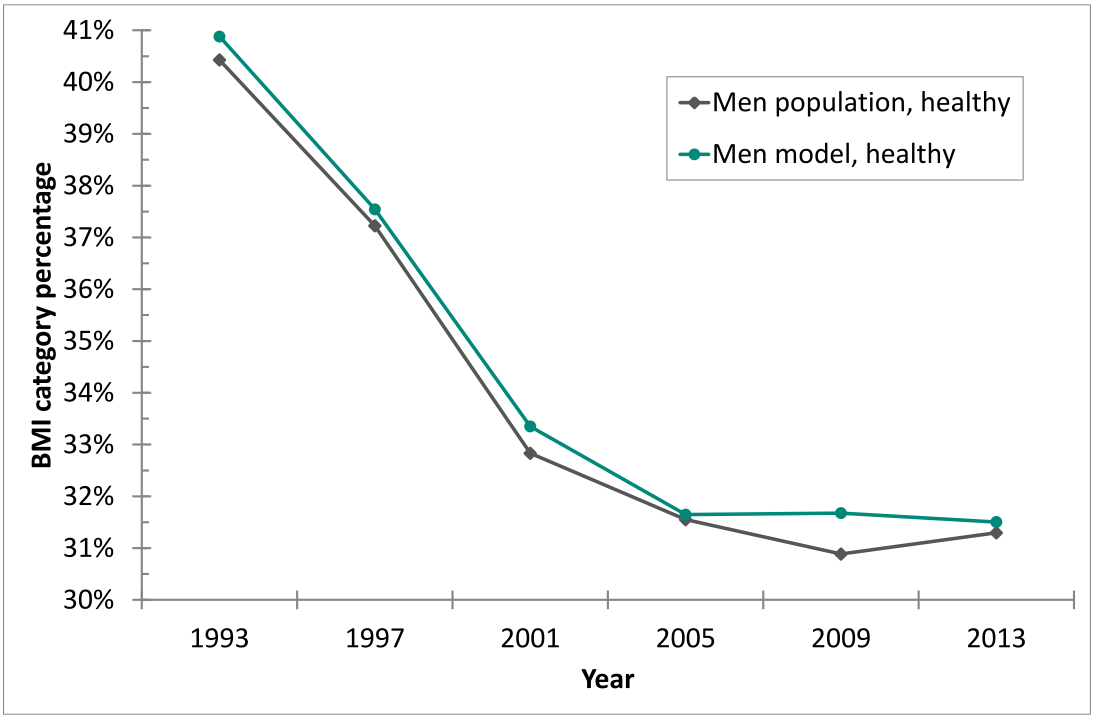


Figure 11.6 Healthy category percentages for population and modelled distributions for men.

11.5 Choosing the probability distribution for women

As explained in Section 11.2.1, lognormal distributions were fitted as part of the range of analysis done in producing the various smoothed age BMI distributions. For the lognormal parameters, the mean and standard deviation from the actual distribution were used and minimum was chosen by optimising the fitness function of the squared differences from the actual distribution.

Overall, the lognormal works well for modelling the distributions for women. Examples for ages 18, 36 and 54 for 1999-2014 are in Figures 11.7, 11.8, and 11.9. For a few of the younger ages data sets, such as age 18 in Figure 11.7, there is a slight difference in pattern where the lognormal peak is to the left of the peak of the actual data. This applies more to the 1991-1998 data than to 1999-2014. However, for most of the single year ages and also for the age groups the lognormal is a very good model and so this was chosen with no adjustments.

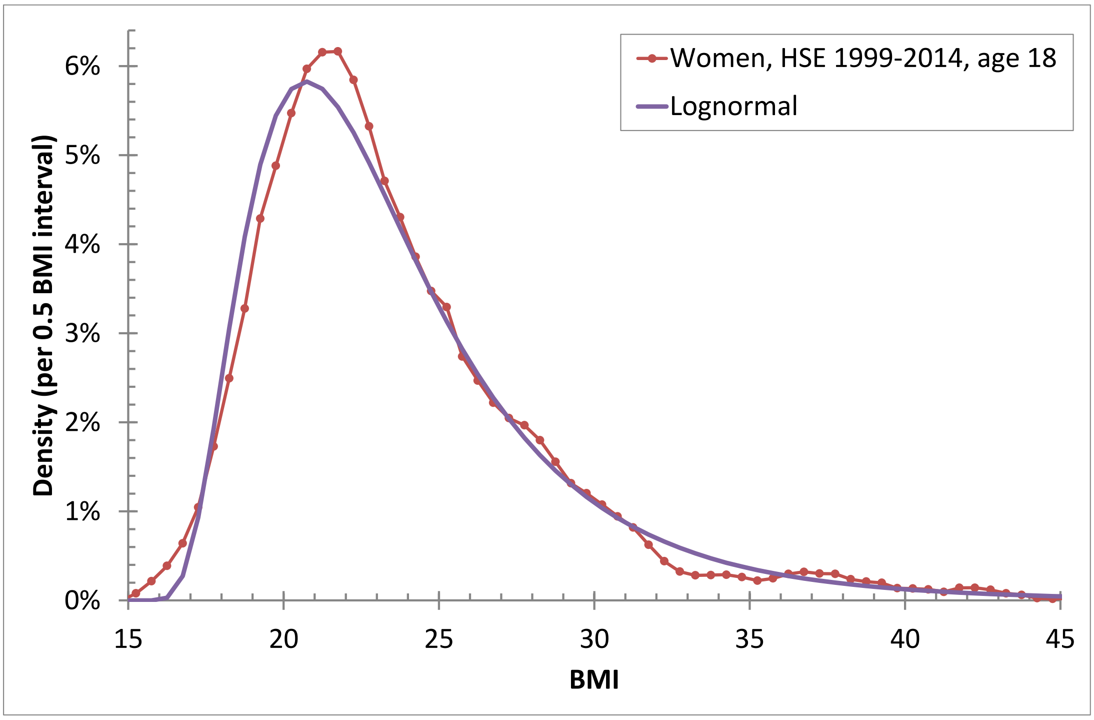


Figure 11.7 Lognormal distribution for age 18 for women for 1999-2014.

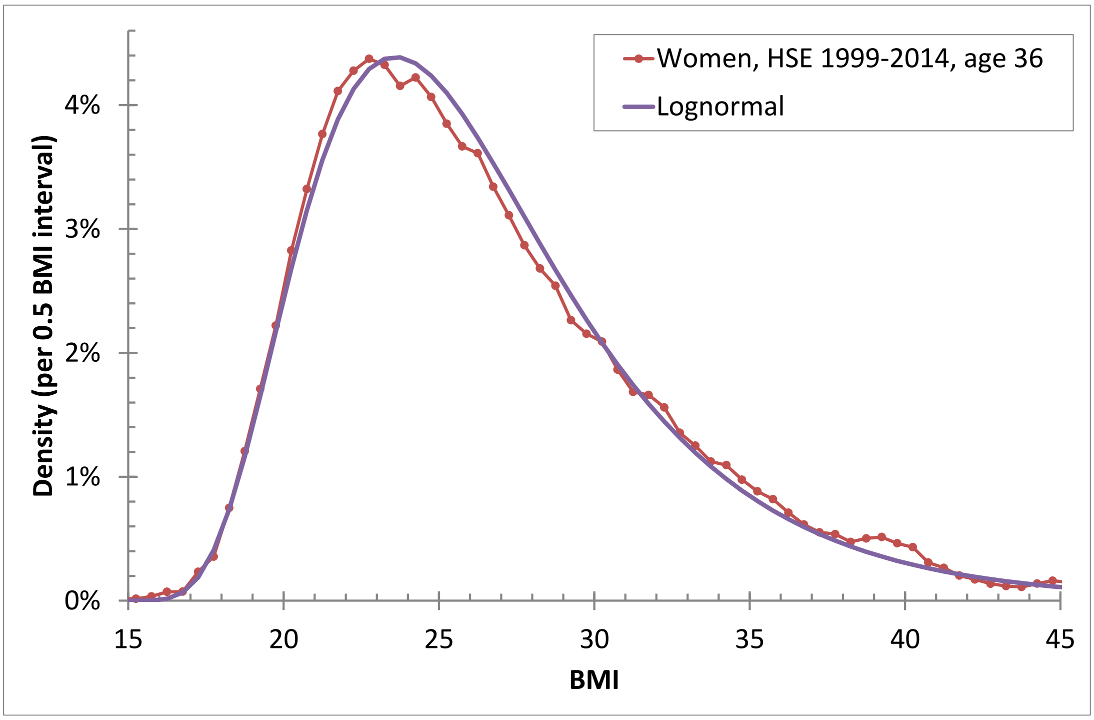


Figure 11.8 Lognormal distribution for age 36 for women for 1999-2014.

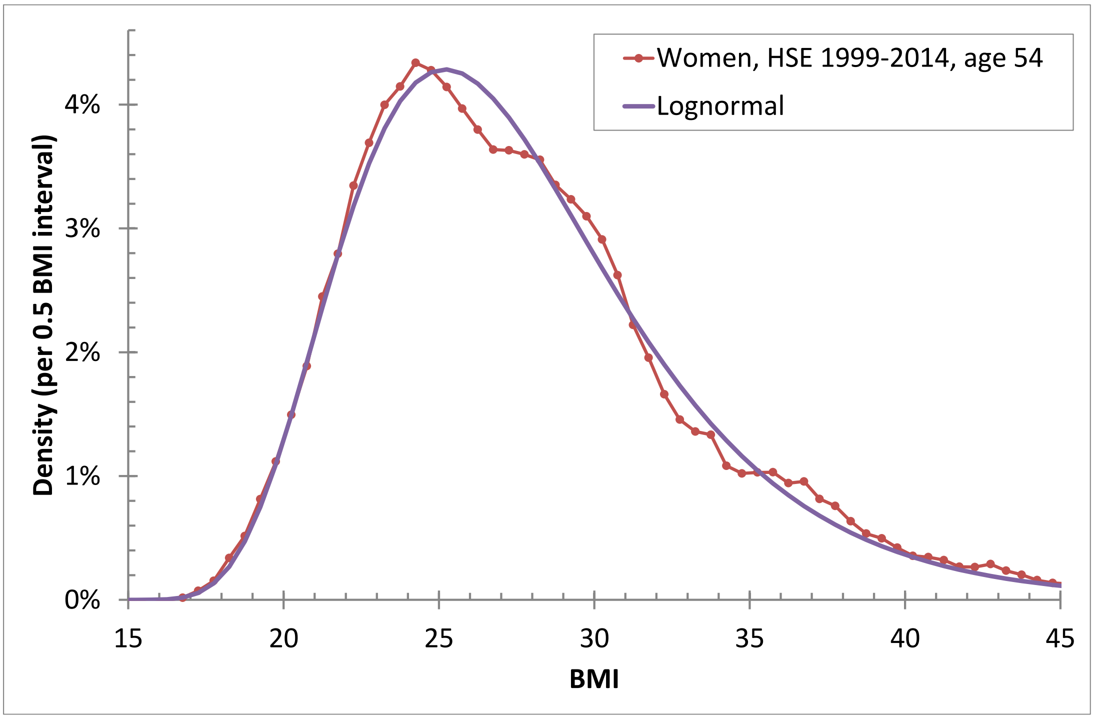


Figure 11.9 Lognormal distribution for age 54 for women for 1999-2014.

11.6 Choosing the minimum values for women

Minimum values need to be set for the model. The model was used first for producing the historical distributions (Section 11.7), and so values needed to be chosen for this. As explained in Section 11.3, minimum values were calculated using Excel solver when fitting lognormal distributions for:

- Each individual year of age from 18 to 90 for 1999-2014
- Each individual year of age from 18 to 90 for 1991-1998
- Age groups (Chapters 8 and 9) for 1991-1998, 1999-2006, 2007-2014

The age group analysis was done with both the weightings from Chapters 7-10 and from Chapters 3-4 although this made little difference (the biggest effect being slightly lower mean BMI values by about 0.4 for the Chapter 3 and 4 weightings for age 75+). All the analysis gave generally quite consistent results for the minimum values. There is naturally some variation between each individual year of age given the relatively small sample sizes but the minimum values are similar to those for the age groups.

As was the case for the data for men, the chosen approach was to use the minimum values from the age groups as they are based on relatively large amounts of data. The values are taken from the age group analysis using the Chapter 3 and 4 weightings (since the comparison is with the population distributions from those chapters) and are shown in Table 11.3. Unlike the minimum values for the data for men, the minimum values for age 18 were not especially different to those for other ages and so separate values were not used for this age.

Table 11.3 Minimum values used for modelling the historical distributions for women.

Distribution year	1993	1997	2001	2005	2009	2013
Age 18-24	15.05	15.05	15.39	15.39	14.49	14.49
Age 25-34	15.78	15.78	15.23	15.23	15.17	15.17
Age 35-44	15.82	15.82	15.20	15.20	14.57	14.57
Age 45-54	15.25	15.25	14.27	14.27	14.11	14.11
Age 55-64	13.08	13.08	13.27	13.27	12.25	12.25
Age 65-74	4.99	4.99	8.71	8.71	10.70	10.70
Age 75+	0.00	0.00	0.00	0.00	3.33	3.33

11.7 Modelling the historical distributions for women

11.7.1 Method used for the historical distributions for women

As described for the data for men in Section 11.4, the modelling approach was tested by trying to reproduce the population BMI distributions for 1993, 1997, 2001, 2005, 2009, 2013 from Chapter 4.

Each year of age was modelled using the lognormal distribution with the minimum values from Table 11.3 and the mean and standard deviation values from the HSE data. The distributions for each year of age, in 0.5 BMI intervals, were combined together with weightings of the population age percentages from Section 3.3. This produces a population distribution that can be compared with the corresponding Chapter 4 distribution.

11.7.2 Results for the historical distributions for women

The model does well in reproducing the Chapter 4 distributions. The modelling approach ensures that the mean and standard deviation will basically be the same as the target distributions. The shape is matched well and Figure 11.10 shows the results for 2013. All the distributions are shown in Appendix 11.III. The model's performance is similar for each of the distributions from 1997 to 2013. For 1993 there is a bit more of a difference with the model having a slightly lower peak and the peak extending a bit more to the left compared to the population distribution. The fitness values of the sum of squared differences of the interval frequencies are, in units of 10^{-5}: 6.47, 2.66, 1.71, 2.06, 2.26, 2.47. Apart from 1993, these are at a similar level to the fitness values for the model distributions for men (Section 11.4.2).

The category percentages were also compared as a measure of how well the model represents the changes over time. The healthy category percentages are shown in Figure 11.11. The model values are close to the population values. The mean absolute differences for each of the BMI categories across the six years are all small: underweight 0.43%, healthy 0.24%, overweight 0.47%, obese I 0.24%, obese II 0.40%, severely obese 0.09%. All the statistics are given in Appendix 11.III.

Age group distributions were also calculated and compared for the time interval 2007-2014 in the same way as for the data for men (Section 11.4.2). This also uses the case weightings from Chapters 3 and 4. The model lognormal distributions for each year of age use the actual mean and standard deviation from the HSE data and the minimum values from Table 11.3. They are combined together into age group distributions using weightings of the population age profile percentages. Again, the model works well in matching the data. Charts and statistics are in Appendix 11.IV. There are seven age groups and the mean absolute deviation of each category percentage can be calculated and averaged across the age groups. The highest value is the overweight category at 1.19%, whilst all the others are 0.50% or less.

As for the model for men, the model here does well in matching the population distributions. It is therefore considered suitable for generating future scenarios. Some scenarios are explored in Chapter 12.

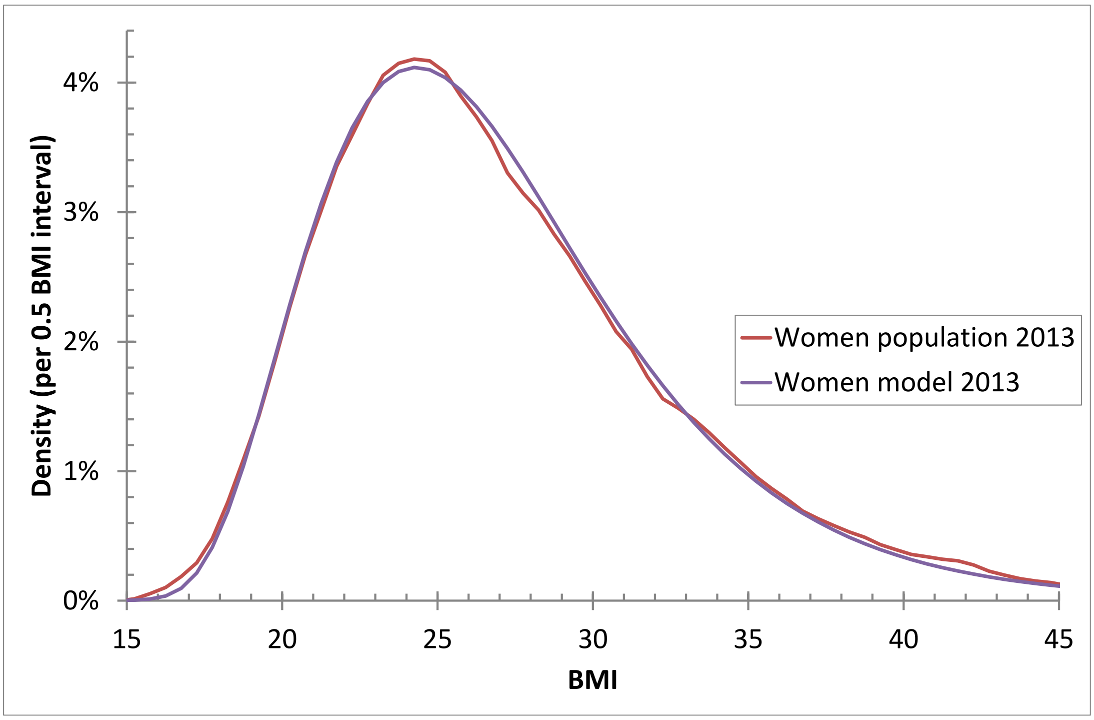


Figure 11.10 Population BMI distribution and modelled distribution for women for 2013.

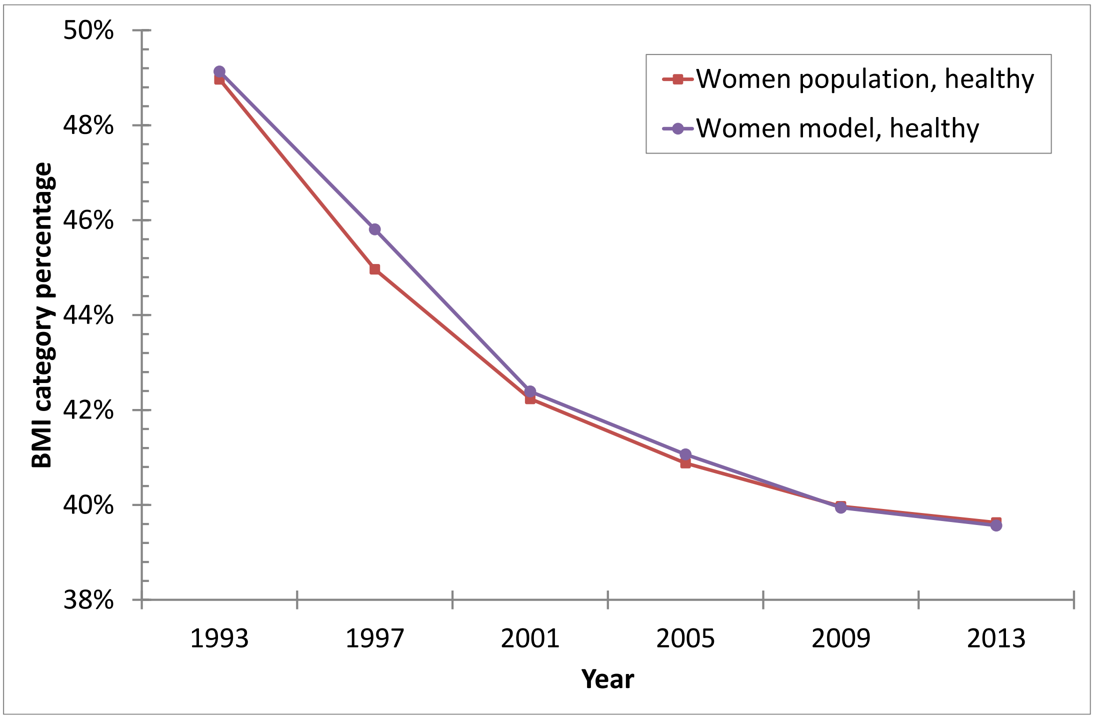


Figure 11.11 Healthy category percentages for population and modelled distributions for women.

11.8 Nature and advantages of the modelling approach

The modelling approach set out in this chapter has a detailed structure. It is based on using a lognormal distribution for the BMI of each year of age from 18 upwards (along with a cosine curve adjustment for men). The lognormal distribution is shown to work well in matching the historical data for the distributions for single years of age.

A scenario is constructed by specifying the lognormal parameters for each year of age, which are the mean, standard deviation, and minimum value (along with the cosine period and amplitude for the cosine function adjustment for men). This chapter has shown that this modelling approach is able to reproduce the historical population BMI distributions well and this gives confidence in the validity of the approach.

This approach does require a lot of parameter values to be chosen. An important advantage of this is that the assumptions made in constructing any scenario are clear. In particular, the shape of the BMI distribution for each year of age is specified and the way that the parameters vary with age, especially the BMI mean and standard deviation, are also set out. This enables clear comparisons to be made between the parameters used in a scenario and any historical data, and hence the specific changes that are being assumed in the scenario. It will also help in identifying the reasons for any differences between a scenario generated now and actual BMI distributions in the future. This is done in Chapter 14 in Section 14.8 where the distributions for the "current trend scenario" in Chapter 12 are compared with the actual distributions for 2017 obtained from newer data.

Other scenario modelling approaches could be used. For example, a simpler approach could be to project past data such as trends in the BMI category percentages. One difficulty with this is deciding the shape of the curves of how the percentages might change in the future. A category percentage is a combination of different proportions from different ages in the population and it is hard to assess how this might change over time when considering it just as a single value, without taking account of how BMI varies with age and how this has changed over time (Chapters 7-10). There are interactions between the categories in that changes in BMI will effectively move people between the categories and again it can be difficult to evaluate the combined effect of an increase in BMI moving some people into the category from a lower one, and some people out into a higher one. One way that this could perhaps be done would be with a linear scaling transformation, given that the changes in BMI in the past have followed this pattern (Chapters 4 and 5). Another issue is that, as discussed in Section 1.2.4, the categories are broad ranges and so changes in the category percentages will depend on exactly how BMI is distributed within the categories. The consequence is that it would be quite hard with such an approach to understand the precise assumptions being made and to assess how plausible a scenario is.

Scenarios can be put together for various purposes. One purpose can be to provide an estimate of future BMI distributions given an assumption of future conditions. The next chapter includes an example of this with a "current trend scenario". This is based on a general assumption that conditions stay the same regarding the obesity factors (Chapter 6) as they were in the most recent HSE years used in the work here. In particular, this assumes that mean BMI will follow the aging model curve from Chapter 7, which is based on projecting the mean values of the youngest cohort using the general pattern of older cohorts.

Another purpose can be to identify what is required to achieve a better population BMI distribution. This is the other example used in Chapter 12 and is called the "low BMI scenario". This scenario is constructed to give an approximate match to the 1993 BMI distributions. It demonstrates one way of getting back to those much better distributions. The parameters for the individual BMI age distributions could then be used as targets to aim for.

Chapter 12: Future scenarios

Key points from this chapter:

- The model from Chapter 11 is applied to generate two example future scenarios. A scenario is specified by values for the mean, standard deviation, and minimum for each year of age. Parameter values for the cosine adjustment are also required for the model for men.
- These are scenarios rather than specific forecasts. They apply a given pattern of parameter values, particularly the mean BMI value for each age, to produce an overall population BMI distribution. The purpose of the scenarios is to calculate the BMI distribution that would result from a particular set of parameter values, given the assumptions made in the model structure.
- What actually happens in reality in the future will depend on the many factors that affect BMI, such as those discussed in Chapter 6. This will include the effectiveness of any public health measures designed to try and reduce the prevalence of obesity.
- One scenario investigated is a "current trend" scenario, which applies the aging model curves from Chapter 7. For men and for women, this curve for mean BMI was specified in Chapter 7 on the basis that the age cohorts appeared to be moving towards the curve. It therefore represents an estimated long term BMI age profile if this is the case and if conditions remain the same as current conditions. The curves are slightly above the actual data used for producing the 2013 BMI distributions. Consequently, the scenario results give a slightly worse situation for BMI than the 2013 distributions. For men, the scenario has a mean BMI value of 27.87 which is 0.46 higher than for 2013, and a healthy category percentage of 28.3% which is 3.0% lower than for 2013. For women, the scenario has a mean BMI value of 27.50 which is 0.37 higher than for 2013, and a healthy category percentage of 37.1% which is 2.5% lower than for 2013.
- The other scenario examined is a "low BMI" scenario. The aim with this scenario was to apply a suitable asymptotic mean value curve from Chapter 7 to generate population BMI distributions similar to 1993. For both men and women, the 3rd curve from Chapter 7 (Section 7.5) was used, along with parameter values for the standard deviation and minimum similar to those used for modelling 1993. Very similar distributions to 1993 were produced. The scenario therefore indicates one way of getting back to 1993 BMI levels. The mean and standard deviation values for each age could therefore be used as target values. The benefits would be considerable. For example, the healthy category percentages in this scenario are 8.5% higher than in 2013 for men, and 8.3% higher for women. There are considerable reductions in each of the obese category percentages.

12.1 The nature of the scenarios

Chapter 11 established a modelling approach for generating BMI distributions. The basis is to use separate distributions for each year of age, based on the lognormal. These are then combined together into distributions for the whole population or into age group distributions. The main parameter values required are the mean and standard deviation for each year of age. Values are also needed for the minimum value for each age. For the model for men a small cosine function adjustment is made to the lognormal and values for the cosine period and amplitude also need to be chosen for this.

Scenarios are set up by specifying values for the parameters. In particular, by choosing how the mean and standard deviation of BMI each vary with age. The model then generates the distributions.

Any pattern could be used for a scenario. Scenarios can therefore be used to explore a wide variety of different situations. This can include patterns representing what might happen based on current BMI trends, and patterns representing what might happen if the situation improves.

These are scenarios rather than forecasts. What actually happens in the future depends on the changes over the long term in the many factors affecting obesity, as discussed in Chapter 6. For example, there could be major changes in the food industry, economic conditions, work and leisure activities, social attitudes and norms, etc. In particular, this is not a passive situation but is one in which national and local government, and public health bodies are trying to change the situation by developing and supporting interventions to reduce obesity. The overall aim is to reduce the prevalence of adverse health conditions related to obesity and thus improve public health. If these initiatives are successful then this will alter current trends and reduce BMI values.

12.1.1 Current trend scenario

I have called the initial scenario the "current trend scenario". This uses the mean BMI values from the aging model (upper asymptotic curves) for men and for women from Section 7.5. These curves are a subjective estimate of the trajectory that the youngest adults are on and that older adults are moving towards. As described in Chapter 7, it was constructed by following the data of the youngest cohort and then continuing the curve based on the shape of the patterns of older cohorts.

The interpretation of the cohort data suggested in Chapter 7 is that the effect of recent modern conditions is to move the population towards this curve. For men and for women, it therefore represents a subjective estimate of what the population mean BMI values will be in the long term if conditions stay the same as they were in the recent HSE years used here.

Each cohort is generally at a higher level than older cohorts were at the same age (Figures 7.5 and 7.6). The assumption with the curve is that the youngest cohort will follow the general shape of the data for older cohorts but at a higher level since they are starting at a higher level. In that sense it is a slightly pessimistic scenario. The curve is close to the most recent data for the three youngest cohorts (right ends of the blue, red and green data series in Figures 7.5 and 7.6).

This scenario is also based on new adults in the future following the same path. No account is taken of how obesity or BMI has changed in children in recent years as this is outside the scope of this study. This is a long term scenario in that it will obviously take a long time for the youngest cohort to follow the full curve through to old age.

Of course, if the obesity factors change significantly in the future for better or for worse then BMI could be much higher or lower. As explained in Section 12.1 this is not a forecast but an exploration of what the BMI distribution would look like given this curve and the assumptions of the model.

Values also need to be chosen for the other parameters, including the standard deviation and minimum values. Both of these affect the shape of the BMI distributions and hence the proportions in the BMI categories. The model values were chosen based on the values for the most recent populations and are explained in Sections 12.2.1 and 12.4.1.

One possible use of the current trend scenario is as a benchmark to compare with future data. This is actually done in Chapter 14 in Section 14.8, where the 2017 distributions (from the 2015-2018 HSE data) are compared with the current trend models from this chapter. The 2017 distributions were produced after the current trend model and so this makes it an interesting comparison between the current trend model and the next distribution in the future.

12.1.2 Low BMI scenario

The "low BMI scenario" gives an example of how the model can generate alternative scenarios. The aim of this scenario is to model an improved situation where BMI values are at a similar level to the early 1990s. It gives an optimistic situation and could provide a target to aim for. It is potentially achievable in the sense that the BMI distribution existed only about 30 years ago, although it would be challenging given the big increase in BMI since then.

The scenario uses mean and standard deviation values similar to those in 1991-1994. It uses a mean BMI curve of the asymptotic shape from Chapter 7 on the basis that this is considered a realistic trajectory for modern lifestyles. This is slightly different to the curve of actual BMI values for 1991-1994.

With the design of the scenario the resulting distributions for men and women are inevitably similar to the 1993 population distributions. The results are therefore described fairly briefly.

12.2 Current trend scenario for men

12.2.1 Parameter values for the current trend scenario for men

This scenario uses the mean values from the upper asymptotic curve in Figure 7.5. The scenario can be compared with the most recent population BMI distribution from Chapter 4, i.e., the 2013 population generated from the 2011-2014 HSE data.

Figure 12.1 shows the mean values along with the mean values from the 2011-2014 data (using the weightings from Chapters 3 and 4). The model values are similar to or just slightly above the data values for most of the ages up to age 55. They are quite a bit higher for the oldest ages although those ages are a relatively small proportion of the population. For example, with the population profile that is used here (Section 3.3), ages 80 and over are only 4.6% of the population and ages 70 and over are 13.3%. Based on the values in Figure 12.1, the current trend model BMI distribution can be expected to have a higher mean than the 2013 population but with only a small difference.

It could be realistic to assume that BMI decreases a little in older ages. One factor that might cause this is those with a lower BMI tending to be healthier and living longer. Reduction in muscle mass and bone density for older people might also reduce BMI. The curve could be adjusted to have this pattern but it was decided to leave it as it is for this scenario.

The standard deviation values chosen are based on the 2011-2014 data since the scenario is trying to reflect recent conditions. Figure 12.2 shows the standard deviation values. The range of most interest in the data is from ages 18 up to about 55 since, from Figure 12.1, the current trend model is close to the actual data over those ages. There is some variation from one year of age to the next but for this age range the standard deviation values are similar with no indication of a trend. Regression applied to the data up to age 55 has a *p*-value of 0.75 and so a linear trend is certainly not statistically significant. The average of the standard deviation values up to 55 is 4.68. On this basis a value of 4.7 was chosen for the standard deviation for all ages in the current trend model.

The standard deviation obviously tends to get lower for older ages in the data. However, in the current trend model the assumption is that the older ages in the future will continue the pattern of the current younger cohorts. In Figure 12.1 the mean does not change much in the model above 55 and the assumption is that the age distributions will also be a similar shape. Therefore, the standard deviation value of 4.7 was applied to all the ages in the model.

The model also requires minimum BMI values for each age. Again, the approach was to base the values on the recent data. The minimum values used for modelling the 2007-2010 and 2011-2014 populations based on the age group distributions for 2007-2014 (apart from age 18 which is based on the single year distribution) are given in Table 11.1 in the previous chapter. These are a value of 14.63 for age 18 and then values between 12.90 and 10.85 up to age 55 (and smaller values above that). The average value for ages 19 to 54 is 12.17. A different value is considered to be needed for age 18 but the other values and distributions are assumed to be similar. Therefore, rounding to 1 decimal place, the minimum values used for the scenario are 14.6 for age 18 and 12.2 for all other ages. In the experimentation with the model the effects of changing the standard deviation and minimum values were considered and this is discussed in Section 12.2.3.

In the model for men there is also the cosine adjustment. The parameter values used for this are those for modelling the 2013 distribution shown in Table 11.2. The values are 0.6235 for the cosine period and 0.3457% for the amplitude. The cosine adjustment reduces the standard deviation slightly and the standard deviation values for each age end up being about 4.61.

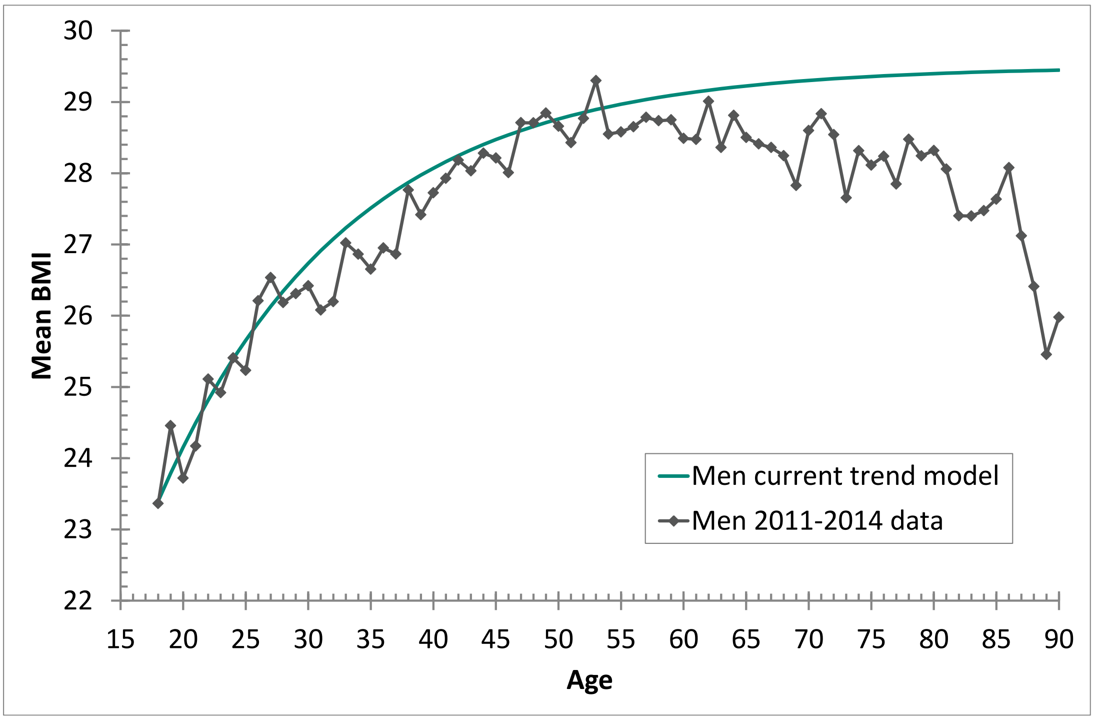


Figure 12.1 Mean values for the current trend model and 2011-2014 data for men.

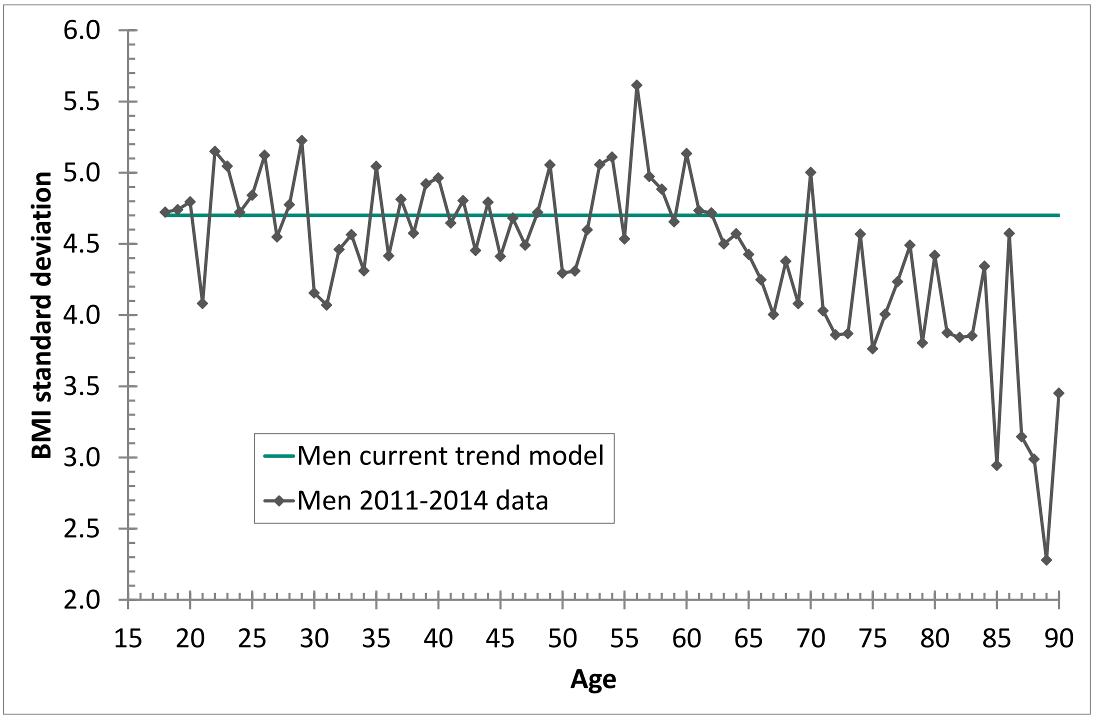


Figure 12.2 Standard deviation values for the current trend model and 2011-2014 data for men.

12.2.2 Results for the current trend scenario for men

The model was applied using the parameter values specified in Section 12.2.1. This generated lognormal distributions with the cosine adjustment for each year of age with a percentage frequency for each 0.5 BMI interval. These were then combined together into a complete population BMI distribution using the population percentages explained in Section 3.3 and that were applied for the distributions in Chapter 4. The resulting BMI distribution was compared with the 2013 population distribution from Chapter 4. Figure 12.3 shows the two distributions and Table 12.1 shows the statistics.

In Figure 12.3 the model distribution shows an increase in BMI compared with the 2013 population. The general shape of the distribution is similar but shifted a little to the right as well as a slightly lower and broader peak. The mean BMI value is 0.46 higher than for 2013 but with similar values for the standard deviation and skewness. In a historical context this is similar to the largest four year increase in mean BMI between the Chapter 4 distributions from 1997 to 2001 of 0.49. By contrast the most recent change in mean BMI between 2009 and 2013 is only 0.03. However, the current trend scenario is a long term one and is not particularly intended to represent changes that might happen within a few years.

The change in the BMI category percentages in Table 12.1 are shown in Figure 12.4. The healthy percentage is 3.0% lower. The main increases are 2.4% in obese I and 0.9% in obese II. Again, this is a reasonably large change in the context of recent history. For the Chapter 4 populations (Table 4.1) the healthy percentage fell the most by 4.4% from 1997 to 2001. It actually increased by 0.4% most recently from 2009 to 2013.

For reference, Table 12.1 also shows the values for the 2013 model from Section 11.4.2. The model fitted the 2013 population well but some of the changes from the 2013 population values will be due to the minor limitations of the model. In particular the 2013 model only has 0.8% in the underweight category compared to 1.2% in the population distribution. The current trend model also has 0.8% in the underweight category and so this is due to a limitation of the model rather than anticipating that the 2013 population percentage value for the underweight category will reduce in this scenario.

Overall, the model results are that the current trend scenario would result in a worse situation for BMI than at present with a mean value of 27.87 and a percentage in the healthy category of just 28.3%.

Table 12.1 Descriptive statistics for the 2013 population and the current trend model.

	2013 population	Current trend model	Change	2013 model
Mean	27.42	27.87	0.46	27.42
Median	26.87	27.36	0.49	26.91
Modal interval mid-point	26.25	26.75	0.50	26.25
Standard deviation	4.84	4.88	0.04	4.75
Skewness	0.89	0.80	-0.09	0.85
Underweight (BMI < 18.5)	1.2%	0.8%	-0.4%	0.8%
Healthy (18.5 ≤ BMI < 25.0)	31.3%	28.3%	-3.0%	31.5%
Overweight (25.0 ≤ BMI < 30.0)	42.4%	42.2%	-0.2%	42.5%
Obese I (30.0 ≤ BMI < 35.0)	18.2%	20.7%	2.4%	18.5%
Obese II (35.0 ≤ BMI < 40.0)	5.2%	6.1%	0.9%	5.1%
Severely obese (BMI ≥ 40.0)	1.7%	2.0%	0.3%	1.6%

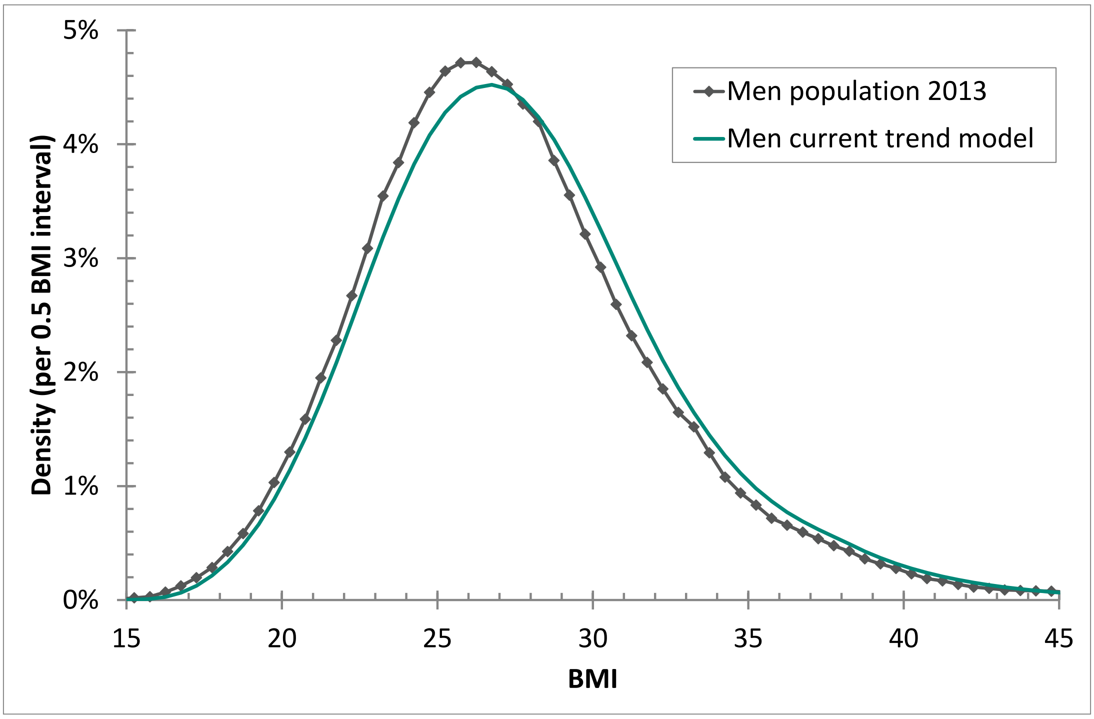


Figure 12.3 BMI distributions for the 2013 population and the current trend scenario for men.

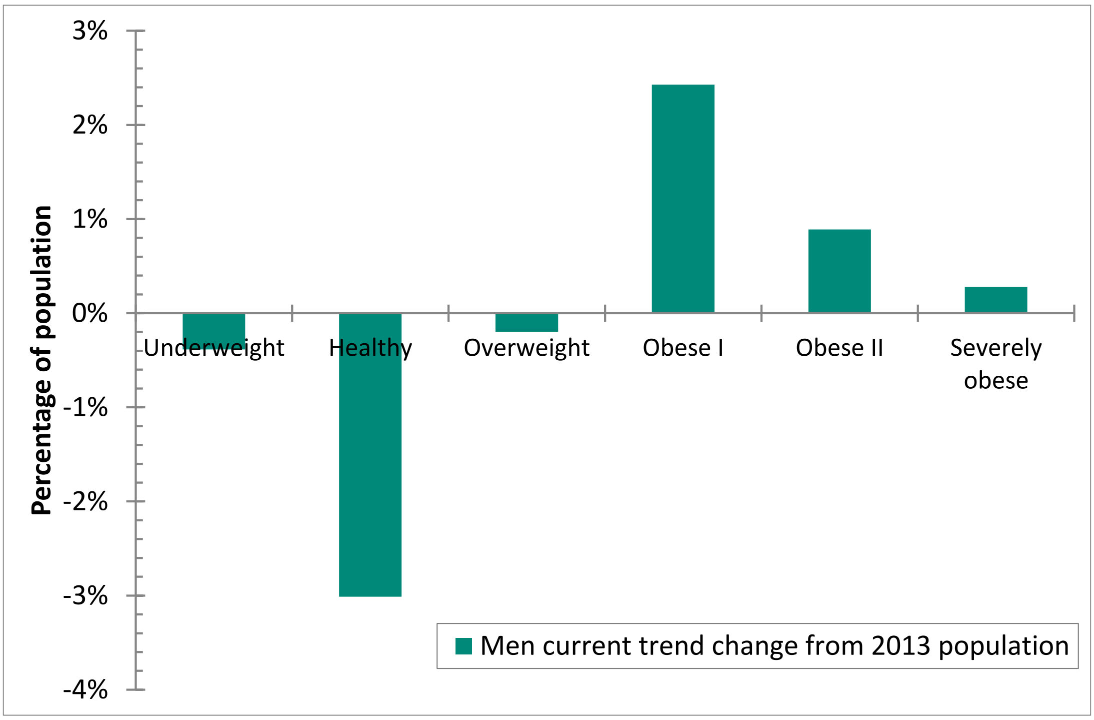


Figure 12.4 Change in BMI category percentages for current trend scenario from 2013 for men.

12.2.3 Sensitivity analysis on the current trend scenario for men

The current trend scenario uses standard deviation and minimum values based on the recent data following the rationale explained in Section 12.2.1. Some experimentation was done to see how sensitive the scenario is to changes in the values for these two parameters. Changing these values has no effect on the mean of the distribution but will affect the spread and the shape.

Altering the standard deviation first, the value was changed from 4.7 to lower and higher values Reducing the standard deviation makes the distribution for each age narrower. The effect on the overall combined distribution is of course also to reduce the standard deviation giving the distribution a taller and narrower peak. It is also less skewed. In terms of the category percentages, the effect is that the percentages increase for the categories of overweight and obese I that are near the peak of the distribution, and those for the other categories that are further from the peak decrease. Increasing the standard deviation naturally has the opposite effect.

Some sensitivity analysis was done to see how large these differences are. The lognormal standard deviation was altered from 4.7 to 4.4, 4.6, 4.8 and 5.0. One way to assess the changes is with the category percentages. Changing the standard deviation has the biggest effect on the overweight category percentage since that is nearest the peak. However, the category of most interest is probably the healthy category. With the chosen model the healthy change from 2013 was -3.0%. The four sensitivity values for the healthy change for standard deviation values of 4.4, 4.6, 4.8, 5.0 are as follows: -4.3%, -3.4%, -2.6%, -1.8%. So, changing the standard deviation by 0.1 to 4.6 or 4.8 has the effect of altering the healthy percentage value by 0.4%. This gives an indication of the margin of uncertainty in the model as 4.6 and 4.8 could be considered reasonable values from the data. The result across this range is therefore that the healthy category percentage reduces by between 2.6% and 3.4%. Values of 4.4 and 5.0 are more extreme values compared to the data and naturally they produce results that have a larger difference from the chosen model.

Changing the minimum value does not alter either the mean or standard deviation of the age distributions but it does alter the shapes. A larger minimum value makes the age distributions and the overall distribution more positively skewed with a higher peak that is further left and a longer right tail. A smaller minimum makes it more symmetrical with a peak further right. With the change in shape the effect on the category percentages is mixed. Increasing the minimum increases the percentages of healthy, overweight, and severely obese, whilst reducing underweight, obese I and obese II. In this scenario, increasing the minimum within a certain range of values brings the distribution closer to the 2013 population (a minimum of 15.89 giving the closest match).

Again, some sensitivity analysis was done. In the chosen model a minimum value of 12.2 was used except for 14.6 for age 18. Values of the minimum for all ages of 10, 11, 13, and 14 were tried, which is considered quite a wide range. Comparing 10 and 14 the difference in the distribution shapes is noticeable but the effects on the category percentages are quite small. To some extent the change in shape has a mixed effect within the categories and so the percentages don't change a great deal. The healthy category change from the 2013 population for the four minimum values is -3.5%, -3.3%, -2.9%, -2.7% and so not that much difference from the -3.0% of the actual model. Hence, the main results are not overly sensitive to the minimum value parameter.

The other parameters here are for the cosine adjustment. The amplitude was changed from 0.3457% to 0.25% and 0.45%. The effect is quite small. A higher value makes the peak a bit higher and narrower. The healthy category changes from 2013 for the two values are -2.8% and -3.3% respectively.

The results would also be affected by changes to the model structure. In particular, altering the choice of distribution of the lognormal. This is hard to test and the model does well in matching the historical distributions and so the model is considered to have a good structure.

The model parameter values set out in Sections 12.2.1 and 12.2.2 are those considered most appropriate for the scenario. Small changes in the values would naturally alter the results but the effect is fairly small and the main pattern of the changes from the 2013 population stays the same.

12.3 Low BMI scenario for men

The idea of this scenario is to use an asymptotic curve to produce a BMI distribution similar to that of 1991-1994. The 3rd curve from Chapter 7 (Section 7.5) would seem to be a good choice as shown in Figure 12.5. The values used are plotted as the green curve which on average looks to be at about the same level as the 1991-1994 data. The 2011-2014 data values are shown for comparison.

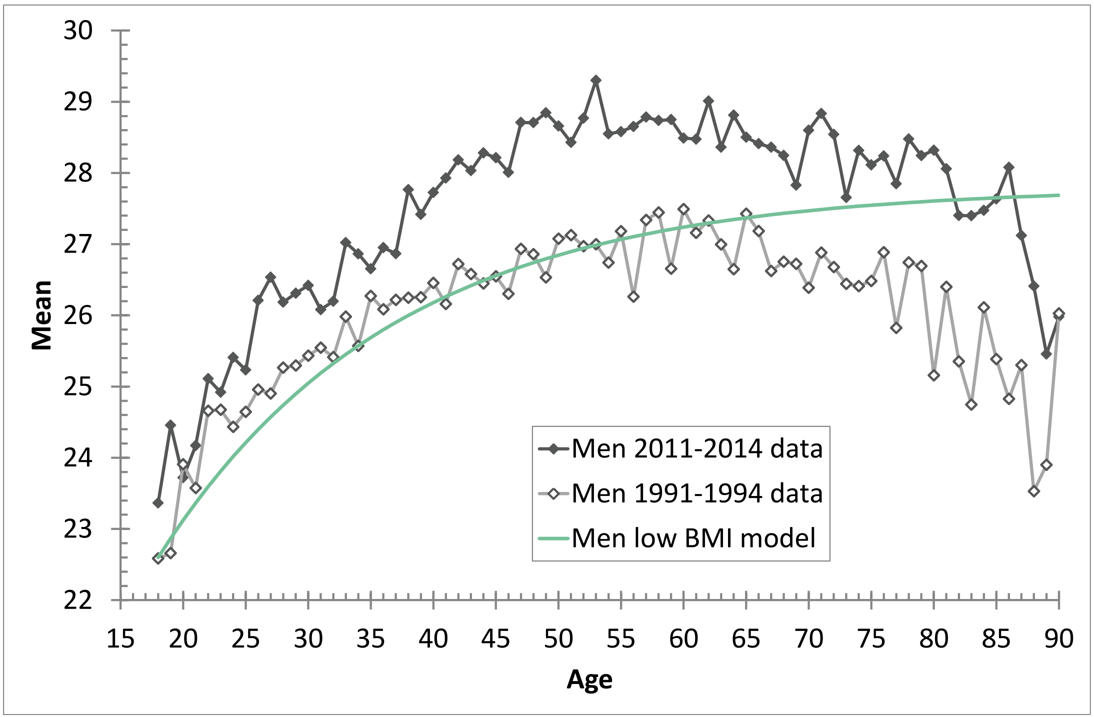


Figure 12.5 Mean BMI values for the low BMI scenario for men.

For the 1991-1994 data the standard deviation values for each age year are mostly between 3.2 and 4.0 with not much trend or change in levels across the whole age range. The average value for all ages is 3.68 and, on this basis, a value of 3.7 was used for this scenario.

The minimum values chosen were based on the values used for modelling the 1991-1994 and 1995-1998 populations (shown in Table 11.1). These values vary quite a bit and the choice is fairly arbitrary. This scenario is based on obtaining low BMI values in modern conditions and so the younger cohorts are considered more relevant. The average minimum value up to age 44 is 10.95 and on this basis a value of 11 was used. The cosine parameters values used were those from modelling 1991-1994 and are amplitude 0.5562% and cosine period 0.5969. The cosine adjustment has the effect of reducing the standard deviation values for each age to about 3.62.

A chart of the resulting BMI distribution for this scenario is shown in Figure 12.6. The 2013 and 1993 distributions are also shown. The aim of the scenario was just to model reasonably similar values to 1993 rather than getting a strong correspondence. As it turns out, the model does match the 1993 distribution very closely. Consequently, the statistics are very similar and, for example, the model mean is 26.13 compared to 26.05 for 1993. Examples of the individual component age distributions along with the overall distribution are shown in Figure 12.7 (note that the y-axis scales are different in Figures 12.6 and 12.7and therefore the overall distribution appears to have a slightly different shape in the two charts although this is not actually the case).

The category percentages and the changes from the 2013 distribution are shown in Figures 12.8 and 12.9. These are inevitably similar to the differences between the 2013 and 1993 distributions (the statistics are in Chapter 4 in Table 4.1). For example, the model healthy category percentage is 8.5% higher than 2013 at 39.8% (the 1993 value is 40.4%). The severely obese category has only 0.3% (the same as 1993) compared to 1.7% in 2013.

What this scenario implies is that the BMI aging pattern shown in Figure 12.5, combined with the other parameter values used particularly the lower standard deviation value, would get the overall population BMI distribution back to early 1990s levels. This scenario could therefore be used as target levels to aim for across the age range. For example, the mean values for ages 20, 25, and 30 in the model are 23.12, 24.21, and 25.05 respectively, each with a standard deviation of 3.62. The younger ages do have a significant proportion in the underweight category (age 20 has 6.5%), and this aspect is presumably not desirable. However, the distributions in Figure 12.7 show generally what is required to get back to the BMI population distribution of the early 1990s.

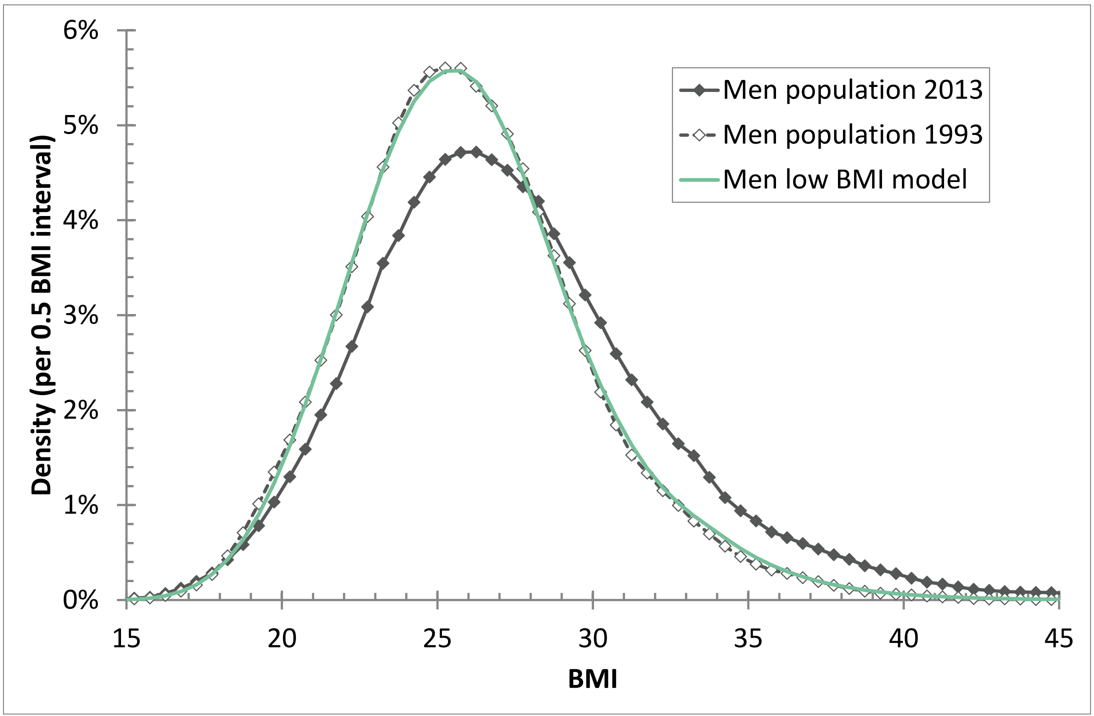


Figure 12.6 BMI distribution for the low BMI scenario for men.

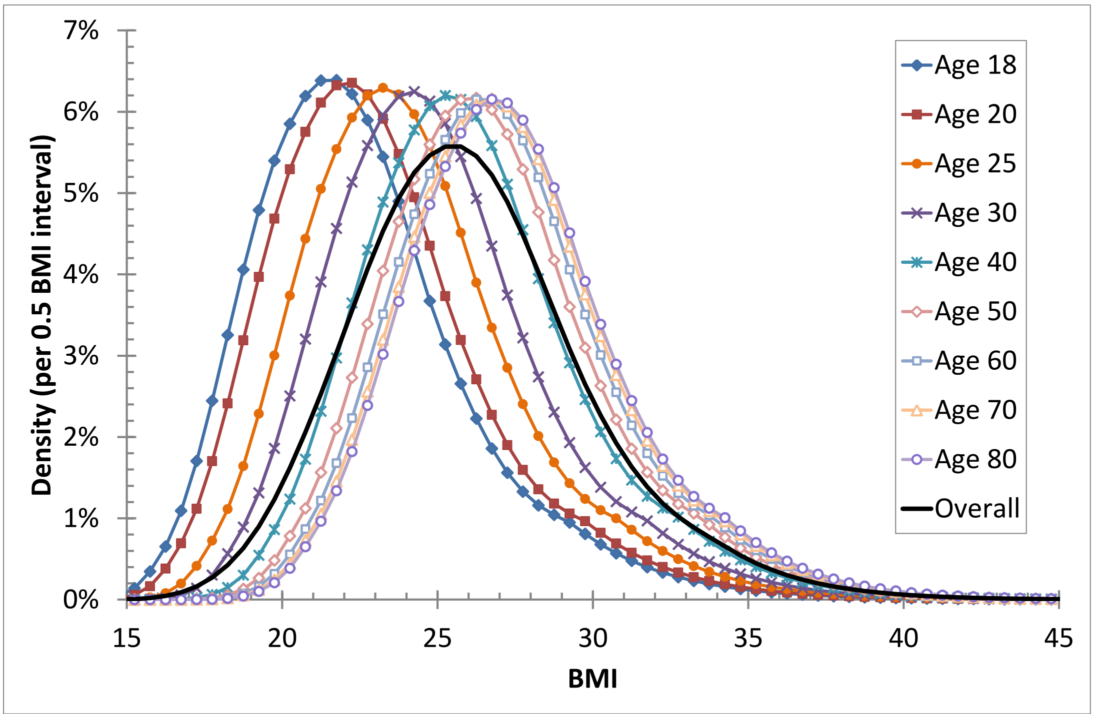


Figure 12.7 Example age BMI distributions for the low BMI scenario for men.

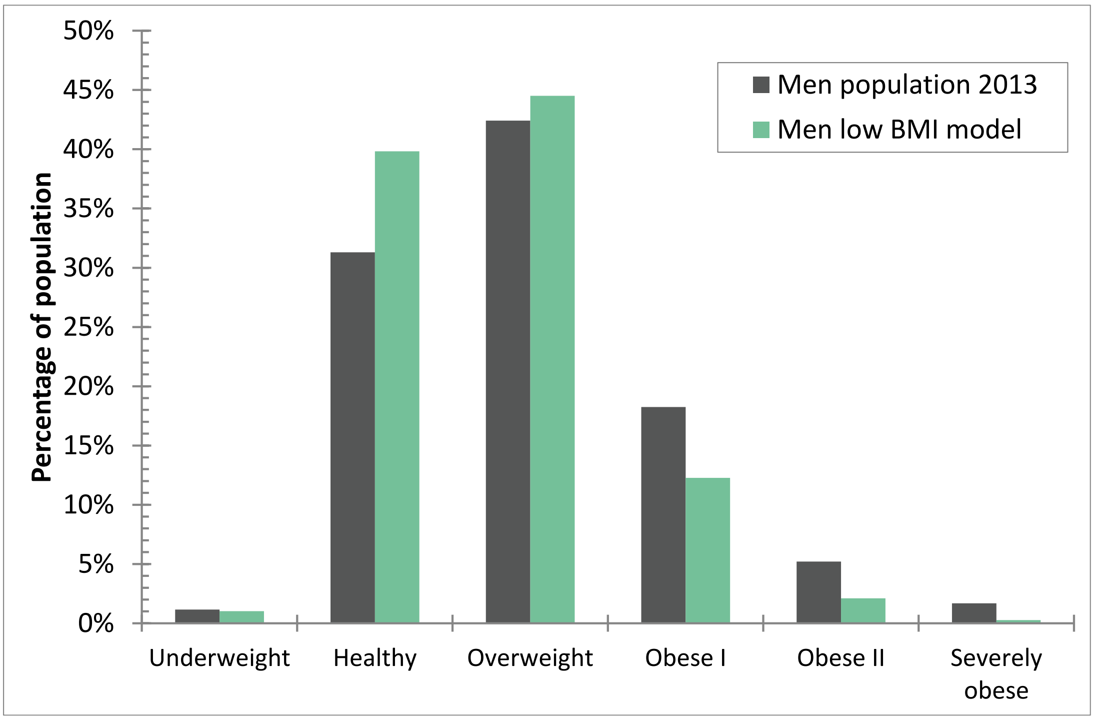


Figure 12.8 BMI category percentages for the 2013 population and the low BMI scenario for men.

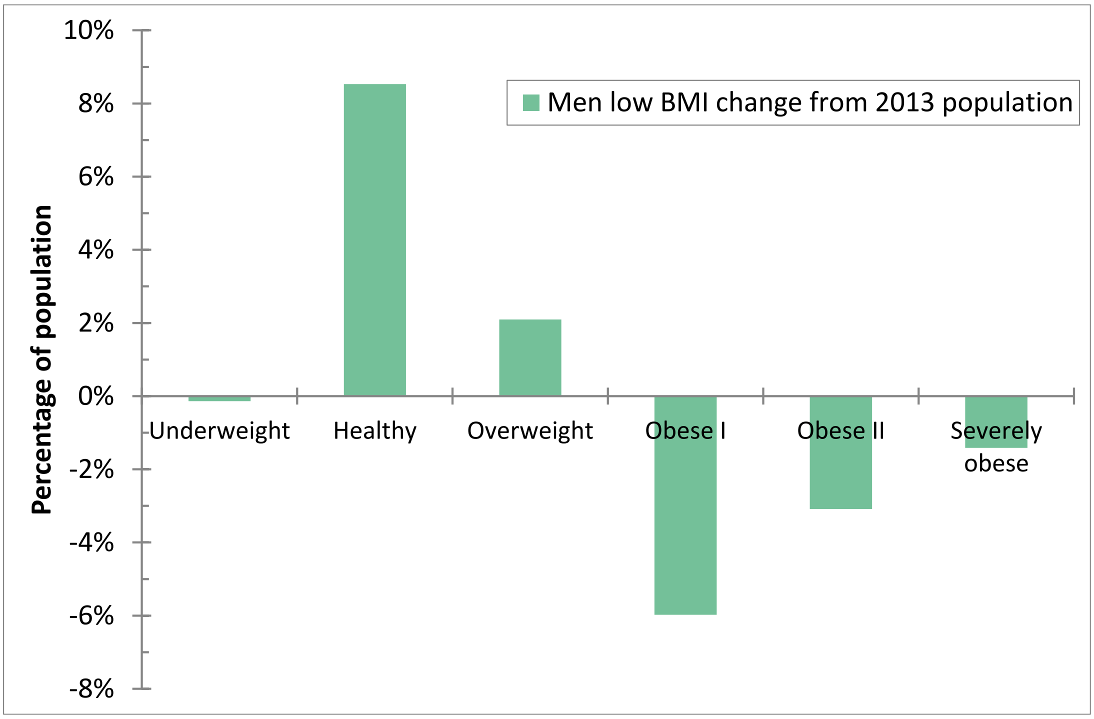


Figure 12.9 Change in BMI category percentages for low BMI scenario from 2013 for men.

12.4 Current trend scenario for women

12.4.1 Parameter values for the current trend scenario for women

This analysis for the current trend scenario for women follows the same approach as set out in Section 12.2, in using the mean values from the aging model curve in Chapter 7 (Figure 7.6). This curve is shown in Figure 12.10 along with the 2011-2014 data values used for the 2013 population distribution in Chapter 4. The values for the model are mostly just a little above the data values. The suggested interpretation of the cohort data in Chapter 7 is that most of the cohorts are quite close to following the trend which is reflective of modern lifestyles. Consequently, the current trend model can be expected to have just a small increase in BMI compared to the 2013 population distribution.

The standard deviation value that was used is based on the 2011-2014 data. The actual data values are at a fairly constant level up to about age 65, and are shown in Figure 12.11. The average up to that age is 5.73, and on that basis a value of 5.7 was used for this scenario for all ages.

The minimum value for each age was set from the values used for modelling the 2009 and 2013 distributions in Table 11.3, which are based on the 2007-2014 age group distributions. The values are quite similar up to age 54 being between 14.11 and 15.17. The average up to age 54 is 14.60 and so a value of 14.6 was used in the scenario for all ages. Unlike the model for men, there is no cosine adjustment in the model for women.

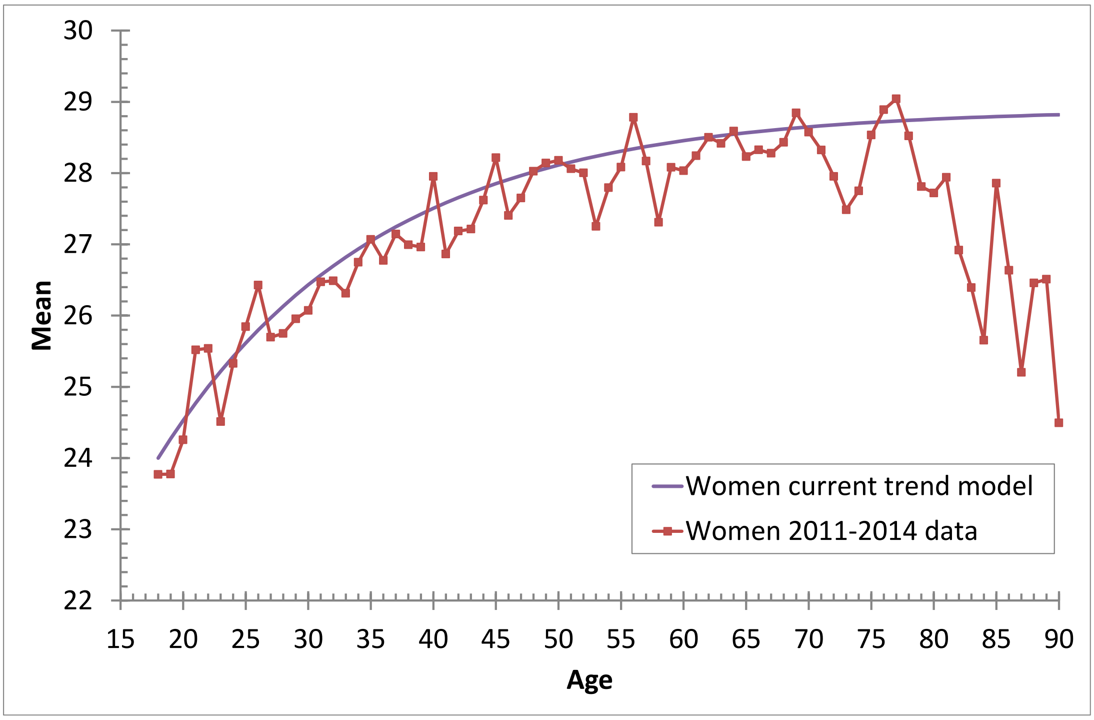


Figure 12.10 Mean values for the current trend model and 2011-2014 data for women.

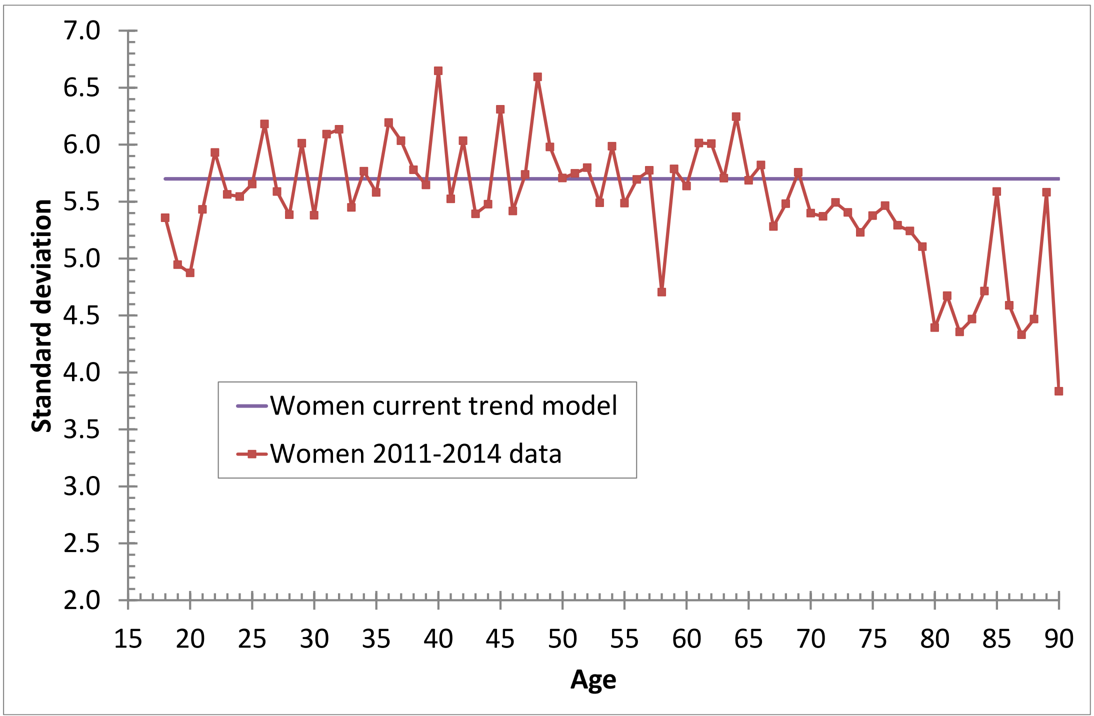


Figure 12.11 Standard deviation values for current trend model and 2011-2014 data for women.

12.4.2 Results for the current trend scenario for women

The population distribution for the current trend model was generated by combining the lognormal distributions for each age (using the parameter values explained in Section 12.4.1) weighted by the population age profile percentages explained in Section 3.3. The resulting distribution is shown in Figure 12.12 and the statistics are in Table 12.2. The changes in the category percentages for the current trend scenario compared to the 2013 distribution are shown in Figure 12.13.

Table 12.2 Descriptive statistics for the 2013 population and the current trend model.

	2013 population	Current trend model	Change	2013 model
Mean	27.13	27.50	0.37	27.13
Median	26.07	26.46	0.39	26.13
Modal interval mid-point	24.25	24.75	0.50	24.25
Standard deviation	5.77	5.84	0.07	5.76
Skewness	1.06	1.29	0.23	1.27
Underweight (BMI < 18.5)	1.9%	1.1%	-0.8%	1.5%
Healthy (18.5 ≤ BMI < 25.0)	39.6%	37.1%	-2.5%	39.6%
Overweight (25.0 ≤ BMI < 30.0)	32.7%	34.6%	1.9%	33.6%
Obese I (30.0 ≤ BMI < 35.0)	16.0%	17.0%	1.0%	16.3%
Obese II (35.0 ≤ BMI < 40.0)	6.4%	6.5%	0.2%	6.0%
Severely obese (BMI ≥ 40.0)	3.4%	3.6%	0.2%	3.1%

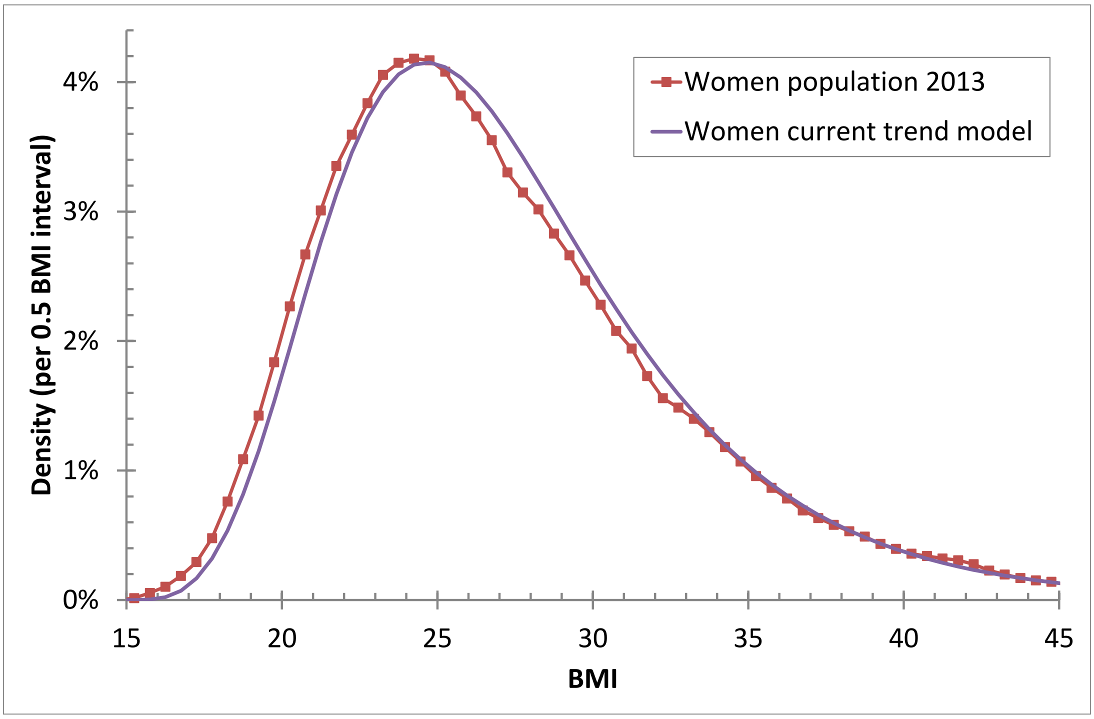


Figure 12.12 BMI distributions for the 2013 population and the current trend scenario for women.

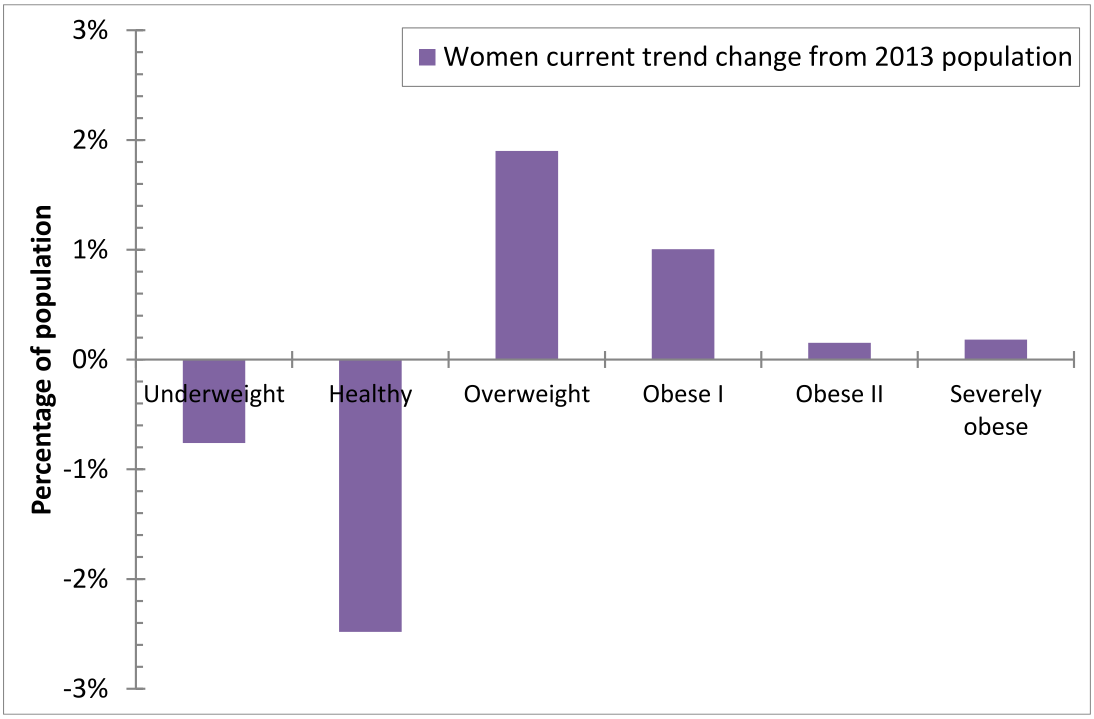


Figure 12.13 Change in category percentages for current trend scenario from 2013 for women.

From Figure 12.12, the current trend model has higher BMI values than the 2013 distribution by being shifted slightly to the right in the peak and most of the two tails. The end of the right tail is very similar to the 2013 distribution. The statistics in Table 12.2 show that the mean BMI is 0.37 higher than for the 2013 population distribution and the healthy category percentage is lower by 2.5%. The main increases in the category percentages for the model compared to 2013 in Table 12.2 and Figure 12.13 are overweight being 1.9% higher and obese I being 1.0% higher.

For reference, Table 12.2 also shows the values for the 2013 lognormal model. As for the model for men (Section 12.2.2), the underweight percentage for the 2013 model is less than for the actual 2013 distribution. This is a slight limitation of the model and is part of the reason that the current trend model has a lower underweight category percentage than the 2013 distribution.

The magnitude of the changes of the model from 2013 is similar to that for the current trend scenario for men. It is also comparable with the largest four year changes for the distributions for women in Chapter 4 from 1993 to 1997 and 1997 to 2013, which have increases in mean BMI of 0.41 and 0.48 respectively. However, this is a long term scenario and so the increases are quite small when viewed as long term changes.

Overall, the model results are that the current trend scenario would result in a slightly worse situation for BMI than at present with a mean value of 27.50 and a percentage in the healthy category of 37.1%.

12.4.3 Sensitivity analysis on the current trend scenario for women

Sensitivity analysis was applied to this scenario in the same way as for the scenario for men (Section 12.2.3). The standard deviation and the minimum values were altered since there is some subjectivity in the choice of these. Changing these values does not alter the mean but it does change the shape and spread of the distribution.

The lognormal standard deviation was altered from 5.7 to values of 5.4, 5.6, 5.8, 6.0. The effect is very similar to that for the model for men, with a smaller standard deviation narrowing the peak. This makes the overall standard deviation and the skewness smaller. It also increases the percentages in the categories near the peak and reduces those away from the peak. The effect on the category percentages of a smaller standard deviation is therefore to make the overweight and obese I percentages larger, and the healthy, obese II and severely obese percentages smaller. Since a smaller standard deviation makes the healthy percentage smaller, the reduction compared to the 2013 distribution is larger. The change in the healthy percentage from 2013 for the four standard deviation values is -3.5%, -2.8%, -2.2%, -1.6%, compared to -2.5% for the chosen standard deviation value of 5.7.

The minimum value was changed from 14.6 to 13, 14, 15, 16. Changing the minimum doesn't change the mean and has little effect on the standard deviation. However, the shape does change. Reducing the minimum from 14.6 to 13 moves the left tail a bit to the left and lowers the peak. The end of the right tail stays much the same. The main effects of this are to increase the underweight and obese I category percentages, whilst reducing the healthy category percentage. Increasing the minimum has the opposite effect with the left tail moving to the right and the peak of the curve getting higher. The change in the healthy percentage from 2013 for the four minimum values is -3.5%, -2.9%, -2.2%, -1.6%, compared to -2.5% for the chosen value. So, the effect on the healthy category percentage is very similar to the sensitivity changes in the standard deviation.

Overall, these experiments show that the results only change to a small extent with the different values experimented with for the standard deviation and the minimum.

12.5 Low BMI scenario for women

As discussed in section 12.3, the general aim of this scenario is to use an asymptotic curve that gives a BMI distribution similar to the actual distribution for 1993. As for men, the 3rd curve from the top from Chapter 7 (Table 7.1 and Figure 7.6 in section 7.5) would seem to be a good choice. The values used are shown as below in Figure 12.14 which on average looks to be at about the same level as the 1991-1994 data. The 2011-2014 data values are shown for comparison.

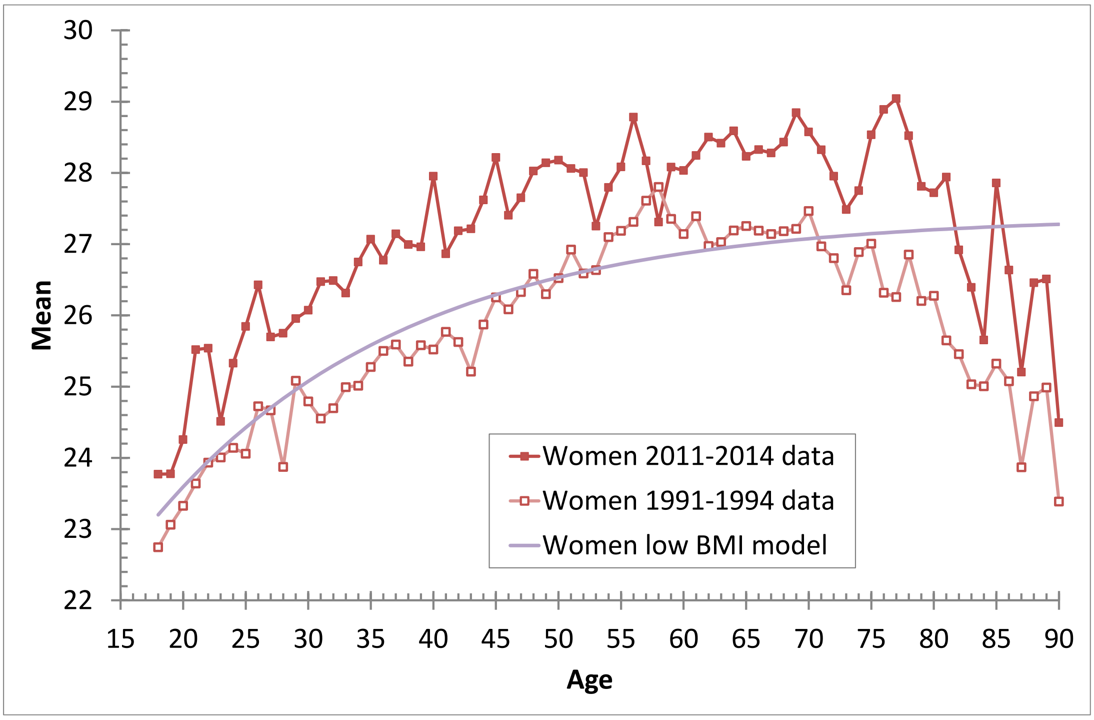


Figure 12.14 Mean BMI values for the low BMI scenario for women.

The parameter values for the standard deviation and minimum values are, to some extent, arbitrary choices for this scenario. The values used were based on the 1991-1994 data, particularly the values for the younger ages. For the standard deviation, the data values are at a similar level for most ages without any apparent trend. The values for 18 to 20 and above 79 are slightly lower than for the ages in between. The average for ages 21 to 65 is 4.88 and, on this basis, a value of 4.9 was chosen for the scenario.

The minimum values used for modelling 1991-1994 and 1995-1998 are shown in Table 11.3 in Chapter 11. The values are quite similar up to age 54. The average of these values is 15.51 and, on this basis, a value of 15.5 was chosen for the scenario.

The BMI distribution was generated using these parameter values and is shown in Figure 12.15. The scenario was just aiming for a distribution that was fairly similar to 1993 but the result does turn out to be very close to the 1993 distribution. The end of the left tail is a bit to the right of the 1993 distribution and the overall mean is slightly higher at 26.04 compared to 25.76. Examples of the individual component age distributions and the overall distribution are shown in Figure 12.16 (with a different y-axis scale to Figure 12.15 and so the peak of the overall distribution looks lower on Figure 12.16).

The category percentages and the changes from the 2013 distribution are shown in Figures 12.17 and 12.18. Since the model distribution is close to the 1993 distribution, these are similar to the differences between the 2013 and 1993 distributions (statistics in Chapter 4 in Table 4.2). For example, the model healthy category percentage is 8.3% higher than 2013 at 47.9% (the 1993 value is 49.0%). As shown in Figures 12.17 and 12.18, the scenario has considerably lower percentages in the underweight, obese I, obese II, and severely obese categories. The changes are -0.7%, -3.7%, -2.5%, -1.7%.

Overall, this scenario shows one way of getting the overall BMI distribution back to that of the early 1990s with a mean BMI pattern of the shape of the aging model curves in Chapter 7. As also noted in Section 12.3, this scenario could therefore be used as target levels to aim for across the age range. For example, the mean values for ages 20, 25, and 30 in the model are 23.59, 24.43, and 25.08 respectively, each with a standard deviation of 4.90. As mentioned in Section 12.3, the younger ages do have a significant proportion in the underweight category in the model (age 20 has 6.5%), and so this aspect is presumably not desirable. However, the distributions in Figure 12.16 show generally what is required to get back to the better BMI population distribution of the early 1990s.

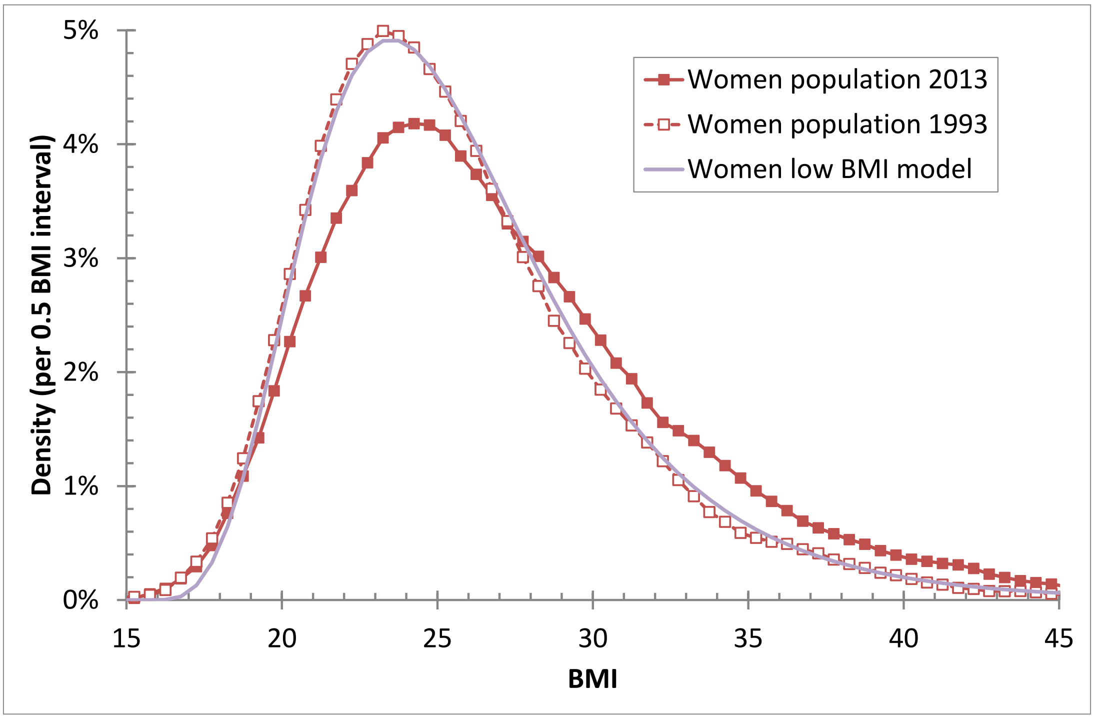


Figure 12.15 BMI distribution for the low BMI scenario for women.

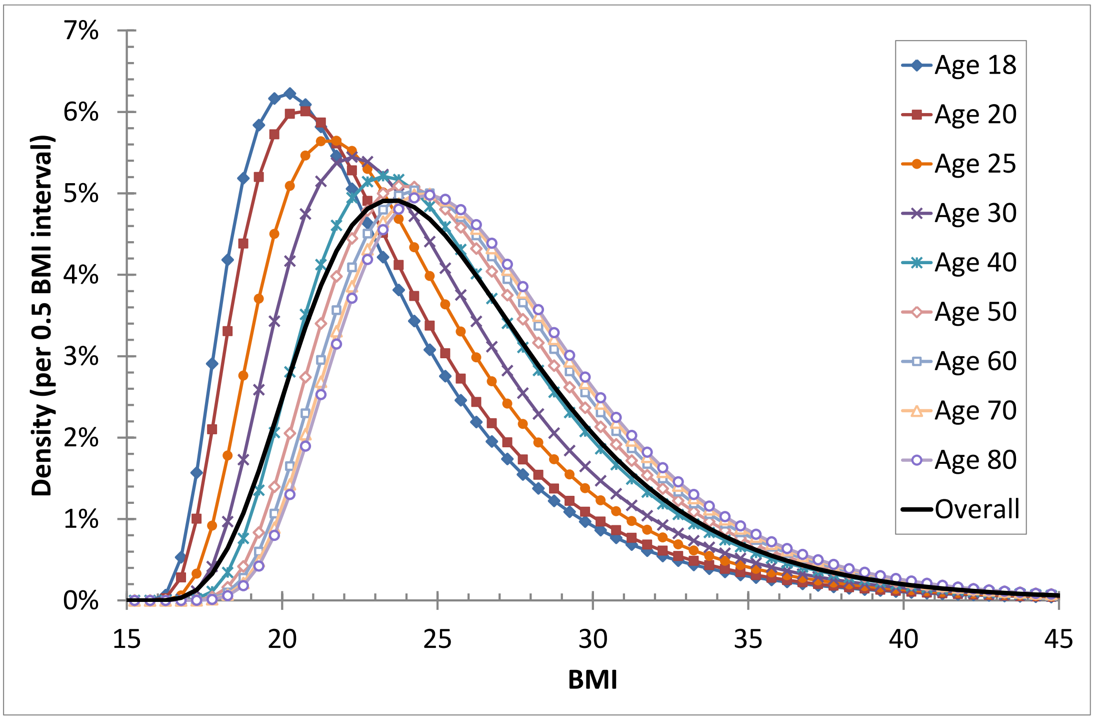


Figure 12.16 Example age BMI distributions for the low BMI scenario for women.

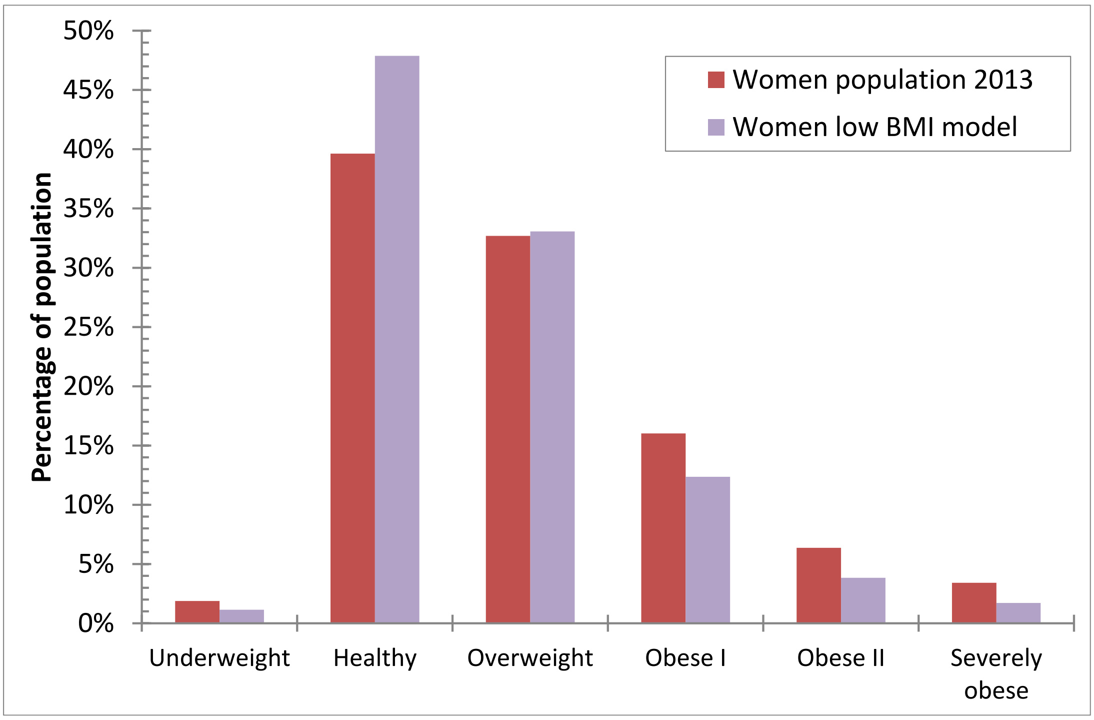


Figure 12.17 BMI category percentages for the 2013 population, low BMI scenario for women.

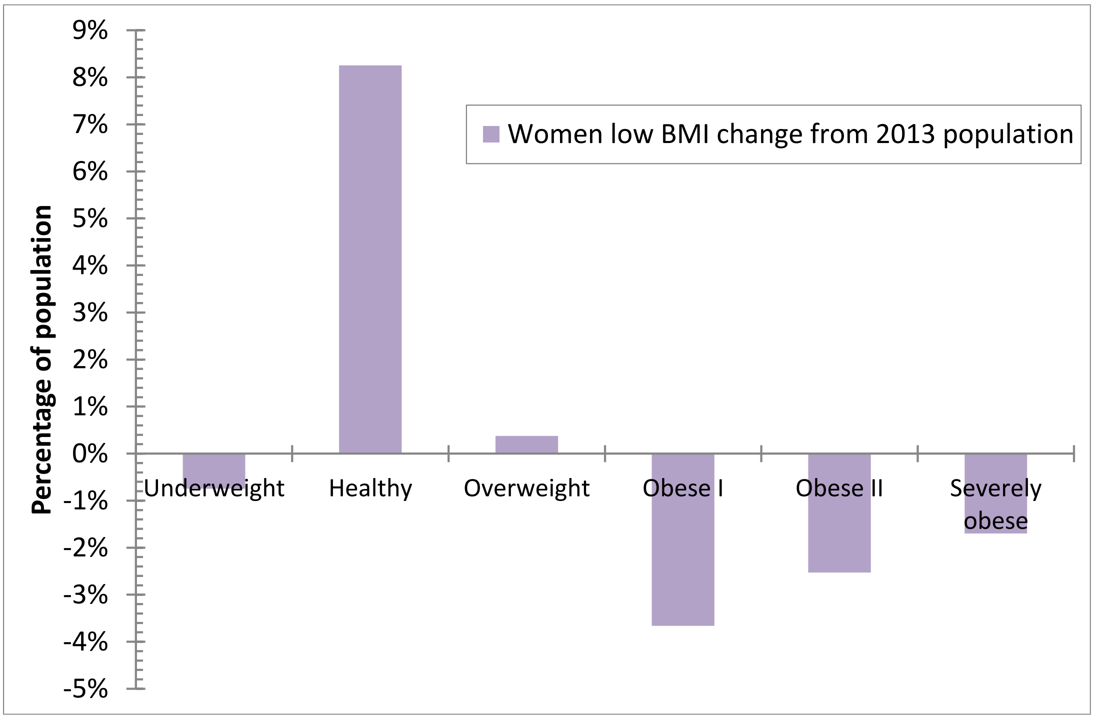


Figure 12.18 Change in BMI category percentages for low BMI scenario from 2013 for women.

Chapter 13. Converting BMI changes into calories

Key points from this chapter:

- Energy balance equations can be applied to convert BMI values and BMI changes into calories. At equilibrium, energy intake equals energy expenditure. Equations can be used to calculate energy expenditure, which therefore gives energy intake at equilibrium.
- The approach here follows that of a 2012 report by The Scientific Advisory Committee for Nutrition (SACN). Energy expenditure is calculated as basal metabolic rate (BMR) multiplied by a physical activity level multiplier (PAL). The Henry (2005) equations are used for BMR. Applying these equations gives values for the calorie difference at equilibrium for a 1 unit difference in BMI, depending only on age category, height, and activity level (Table 13.4). The equations can be applied to any of the BMI values or the BMI changes in this report. The calorie amounts are calculated for two examples taken from earlier chapters in this report.
- In the first example, the calories are calculated for the increase in BMI from 1993 to 2013 for the combined scaling models for men and for women from Section 5.8.1. The general pattern is exactly the same as that of the BMI increases, since they are just multiplied by a constant to convert to calories. Following the BMI pattern, the increase in calories is higher for higher starting BMI values. For common 1993 BMI values between 25 and 30, the extra calories per person calculated at equilibrium for the median activity level vary from about 50 to 120 kcal / day for men, and 40 to 80 kcal / day for women. These are fairly small amounts of calories and yet the change in the BMI distributions is considerable.
- In the second example, the increase in BMI from age 18 to 24 from Sections 8.5.2 and 9.5.2 is converted into calories. This is the age range with the greatest increase in BMI with age. The percentile BMI values are used. The calorie amounts calculated are again reasonably small. Most of the values for men for the median activity level are about 170 kcal / day. The values vary for women up to a maximum of about 140 kcal / day.
- These are population level values and so they are estimates of average values. They should therefore not be applied for individuals. The accuracy of the results here depends on the validity of the values and models used. Alternative models are compared from other sources, with some differences in the results.

13.1 Energy balance equations and equations for energy expenditure

13.1.1 Energy balance

Energy balance is based on the basic physics principle of conservation of energy (the first law of thermodynamics). The implication for the human body is that if there is a difference between energy intake and energy expenditure then this excess or deficiency will be used in adding or removing body tissue which stores or releases the difference in energy. Hence, a difference between energy intake and expenditure will result in a change of weight. At equilibrium, where there is no change in weight, energy intake will equal energy expenditure. A detailed review of the literature on this is outside the scope of this report but, for example, the energy balance approach is discussed in Marlett and Ravussin (2018).

Most of the work in this report is looking at different populations, particularly at different points in time, rather than changes in individuals over time. Therefore, in this context the interest is in comparing the equilibrium calorie intake values for the different populations.

Energy intake is the calories consumed but this can be very difficult to measure. However, there are methods for calculating energy expenditure and so at equilibrium this also gives the value for energy intake. Calculating energy expenditure is therefore the approach used in the analysis method described here.

It is important to take into account weight when estimating energy expenditure, because energy expenditure increases with weight. This is because, if weight increases, extra energy is required both for maintaining the extra tissue and for doing activities as these involve moving the extra weight.

This also means that changes of intake, expenditure, and weight for an individual over time are a dynamic process with changes in weight affecting expenditure as well as differences between intake and expenditure affecting weight. The potential interactions if someone in equilibrium increases their energy intake (i.e., eats more calories) are:

1. Energy intake now exceeds energy expenditure.
2. This results in the formation of extra tissue and therefore extra weight.
3. The extra weight means that energy expenditure increases.
4. Steps 2 and 3 continue until energy expenditure matches energy intake and a new equilibrium is reached.

In the same way, energy intake being less than energy expenditure will result in weight loss until a new equilibrium is achieved. There are also other complexities between the different types of body tissue. Changes for individuals particularly over a short timescale are not the focus of this chapter or this report, although there are some comments in Section 13.6.2 on a simulator from the literature that implements a model of this.

13.1.2 SACN approach

The approach in this chapter follows that used by The Scientific Advisory Committee for Nutrition (SACN) in their report published in 2012 that looks at energy requirements (report titled “Dietary Reference Values for Energy” SACN, 2012 [36]). This is also referenced earlier in this report in Section 6.3.1. There are also cross references in several other chapters to the SACN mean height data in this chapter.

SACN is an advisory group to the U.K. Government [37]. The 2020 annual report for SACN explains their role as follows (page 3 of the annual report [38]): “The role of SACN is to provide scientific advice on, and risk assessment of, nutrition and related health issues. It advises the four UK health departments and other government departments and agencies. Members are appointed as independent scientific experts on the basis of their specific skills and knowledge. The committee also includes 2 lay members.”

The SACN (2012) report was done in the context of increasing rates of obesity. The aim was to derive dietary reference values (DRVs), and for adults these are based on a BMI of 22.5. This is applying a value from within the healthy BMI range, and the report preface on page 1 says “Using this approach, if overweight groups consume the amount of energy recommended for healthy weight groups, they are likely to lose weight, whereas underweight sections of the population should gain weight towards the healthy body weight range”. However, page 2 of the preface explains that the reference values produced are at a population level and should not be used for individuals: “It is important to note that DRVs should be used to assess the energy requirements for large groups of people and populations, but should not be applied to individuals due to the large variation in physical activity and energy expenditure observed between people”.

A key part of the work in the SACN report is the use of equations to calculate energy requirements, which gives the energy intake required at equilibrium and hence the dietary reference values. The reference values produced are called estimated average requirements (EAR).

The SACN report is a large document and includes a detailed discussion of the processes involved in energy intake, expenditure, and the effects of imbalances. It covers both children and adults, with consideration given for different situations such as pregnancy and lactation.

[36] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf
[37] https://www.gov.uk/government/groups/scientific-advisory-committee-on-nutrition
[38] https://www.gov.uk/government/publications/sacn-annual-report-2020

The overall energy used is called the total energy expenditure (TEE). The SACN report calculates this as the basal metabolic rate (BMR) multiplied by a physical activity level multiplier (PAL). BMR is the energy requirements for a person at rest and so represents the energy needed for the basic functions of the body. The SACN report uses the equations for BMR derived by Professor Jeya Henry, as set out in Henry (2005). These are labelled in the SACN report and in this chapter as the Henry equations, although Henry (2005) calls them the "Oxford equations" since Professor Henry was working at Oxford Brookes University. SACN considered the Henry equations to be the most suitable equations for BMR, as explained in paragraphs 80 and 81 on page 42 of their report.

Henry (2005) gives some of the history of measuring BMR, which dates back over 100 years, and describes some of the methods used. The paper also discusses the conditions in which BMR should be measured. The Henry paper evaluates the different data sets of BMR and selects those considered most reliable. This results in 166 studies being used giving values for 5794 men and 4702 women. The Henry equations were derived from these cases using linear regression with variables for height and weight. Separate equations were produced for men and for women, and for different age groups. The equations with both height and weight are used here since these are the versions used by SACN (2012), although Henry (2005) finds that the difference is not significant compared to the equations with just weight. The details and parameter values are given in the following sections. Since the equations are based on regression they represent average values. The charts in Henry (2005) show that there is quite a bit of variation around the regression line between the different cases. Therefore, values for BMR may vary considerably on an individual level.

The PAL activity multiplier values were derived by SACN from a dataset put together from two U.S.A. studies (called OPEN and Beltsville), as described in pages 57-60 of the SACN (2012) report. The PAL values were obtained from measured values of TEE in the studies (using the doubly labelled water method), comparing these with measured BMR in the Beltsville data and with the Henry value for BMR for the OPEN data. The PAL values were then calculated by PAL = TEE / BMR. Some trimming was applied to the data and then the quartiles and the median used for the PAL values. The resulting PAL values are given in Section 13.3.

13.2 Approach based on SACN (2012), using the Henry (2005) equations

As explained in the previous section, the SACN (2012) report uses the Henry equations (Henry, 2005) for basal metabolic rate (BMR) along with a physical activity level multiplier (PAL).

The SACN approach is to calculate total energy expenditure (TEE) as:

$$\text{TEE} = \text{PAL} \times \text{BMR} \qquad (13.1)$$

The Henry equation for BMR in kcal / day using weight in kg, w, and height in metres, h, for adults has the form (using coefficients α, β, γ)

$$\text{BMR} = \alpha + \beta h + \gamma w \qquad (13.2)$$

Since BMI = w / h^2 we have w = BMI × h^2 and so:

$$\text{BMR} = \alpha + \beta h + \gamma h^2\text{BMI} \qquad (13.3)$$

$$\text{TEE} = \text{PAL} \times (\alpha + \beta h + \gamma h^2\text{BMI}) \qquad (13.4)$$

From equation (13.4):

$$\text{Difference in TEE for each unit difference in BMI} = \text{PAL} \times \gamma h^2 \qquad (13.5)$$

Under equilibrium the energy intake will equal energy expenditure. Therefore, the estimated difference in calorie intake is the same as Equation 13.5:

$$\text{Difference in calorie intake for each unit difference in BMI} = \text{PAL} \times \gamma h^2 \qquad (13.6)$$

This means that, for a given height, PAL value, and coefficient γ, the difference in equilibrium calorie intake per unit difference in BMI is a constant.

13.3 SACN (2012) and Henry (2005) parameter values for different ages

In the SACN (2012) report the Henry parameters are given in Table 18 on page 105, the mean heights are in Tables 11 and 12 on page 61 (derived from the 2009 Health Survey for England), the PAL values are in Tables 10, 11, and 12, and Figure 3 on pages 59-61. Several age categories are used in the report, although the PAL values are the same for all adults. The Henry values in the SACN report are those from Table 15 on page 1146 in Henry (2005). The values from the SACN report are reproduced below in Tables 13.1, 13.2 and 13.3.

Table 13.1 Henry (2005) parameter values given in the SACN (2012) report.
[SACN, 2012, Table 18 on page 105]

	Constant (α)	Height coefficient (β)	Weight coefficient (γ)
Men age 18-30	113	313	14.4
Men age 30-60	-137	541	11.4
Men age > 60	-256	541	11.4
Women age 18-30	-282	615	10.4
Women age 30-60	-11.6	502	8.18
Women age > 60	10.7	421	8.52

Table 13.2 Mean height values given in the SACN (2012) report.
[SACN, 2012, Tables 11 and 12 on page 61]

Age	Men mean height in metres	Women mean height in metres
19-24	1.78	1.63
25-34	1.78	1.63
35-44	1.76	1.63
45-54	1.75	1.62
55-64	1.74	1.61
65-74	1.73	1.59
75+	1.70	1.55
All adults	1.75	1.62

Table 13.3 PAL values given in the SACN (2012) report.
[SACN, 2012, Tables 10, 11, and 12 on pages 59-61]

	Value for all adults
PAL 25th centile	1.49
PAL median	1.63
PAL 75th centile	1.78

As shown in Table 13.3, three PAL percentile values are used in the SACN report corresponding to different levels of activity. These are obtained as the quartile values (median, and 25th and 75th percentiles) of the PAL values derived from a dataset put together from two studies, as described in pages 57-60 of the report. A lower PAL value means a lower level of activity. They also found that there was no significant association between PAL and BMI in the data (page 59, paragraph 117) and so these PAL values should be suitable for a wide range of BMI values. There is also only a small association effect with age (page 59, paragraph 118). Hence the use of the same PAL values for all ages. The report does note that PAL values may be very low for some older adults (page 62, paragraph 124).

13.4 Calorie calculations from the Henry (2005) equation

Calorie values can be calculated from Equation 13.6 using the parameter values from Section 13.3. The values needed for Equation 13.6 are the PAL value, the Henry (2005) weight coefficient γ, and the height. Height does not vary much with age and the same PAL values are used for all ages. Therefore, the main difference with age is the Henry equation weight coefficient parameter value.

Taking the example of age 30-60 for men, the Henry weight coefficient is 11.4. Using the average height for age 45-54 of 1.75m, and the median PAL of 1.63, Equation 13.6 gives: difference in calorie intake for each unit difference in BMI = PAL × γh^2 = 1.63 × 11.4 × 1.75^2 = 56.91 kcal / day. In other words, the implication is that at equilibrium a difference in BMI of 1 corresponds to a difference in the number of calories consumed of 57 kcal / day. For the 25th and 75th percentiles for PAL, the values per unit difference in BMI are 52.02 and 62.14 kcal / day, respectively.

For women, the Henry weight coefficient for age 30-60 is 8.18. The average height for ages 45-54 is 1.62m, which for the median PAL gives a difference in calories per unit difference in BMI of PAL × γh^2 = 1.63 × 8.18 × 1.62^2 = 34.99 kcal / day. For the 25th and 75th percentiles for PAL the values per unit difference in BMI are 31.99 and 38.21 kcal / day, respectively. Therefore, the calculated calorie difference per unit BMI is quite a lot lower for women than for men.

These values and those for the other Henry age groups are given in Table 13.4. The height values are selected as considered most suitable from Table 13.2. Results are given for each of the three PAL values from Table 13.3. The Henry weight coefficient value for men and for women is higher for age 18-30 than for older age groups and so the calorie values are higher.

Table 13.4 Calculated change in equilibrium daily calorie intake per unit difference in BMI.

	Weight coefficient (γ)	Height in metres	kcal / day (lower PAL = 1.49)	kcal / day (median PAL = 1.63)	kcal / day (upper PAL = 1.78)
Men age 18-30	14.4	1.78	67.98	74.37	81.21
Men age 30-60	11.4	1.75	52.02	56.91	62.14
Men age > 60	11.4	1.73	50.84	55.61	60.73
Women age 18-30	10.4	1.63	41.17	45.04	49.18
Women age 30-60	8.18	1.62	31.99	34.99	38.21
Women age > 60	8.52	1.59	32.09	35.11	38.34

The actual TEE value can be calculated for any BMI value using Equation 13.4 with the relevant Henry coefficients. This is the approach in the SACN report and their EAR reference values using a BMI of 22.5 for each PAL value are given in Tables 11 and 12 on page 61 of the report. The values are given in MJ / day. Table 16 on page 85 also gives the values in kcal / day for median PAL although from the numbers these are calculated by converting the MJ value given to 1 decimal place into kcal by multiplying by 239. This is slightly different to using the Henry kcal parameters because of the rounding of the MJ value. The values in Table 16 of the SACN report reduce with age. For men they vary from 2772 kcal / day for age 19-24 to 2294 kcal / day for age 75+. For women they vary from 2175 kcal / day for age 19-24 to 1840 kcal / day for age 75+. Again, as explained in Section 13.1.2, these are population level values and are not values for individuals. People may vary considerably in their activity levels and BMR values.

13.5 Calorie values for BMI results

Any of the BMI values and the BMI changes already given in this report can be converted into calories using the approach described in Sections 13.1-13.4. This section gives two examples with substantial changes in BMI. The first example in Section 13.5.1 is the change from the 1993 to 2013 population BMI distributions. The second example in Section 13.5.2 is the increase in BMI from age 18 to 24.

13.5.1 Calories for the changes from 1993 to 2013

In Section 5.8 of Chapter 5, various models are applied for the changes in BMI from 1993 to 2013. All have a similar pattern for the increase in BMI, and the actual percentile value increases are also very similar to the models (Figure 5.15 in Chapter 5, Appendix 5.V). Any of the models or the percentiles could be chosen and calories calculated, and the results would be much the same. The model that is used here is the combined model from Section 5.8.1 that uses the scaling models for each of the four year time intervals and combines them together. As with the other models, the combined model does well in matching the changes from 1993 to 2013.

The increases in BMI for this model are shown in Chapter 5 in Figure 5.13. These can be converted into calories simply by applying the values and equations described in this chapter. The changes are for the whole population, and the parameter values for age 30-60 from Section 13.4 were applied on the basis that they cover the largest part of the population. For example, for the median PAL value, from Table 13.4 the BMI increases need to be multiplied by 56.91 kcal / day for men and 34.99 kcal / day for women. The resulting values are shown in Figure 13.1. The values for the lower and upper PAL are also shown.

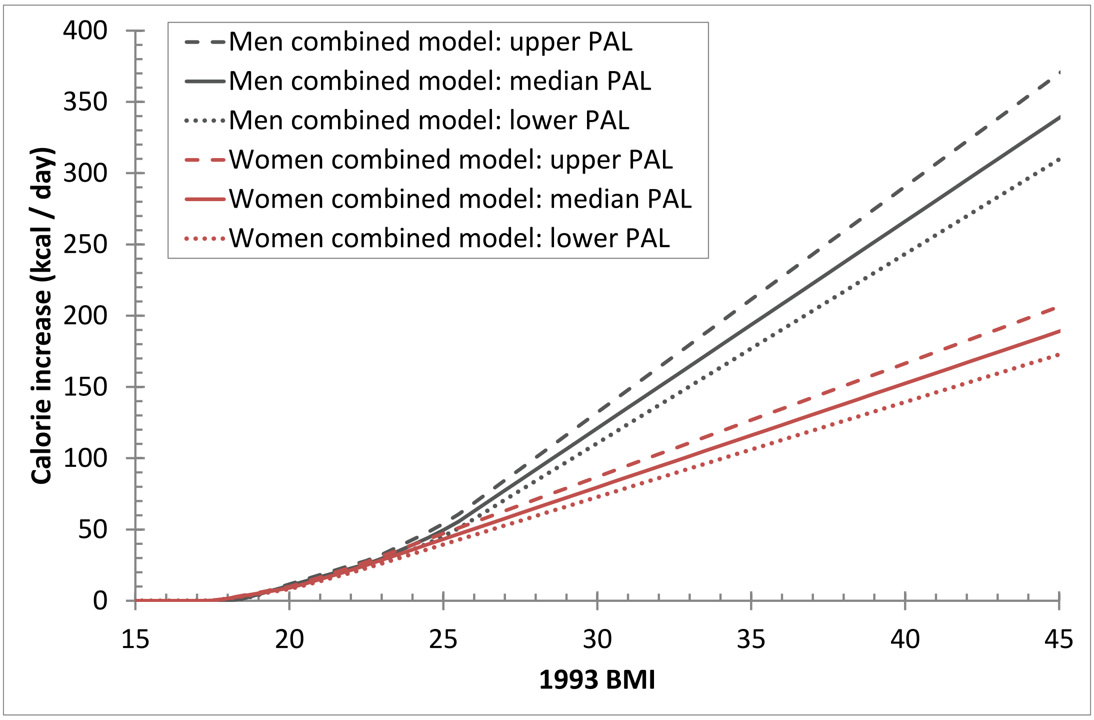


Figure 13.1 Changes in calories for the BMI changes of the combined model from 1993 to 2013.

Figure 13.1 gives the estimated extra calories for each equivalent percentile person in 2013 compared to 1993 using the combined scaling model. The implication is that men in 2013 who would have had a BMI of 25 in 1993, consume on average about an extra 50 kcal / day, with this being about 40 kcal / day for women (calculated values for median PAL of 50 and 43 respectively). Men who would have had a BMI of 30 in 1993 consume about an extra 120 kcal / day on average, while for women this is about 80 kcal / day (calculated values for median PAL of 121 and 80 respectively). One important point is that there are few people with a BMI of over 35 in the 1993 BMI distributions and so the BMI increases and hence the calories for high BMI values are particularly uncertain.

As discussed in Section 13.1.2, this is population level analysis rather than for individuals and is based on equilibrium energy balance. Hence it is an estimate of the average difference in daily calories consumed at the population level. The calorie amounts in Figure 13.1 are quite small for the common BMI values up to 35, particularly for women.

Relating calories to specific food items, there was previously a page on the NHS website that gave a wide range of examples of 100 kcal. However, this does not seem to be on the website anymore. There are other websites that give examples of 100 kcal such as the website for the British Heart Foundation (BHF) [39]. Examples are given on the BHF web page of 100 kcal for various types of food such as fruit and vegetables (e.g., one large apple), and confectionary (3½ squares of milk chocolate, or a "tiny slice" of cake). The various examples given emphasise that the physical dimensions of the portions will vary considerably for different foods.

As presented in Section 6.3.3 on page 105, some analysis was done of the data for average calories per day for the U.K. (from Roser, Ritchie, and Rosado, 2013 using source data from the United Nations Food and Agricultural Organization). Step changes were noticed in the values for different time intervals, with an increase in the average values for 1985-1996 compared to 1974-1984 of 102 kcal / day, and then a further increase of 170 kcal / day for 1997-2019. Hence, the increase from 1985-1996 to 1997-2019 of 170 kcal / day is of a similar level to the increases calculated here for the most common 1993 BMI values between 20 and 35. There is therefore some consistency in these results and this gives additional evidence that this might be the level of changes in calorie consumption. However, there is some uncertainty with the U.K. calorie data since, as discussed in Section 6.3.3, it is based on retail sales rather than actual consumption and there could be differences over time in the levels of food waste. There is also the usual modelling uncertainty with results here, as discussed in Sections 13.6 and 13.7.

Overall, the results of converting the BMI changes from 1993 to 2013 into calories are that the calculated changes in daily calorie amount on a population level are small despite the large changes in BMI.

13.5.2 Calories for the changes from age 18 to 24

In the analysis of how BMI changes with age in Chapters 7-10, the greatest increases were found to be over the youngest ages. One part of the analysis looked at how the BMI distributions change with age in six year gaps (Sections 8.5.1 and 9.5.1) from 18 to 24 to 30 to 36, and so on up to age 54. This used HSE data for 1999-2014.

The largest increases in BMI are from age 18 to 24, with the mean BMI increasing from 23.2 to 25.1 for men and from 23.8 to 25.3 for women. One aspect of examining this was to calculate the BMI increases for each percentile and the results are shown in the Appendices in Figures 8.VI.4 and 9.VI.4 for percentiles from 5% to 95%. This section applies the energy balance equations to convert these BMI percentile increases into calories.

Since this analysis is considering changes over an age range of 18 to 24, the Table 13.4 values for the category of age 18-30 were used. These are higher than for the older age groups. The values for the three PAL values are 67.98, 74.37, 81.21 kcal / day for men and 41.17, 45.04, 49.18 kcal / day for women. Again, as in Section 13.5.1, the BMI changes just need to be multiplied by these values.

The results from converting the percentile values into calories are shown in Figure 13.2 for the three PAL values. These are for the 10% percentiles along with 5% and 95%. Each series is just the same curve as the original BMI percentile values but multiplied by the relevant value given in the previous paragraph. The results give an indication of the changes in equilibrium calories at a population level that lead to the considerable change in the BMI distributions that occur from age 18 to 24.

[39] https://www.bhf.org.uk/informationsupport/heart-matters-magazine/nutrition/weight/what-does-100-calories-look-like

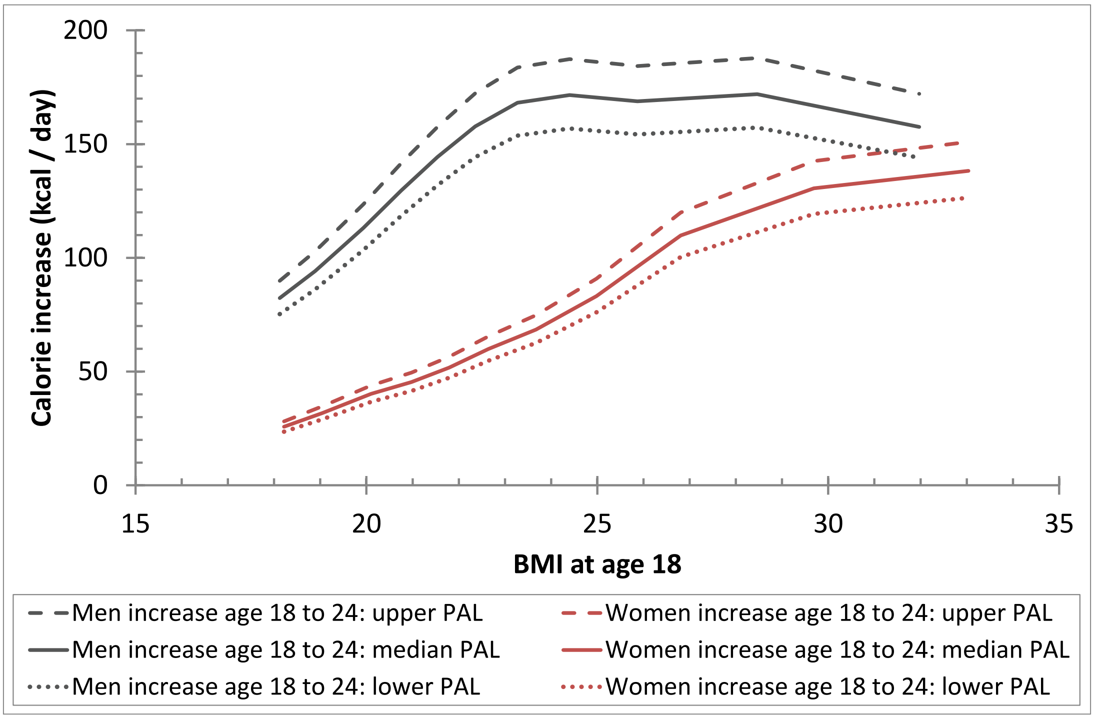


Figure 13.2 Changes in calories for the BMI changes with age from age 18 to 24.

The implication from the results in Figure 13.2 is that for men the typical calorie increase for those of BMI of 23 and above for median PAL is about 170 kcal / day. For women, the calorie increase amounts get larger with higher starting BMI. For those with BMI of 25 and above, for median PAL the calorie increase values are between 80 and 140 kcal / day. As in Section 13.5.1, these are quite small amounts.

Again, these are estimates of calorie differences at a population level assuming equilibrium. As explained in Section 13.1.2, values at an individual level may vary considerably. Section 13.6.2 also looks at some different models, and there is a reasonably large difference in the results.

Age 18 does have a sizable percentage in the underweight category for both men and women. The issue of being underweight is outside the main scope of this report. However, an increase in calories with age for that group would move them into or towards the healthy BMI category and so this is presumably likely to be beneficial.

13.6 Comparison with the Hall simulator and Mifflin-St. Jeor equation

Hall et al. (2011) developed a mathematical model to simulate the relationship between energy intake and body weight. They have made the simulator available as a web-based tool [40] where the user can enter details of current weight, height, sex, age, physical activity level, goal weight, and target date to achieve the goal weight. The tool calculates the amount that should be eaten to maintain the current weight, achieve the new weight at the target date, and then maintain the new weight. The expert mode on the web page produces a chart and values for the daily change in weight. There are also advanced controls where other values can be entered such as "% Calories from carbs". Hall et al. (2011) include some examples of model validation in their paper and the model performs well for those.

13.6.1 Mifflin-St. Jeor equation for BMR

Unless entered directly into the Hall et al. (2011) simulator, the calories for maintaining the current weight are calculated in a similar way to that described in Section 13.2, as PAL × BMR, but using the Mifflin-St. Jeor equation (Mifflin et al., 1990) for BMR (webappendix to Hall et al., 2011). The Mifflin-St. Jeor equation for BMR (which Mifflin et al. refer to as resting energy expenditure, REE) is:

$$\mathrm{BMR} = 9.99w + 625h - 4.92a + 166x - 161 \qquad (13.7)$$

where w is weight in kg, h is height in metres, a is age in years, and x denotes the sex being 0 for women and 1 for men.

Total energy expenditure is obtained by multiplying BMR by PAL:

$$\mathrm{TEE} = \mathrm{PAL} \times (9.99w + 625h - 4.92a + 166x - 161) \qquad (13.8)$$

Some example input values were entered into the web simulator and the starting energy intake values agreed with the answers from this equation.

Since BMI = w / h^2 then w = BMI × h^2 and so from Equation 13.8:

$$\text{Change in TEE for each unit difference in BMI} = \mathrm{PAL} \times 9.99h^2 \qquad (13.9)$$

This applies to men and women of all ages and so the energy increment depends only on height. In other words, the γ parameter value for the coefficient of weight in BMR is always 9.99, whereas in the Henry equations in Section 13.2 the γ value varies for men and women, and for different age groups. This assumes a fixed age. Equation 13.8 has an age variable and so there will be an age effect if modelling people as they increase in age or comparing populations at different ages.

Using the mean height values for age group 45-54 in Table 13.2 of 1.75 m for men and 1.62 m for women, and a PAL value of 1.63, the change in equilibrium energy requirements per unit difference in BMI are 49.87 and 42.73 kcal / day, respectively. These are fairly similar to the values of 56.91 and 34.99 calculated in Section 13.4 using the Henry equations. In fact, the average of the two values here ((49.87+42.73)/2 = 46.30) is very close to the average of the two Section 13.4 values ((56.91+34.99)/2 = 45.95), but the Section 13.4 values have a larger difference between the values for men and women.

Figure 13.3 compares the total energy expenditure values for the two models for BMI from 15 to 45 using these parameter values, with age 45 for the Mifflin St.-Jeor equation and using the Henry (2005) equation for age group 30-60. The values are fairly similar. The slopes of the lines are the values above for the energy per unit BMI and so differ slightly between the equations. As discussed before in this chapter, these are estimated equilibrium values at a population level, and actual values may differ considerably for individuals.

[40] https://www.niddk.nih.gov/bwp

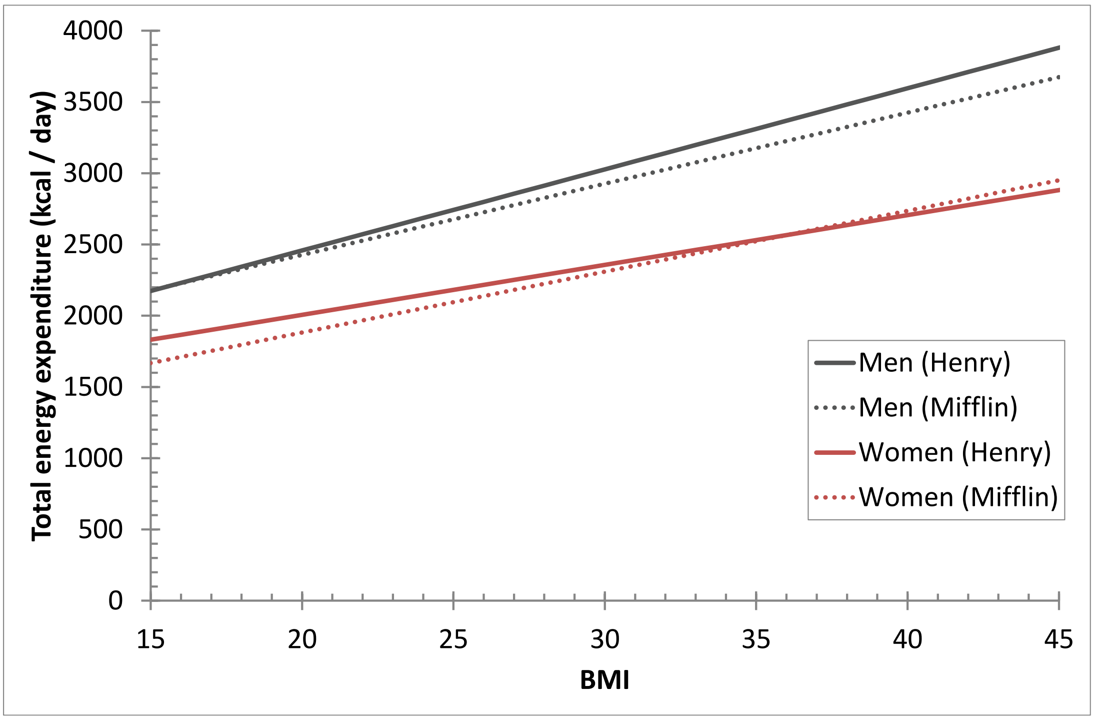


Figure 13.3 Total calorie requirements using two different equations.
[Calculated as PAL × BMR using either the Henry (2005) or Mifflin et al. (1990) equations for BMR. Parameter values for average height, median PAL, and age 45 (i.e., the Henry equation for age group 30-60)]

13.6.2 Hall simulator

The Hall et al. (2011) model calculates what happens when energy intake changes, using a more complicated set of equations that model the amount of fat, lean tissue, glycogen, and fluids, with differential equations for how these change as energy intake or expenditure changes. The details of the model are given in the webappendix to Hall et al. (2011).

The online simulator gives a value for the energy intake to achieve the weight change at the target date as well as the intake to then maintain the new weight. The simulator results, of course, vary depending on the target time for changing weight, particularly for the energy intake in the initial time interval to change to the target weight. Take the example of a man of age 45, height 1.75 m and PAL of 1.63 who is changing his weight to move from a BMI of 26 to 25, with no change in activity level. This corresponds to a weight change from 79.63 kg (26×1.75^2) to 76.56 kg (25×1.75^2). Using the simulator, the initial TEE value given is 2727 kcal / day, matching Equation 13.8. If the target time for the weight change is 30 days (a short time to achieve quite a sizable weight loss) then the simulator result for the energy intake value during those 30 days to achieve the new weight is 2066 kcal / day, and to maintain the new weight after the 30 days is 2657 kcal / day. For 180 days, 500 days and 1826 days (5 years) the energy intakes values during the time interval to change the weight are 2544, 2625, 2648, and the maintenance values after that are 2649, 2648, 2648. By contrast, if a starting weight of 76.56 kg is entered then the maintenance value given is 2677 kcal / day, which is the Mifflin-St. Jeor value from Equation 13.8.

This report is concerned with longer term effects of lifestyle changes and so it is the maintenance values that are of most interest. In this scenario, the change in maintenance values is from 2727 to 2648 for both the 500 and 1826 day scenarios. This is a change of 79 kcal / day which is larger than the values based on Henry of 57 kcal / day or Mifflin St. Jeor of 50 kcal / day (= $1.63 \times 9.99 \times 1.75^2$, or 2727 – 2677).

Consider a similar example for a woman of age 45, height 1.62 m and PAL of 1.63 changing weight from a BMI of 26 to 25, with no change in activity level. This is a weight change from 68.23 kg to 65.61 kg. The initial TEE value from the simulator is 2138 kcal / day, as per Equation 13.8. The simulator results for the energy intake values to change to the new BMI over 30, 180, 500, and 1826 days are 1597, 1982, 2054, 2076 kcal / day with maintenance values 2086, 2079, 2077, 2077 kcal / day. A long term change in maintenance values from 2138 to 2077 is a change of 61 kcal / day which again is higher than the Henry value of 35 kcal / day and the Mifflin St. Jeor value of 43 kcal / day.

For higher BMI values, the change in calories to reduce BMI by one unit is lower. For example, the simulator was used for the same examples as above for a long term change (1826 days) but changing BMI from 36 to 35. The change in maintenance calories for a man is from 3225 to 3163 which is a change of 62 kcal / day compared to 79 for the previous example. The change in maintenance calories for a woman is from 2566 to 2516 which is a change of 50 kcal / day compared to 61 for the previous example. The explanation in Hall et al. (2011) is that this is due to people with a higher weight having a greater percentage of fat. The basis of their equations is that fat has a lower energy requirement than lean tissue and so less energy change is required for people with a higher weight to gain or lose weight. This is of course different to using the Henry or Mifflin St. Jeor equations where the calorie values per unit BMI do not vary for different starting BMI values.

Experimenting with the online simulator for various weight change scenarios shows that the starting maintenance values are those based on the Mifflin St. Jeor equation (Equation 13.8), but the ending maintenance values differ from Equation 13.8. This implies that the steady state conditions for the detailed model equations differ from the Mifflin St. Jeor equations even though the Mifflin St. Jeor equations provide the starting equilibrium conditions. This is not surprising given the different approaches of the equations, but perhaps it reflects that the detailed equations are mainly designed to apply to weight changes made in a fairly short period of time.

13.6.3 The alternative models and the two scenarios

This section considers the two scenarios examined in Sections 13.5.1 and 13.5.2, and compares the Henry results in those sections with results using the Mifflin St. Jeor equation and the Hall simulator (if applicable).

First scenario of the change from 1993 to 2013

The first scenario considered in this chapter in Section 13.5.1 is the change in BMI from the 1993 to the 2013 population distributions. This is comparing the population distribution at two points in time that are quite far apart. The changes in BMI are for the combined scaling model which gives the increase in BMI for people at equivalent BMI percentiles in the two populations. The calorie increase is an estimate of the differences in the equilibrium values in the two populations. With the SACN model, using the Henry equations and height values for age group 30-60 from Table 13.4, the BMI increases were simply multiplied by 56.91 kcal / day for men and 34.99 kcal / day for women (Section 13.5.1).

If the Mifflin St. Jeor equation was used instead then the BMI increases would be multiplied by 49.87 and 42.73 kcal / day, respectively (Section 13.6.1). Hence, the calorie values for men would be slightly lower and for women would be slightly higher. For example, the BMI increase for the combined scaling model for a BMI value of 30 is 2.126 for men and 2.276 for women. As given in Section 13.5.1, with the SACN and Henry approach this is an increase in calories of 121 and 80 kcal / day respectively. With the Mifflin St. Jeor values this is an increase of 106 and 97 kcal / day respectively, and so these are fairly similar values.

The Hall simulator is designed for changes in people over time and so is not as applicable in this context where two different populations are being compared.

Second scenario of the change from age 18 to age 24

The second scenario in Section 13.5.2 is the change in BMI from ages 18 to 24 for the percentiles from 5% to 95%. This is a different type of scenario in that it is considering how BMI changes as the population ages. The Henry values in Section 13.5.2 for median PAL were compared with the Mifflin St. Jeor values and the Hall simulator. The parameters for this scenario are the median PAL of 1.63 and height values of 1.78 m for men and 1.63 m for women. The results are shown in Figure 13.4.

The Mifflin St. Jeor values in Figure 13.4 were calculated as follows. For each percentile there is a BMI value at age 18 and at age 24 from the BMI distributions (Section 8.5.1 for men and Section 9.5.1 for women). These two values were converted into calories using Equation 13.8 giving the equilibrium calorie values at age 18 and age 24. The calorie increase is then the age 24 calorie value minus the age 18 calorie value.

The Hall simulator values were obtained by entering the values into the simulator on the web page. These values are age 18, sex, height, and the starting weight and ending weight from the BMI values. Experimentation with a few examples showed that the end calorie maintenance value was the same whether the timescale was set to 500 days or 2190 days (6 years). Therefore 500 days was used as the simulator runs quicker (as suggested in a message box on the web page when large timescales are entered). The change in calories was then calculated as the end maintenance calorie value minus the start maintenance calorie value (which is the Mifflin St. Jeor value).

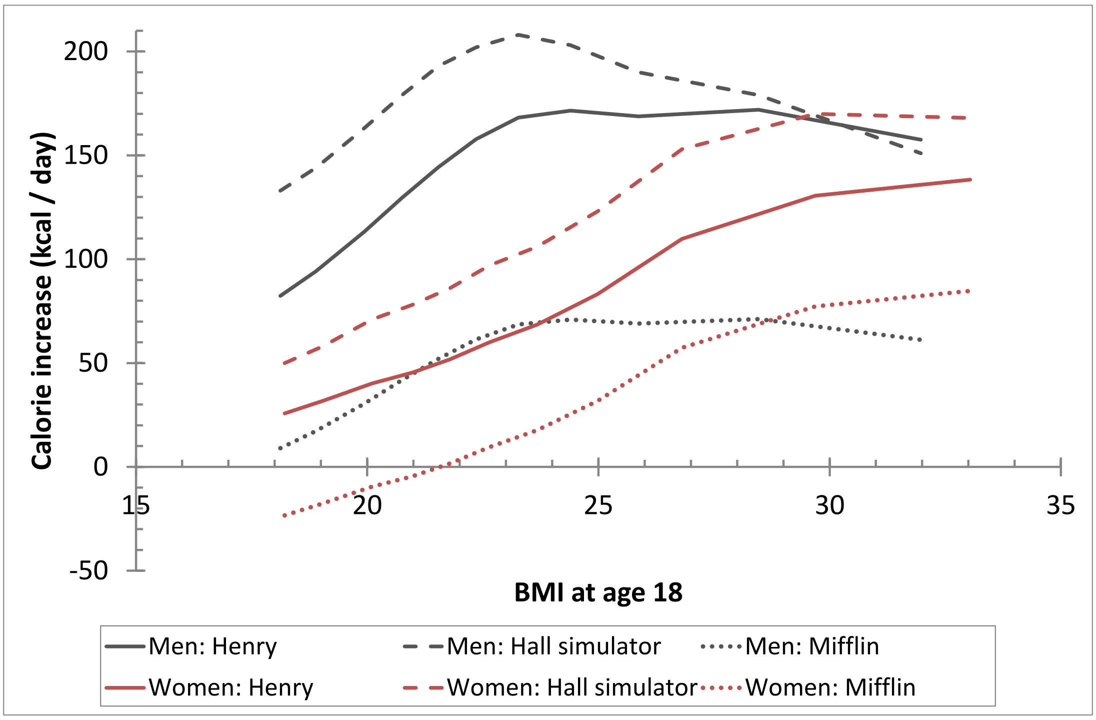


Figure 13.4. Changes in calories for the BMI changes with age from age 18 to 24 for three models.

For the Mifflin St. Jeor values from Equation 13.8 there are two components of the difference between the age 18 and age 24 calorie values. One is the effect of the difference in BMI and hence weight, and the increase in calories for this is given by Equation 13.9. The other is the difference in age since Equation 13.8 has a term for age in years with BMR reducing by 4.92 for each year of age.

This means that there are two elements of the difference between the model using the Henry equations and using the Mifflin St. Jeor equations. The first is the different multiplier per unit difference in BMI. As given in Table 13.4, the Henry equation values are 74.37 kcal / day for men and 45.04 kcal / day for women. From Equation 13.9, the Mifflin St. Jeor multipliers per unit BMI for this scenario of PAL and height values are 51.59 and 43.26 kcal / day respectively. So, a smaller value for men but very little difference in the value for women. The second element is the age term in the Mifflin St. Jeor Equation 13.8. BMR reduces by 4.92 for each year of age. From age 18 to 24 is 6 years, and so the effect when multiplied by the PAL value of 1.63 is a reduction in the equation calorie value of 1.63×6×4.92 = 48.12 kcal / day for both men and women. The total difference between the Henry and the Mifflin St. Jeor calorie values is the sum of the effects of these two elements.

In Figure 13.4 the highest calorie values for the data for men are between a BMI of about 23 and 29. The BMI increases obtained from the percentiles for the BMI distributions for age 18 and 24 are about 2.3 for these BMI values. Hence, the Henry values in Figure 13.4 using the multiplier of 74.37 are about 170 kcal / day (2.3 × 74.37 = 171 kcal / day). The lower multiplier for the Mifflin St. Jeor equation of 51.59 has the effect of making the value about 50 lower (2.3 × 51.59 = 119 kcal / day). The age effect reduces the value further by another 48 kcal / day. Hence, the Mifflin St. Jeor equation values in Figure 13.4 are about 100 kcal / day lower than the Henry values at about 70 kcal / day. Hence, there is quite a big difference between the results from the two equations in the values for men.

For women, the very small difference in the Mifflin St. Jeor multiplier compared to the Henry multiplier alters the calorie increase by less than 6 kcal / day. The overall difference is therefore about 50 kcal / day for all BMI values, due to the Mifflin St. Jeor equation age effect. Hence, the differences between the Henry and the Mifflin St. Jeor values in Figure 13.4 for women are about 50 kcal / day for all BMI values.

For both men and women, the Mifflin St. Jeor values in Figure 13.4 are small values. The highest value across the range of BMI values is 85 kcal / day (the value for women for BMI of 33.0). Hence, this model implies that the average increase in equilibrium calories at a population level to produce the increase in BMI from age 18 to 24 is small.

As shown in Table 13.4, there is some effect of age in the Henry equations since the multiplier values reduce from age category 18-30 to age category 30-60 mainly due to the Henry weight coefficient being smaller. However, ages 18 and 24 are in the same age category and so there is no age effect for the scenario considered here. Hence, a difference between the equations for this scenario is BMR reducing with age from 18 to 24 with Mifflin St. Jeor but not with Henry. I do not know if it is realistic or not for BMR to reduce over this age range. If it is then it presumably implies that, for a constant activity level, calorie intake needs to reduce slightly with age to maintain a constant weight and constant BMI (assuming height does not change).

As shown in Figure 13.4, the Hall simulator mostly produces higher values than those from the Henry equation. This is also the case in the examples above in Section 13.6.2. Those examples also found an effect that the change in calories per unit BMI reduces as BMI increases. The starting BMI values for the percentiles go from 18.1 to 32.0 for men and from 18.2 to 33.0 for women. Calculating the multiplier values from the web simulator results (calorie increase divided by the BMI increase), the multiplier values reduce from 120 to 71 kcal / day for men across the range of BMI values from 18.1 to 32.0, and they reduce from 87 to 55 kcal / day for women across the range of BMI values from 18.2 to 33.0. Increasing the timescale from 500 days to 2190 days does not alter the values and so the simulator presumably does not have an effect of BMR reducing with age over the course of the simulation time. This is probably because the simulator is mainly designed to model weight changes over a fairly short timescale.

Overall, there is quite a large difference between the calorie increase values for the three approaches for this scenario which implies some uncertainty as to the most realistic values.

13.7 Interpretation of the calorie results

The results in this chapter represent estimates of changes in equilibrium calorie intake values at a population level. The approach taken in Sections 13.1-13.5 follows that of the SACN (2012) report. As quoted in Section 13.1.2, the SACN report says that their values should not be applied to individuals, and the same is the case for the values in this chapter. This is because there are large differences between individuals in activity levels and in basal metabolic rate (BMR).

The values in the chapter are estimates of the typical or average changes at a population level. The scenarios examined are comparing populations, and the calorie increase values are the difference in the equilibrium values for the different BMI levels. This is different to someone changing their BMI over a short timescale which, from the Hall simulator results in Section 13.6, is likely to require much larger changes in calorie amounts.

The change in daily calorie amounts for the two scenarios are fairly small, and yet the difference in the BMI distributions is substantial. This indicates that the changes in the BMI distributions may result from quite small changes on average in equilibrium daily calorie consumption. For the changes from 1993 to 2013 this speculatively indicates that a fairly small change in average calories may be sufficient at the population level to return back to 1993 BMI levels. Again, these are equilibrium values and so are long term differences rather than short term changes.

The accuracy of the results here depends on the validity of the values and models used. The scientific authority of the SACN (2012) report was the main reason that their approach, including using the Henry (2005) equations, was followed in the analysis in Section 13.5. However, as shown by the different models in Section 13.6, there is some uncertainty in modelling the relationship between energy intake, energy expenditure, age and body weight and this uncertainty needs to be taken into account in considering and using the results. An assessment of the validity of the models and the energy balance approach is outside the scope of this report

Chapter 14. BMI distributions for 2017

Key points from this chapter:

- The work in the earlier chapters uses the HSE data up to 2014 since that was what was available at the time. This enabled BMI distributions to be produced for 1993, 1997, 2001, 2005, 2009, and 2013, with the results being given in Chapter 4. For the approach used in this report, the BMI distributions are generated from four years of HSE data. The 2017 distributions for men and women require the HSE data for 2015-2018. This became available before the report was completed and so the 2017 distributions could then be produced, analysed, and added to the report.
- This chapter gives the results for 2017 and for other analysis of the HSE 2015-2018 data. Most of the analysis in the earlier chapters in the report was updated and extended using this data. The results are presented in this chapter in the same order as the earlier chapters. This analysis was done after most of the analysis for the earlier data was completed and so one aspect of interest is to compare the new results with the patterns, expectations, and models from the previous results. A limitation with the HSE data for 2015-2018 compared to the earlier HSE data is that the age of the cases is only given in wider intervals, which are mostly five year intervals compared to a single year of age for the previous data. This restricted the analysis that could be done on how BMI varies with age.
- One slight change to the approach for producing the 2017 population distributions was that more smoothing was applied (using double smoothing) than for the earlier distributions in Chapters 3 and 4. The two smoothing versions were compared and they only make a very small difference to the results. The reason for using the extra smoothing here is that it was considered to give a more realistic shape for the distributions, particularly for the peak of the distribution for men.
- Statistics for the HSE data were calculated. The most noticeable aspect is that the mean and median values for women for 2017 and 2018 are considerably higher than the values for previous years, and the healthy percentage values are considerably lower. The data for men just continues the general trend in recent years of the mean increasing slightly with not much change in the median.
- The population BMI distributions produced for 2017 have the same general shape as the previous distributions. They continue the trend of BMI increasing over time with the peak getting lower and the right tail extending. The changes from 2013 to 2017 for men and for women are approximately linear scaling transformations, as found previously for the changes between the earlier distributions. This is evident from the increases in the percentiles being approximately linear and from the changes being able to be modelled well as linear scaling transformations.
- The overall pattern of the changes in the BMI distribution from 2013 to 2017 for men is very similar to the changes from 2005 to 2009. This makes the changes bigger in magnitude than from 2009 to 2013, but smaller than between each pair of neighbouring distributions from 1993 to 2005. Hence, the pattern of the changes generally getting smaller over time has not continued, although the changes are much less than those in the 1990s.
- The changes from 2013 to 2017 for women are fairly high. The magnitude is greater than any of the changes from one distribution to the next between 2001 and 2013, although not as high as from 1993 to 1997 or 1997 to 2001. The increase in BMI for the higher percentiles is particularly large.
- The pattern of the prevalence of very high BMI values is slightly different to the trends in other aspects of the distributions. The prevalence of high BMI values was calculated using BMI cut-off values of 38.6 for men and 41.3 for women, chosen to give a prevalence of about 2.5% (i.e., 1 in 40 people) in 2013. The percentage of people in these categories increases in approximately a linear pattern from 1993 to 2017, although with some variations. The increases from 2013 to 2017 are actually the highest, although only very slightly greater than the next largest, with the percentages in 2017 being 3.10% for men and 3.11% for women.

- The overall increase in BMI from 1993 to 2017 is large. The mean BMI increases by 1.58 for men and 1.71 for women. The healthy category percentage reduces by 9.7% for men and 10.8% for women. Looking at the highest BMI category, the severely obese percentages for men and women in 2017 are 2.2% and 4.1% compared to just 0.3% and 1.4% in 1993 (and 1.7% and 3.4% in 2013). The relationship of the increases in the BMI values of the percentiles against 1993 BMI is approximately linear with small increases in the lower percentiles but very large increases in the higher percentiles. For example, the 90th percentiles have BMI increases of 3.05 for men and 3.57 for women.
- Some of the BMI changes from 1993 to 2017 were converted into weight values using average height. The mean increases in BMI correspond to weight increases of 4.83 kg (10.7 lbs) for men and 4.48 kg (9.9 lbs) for women. The weight values for the percentile increases were also calculated and give large values for the high percentiles. For example, the values for the 90th percentile for men and women are each about 9.3 kg (or about 1½ stone).
- As for the previous distributions, there is a big difference in the shape of the BMI distributions for men and women. Specifically, for the 2017 distributions the frequency percentage for women compared to men is 8.9% higher for BMI < 24.0, 12.8% lower for 24.0 ≤ BMI < 34.5, and 3.9% higher for BMI ≥ 34.5. For the BMI categories, this means a higher percentage of men being overweight or obese (total percentages of 68.2% for men and 59.8% for women), but a higher percentage of women with high BMI values in the obese II and severely obese categories (total of 7.6% for men and 11.5% for women). Hence, the general issue of being overweight or obese is more widespread for men. The narrower issue of very high BMI values has a higher prevalence for women.
- The age group analysis shows particularly large increases in BMI for the 18-24 age group for both men and women for the 2015-2018 data compared to the data for 2007-2014. This is a concern as to what the future trajectory will be for this age group and subsequent cohorts. The mean values are above the aging models from Chapter 7, which are a hypothesised pattern representing current conditions. The samples sizes, though, are fairly small and so it will be important to examine future data as it becomes available. The age analysis that could be done was limited by the wider age groups in the 2015-2018 HSE data. However, the way that BMI varies with time and with age for the age groups for this data is generally consistent with the previous analysis in Chapters 7-10, albeit with some irregularities due to the small sample sizes. Chapters 7-10 give a much more detailed and thorough analysis of the age effect on BMI.
- The 2017 distributions have got closer than expected to the current trend model from Chapter 12. In particular, the distributions for both men and women have a higher severely obese percentage than the model, and this also applies to the obese II category for women. One reason for this is that the current trend model uses standard deviation values based on the 2011-2014 data, but the standard deviation values for the 2015-2018 data are higher than this. Adjusting the standard deviation to the 2015-2018 values brings the right tail of the current trend model to a similar or slightly higher level than the data. This updated version of the current trend model may be a better estimate of future BMI if conditions stay the same.
- The calorie amounts corresponding to the changes in BMI since 1993 from Chapter 13 were updated to use the results for the 2017 distributions rather than the 2013 distributions. The general pattern and the values are similar to those in Chapter 13, but are naturally slightly higher as a result of the BMI increases from 2013 to 2017. The calorie increase amounts are still fairly small, indicating than such levels can result in large increases in BMI at a population level. For example, for a BMI of 30 the calorie increase for the median activity level is 141 kcal / day for men and 102 kcal / day for women. These are average long term population values rather than individual values and depend on the accuracy and validity of the energy balance equations used, as discussed in more detail in the chapter on this (Chapter 13).

14.1 Overview and data availability

The work presented in the previous chapters in this report uses the HSE data for 1991-2014. This is what was available when the work started. As discussed in Chapters 2 and 3, for the method applied here, groups of four years of data are used to generate the BMI distributions and this has enabled distributions to be produced for 1993, 1997, 2001, 2005, 2009, and 2013.

The next distributions for men and women using this approach are for 2017 and these require the HSE data for 2015-2018. This data became available during the course of this work (the 2018 HSE data was released in December 2019), and so the methods presented in the earlier chapters were applied to derive and analyse the 2017 distributions. This chapter describes the results obtained for the 2017 distributions and for other analysis of the 2015-2018 data.

The work here on the 2015-2018 data was done after most of the work described in the previous chapters had been completed. One interesting aspect is therefore to see how the results compare with the changes and patterns identified in the previous chapters. For example, seeing any differences in the trends. Also, comparing the aging model from Chapter 7 and the current trend model from Chapter 12 with the new data, given that these models were produced without knowing the patterns of the new data, other than being generally aware of the HSE statistics produced each year (Sections 1.2.5 and 2.5).

One limitation of the HSE data for the years of 2015-2018 is that the age of each person is not given as an age in single years but only in wider age categories of 18-19 and then five year categories of 20-24, 25-29, etc. The calculation of the 2017 distributions therefore had to be adapted slightly to use this format of data. This also limited some of the age analysis that could be done.

The analysis that was done follows the methods explained in the earlier chapters. This chapter therefore focuses mainly on the results rather than the details of the methodology as this is already covered in the previous chapters of the report.

The sections in this chapter cover the following topics (in the same order as the earlier report chapters):

- 14.2: HSE data
- 14.3: Method for obtaining the 2017 population BMI distributions
- 14.4: Analysis of the 2017 population BMI distributions
- 14.5: Linear scaling transformations
- 14.6: Discussion of the interpretation of the results
- 14.7: Variation of BMI with age
- 14.8: Future scenarios
- 14.9: Calories
- 14.10: Summary

14.2 Initial analysis of the HSE data

14.2.1 Descriptive statistics

In Chapter 2 the general process for extracting the valid cases from the HSE is explained. The detailed procedure is set out in Appendix 2.I. Valid cases for recent data with the latest weighing scales are those with a valid value for the “bmiok” variable, or a value for the “BMIval2” variable. In the 2015-2018 data there are only four cases included that use the BMIval2 variable, which are one case in each year for men in 2015-2017 and one case for women in 2016.

Chapter 2 also has the analysis of the HSE data for each year. This is just initial analysis as the main focus of the work here is generating and analysing the BMI distributions for the four years of data combined. The Chapter 2 HSE results are extended in this section to include the 2015-2018 data.

The descriptive statistics for each year of the HSE data for 2015-2018 are in Appendix 14.I in Table 14.I.1. Figure 14.1 shows the mean and median values for all years from 1991-2018. The most notable aspect in the 2015-2018 values is that the mean and median for the data for women are much higher in 2017 and 2018 than in previous years, with the values for 2017 being particularly high. The samples sizes are fairly small (the samples for 2017 and 2018 are 3563 and 3603 cases) and so part of the difference may be due to sampling variation. In both years the mean for the data for women is slightly higher than for the data for men. This is unusual and the only other year in which this is the case is 2009. The mean values for men for 2015-2018 in Figure 14.1 continue the slight upward trend of recent years, with the gradient of the general trend since the year 2000 looking to be roughly constant. There is not much change in the median values for men from 2001 onwards. The average of the median values for 2015-2018 is very slightly higher than the average of the median values for 2001-2014 (27.05 compared to 26.86).

The healthy category percentages are shown in Figure 14.2 and have a pattern that is consistent with the mean and median values. The 2017 value for the data for women of 36.0% is much lower than the values for the earlier years. The value for 2018 is also low at 37.3%. For men the healthy percentages for 2015-2018 are similar to the other values since 2000 at about 30%. The healthy percentage in every year for men is much lower than the healthy percentage for women.

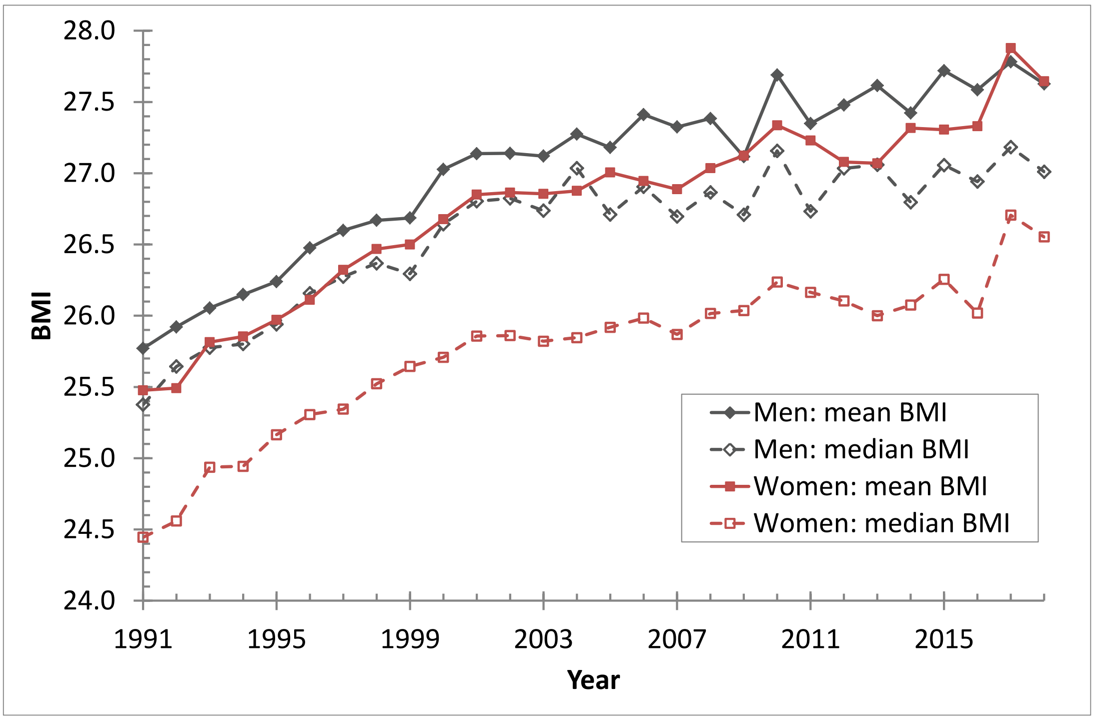


Figure 14.1 Mean and median BMI values for the HSE data for each year from 1991 to 2018.

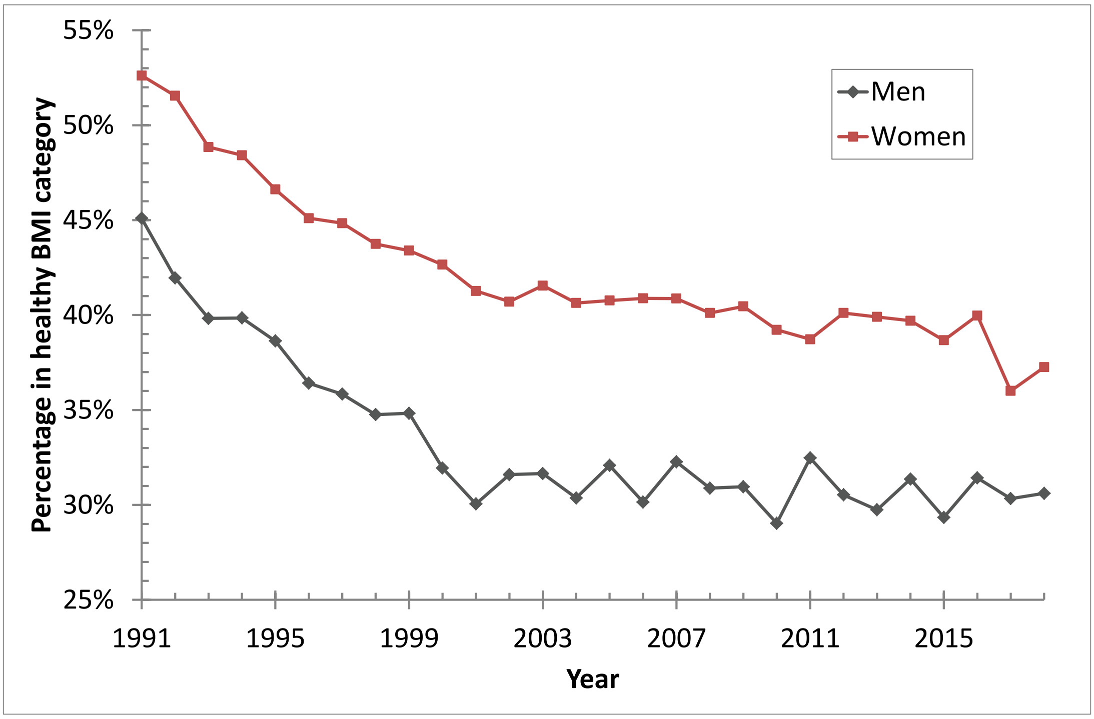


Figure 14.2 Percentage in the healthy category for the HSE data for each year from 1991 to 2018.

14.2.2 Missing values

Following the HSE analysis process that is explained in Chapter 2, the missing values were calculated and analysed for 2015-2018. The results are presented in Appendix 14.I. The patterns are consistent with those for 1991-2014 (Section+ 2.4) in the way that missing values are related to age. The missing value percentages are roughly constant up to about age 65 and then increase considerably as age increases. This is presumably due to the difficulties in obtaining measurements for some older people, particularly those who are unsteady on their feet. As discussed in Section 2.4.3, the consistency in the patterns of the missing values means that there is no particular cause for concern regarding the effects of the missing values on the analysis here, especially the comparisons over time.

The values for 2015-2018 also continue the patterns of the earlier data. The trend is for the percentage of missing values to increase over time with a strong correlation between the percentage of missing values for each year for men and women. The missing value percentages are high for 2016-2018 for men and women, all being over 17% (whereas the 2015 percentages are 13.9% for men and 15.3% for women). The speculation given in Section 2.4.3 is that the general tendency for an increase in missing values over time might be related to the greater prevalence of high BMI values and a possible reluctance of some people with high BMI values to be measured. This may mean that the proportion of high BMI values is underestimated slightly. However, it is not known if this is actually the case.

14.2.3 Trend Table values

For each year of the HSE data, work was done to try and reproduce the Trend Table / Data Table statistics, as explained in Section 2.5. This is to provide a check that the correct data had been extracted from the HSE. Most of the statistics match exactly for 2015-2018. The only difference is that there is one case in 2017 with a BMIval2 variable value (a case for the data for men with a BMI of 37) that is not included in the Trend Table statistics. It was decided to leave this in the data used here, for consistency with the way that the other years are treated and also as one case will make a negligible difference to the analysis.

14.2.4 High weight and BMI values

The statistics for cases over 130 kg were examined for 2015-2018 to extend the analysis in Section 2.6 in Chapter 2. This is mainly to consider the effect of changes in the weight limit of the weighing scales but also gives some information on high weight values. The chart of values is in Appendix 14.I (Figure 14.I.4) and shows that the percentage of such cases generally continues the increasing trend over time, although with some sizable variation from year to year in the 2015-2018 values for men. As discussed in Section 2.6, the consistency of the general trend without any big step changes gives confidence in the data even though the way that high weight values have been included in the HSE data has changed at certain points in time.

The prevalence of high BMI values of 40 and over for men and 45 and over for women were also calculated for 2015-2018 to extend the analysis and the chart from Section 2.7. The updated chart is given in Appendix 14.I as Figure 14.I.5. Again, the data continues the trend of previous years, which in this case is a roughly linear increase in the percentage of cases in these categories.

The main focus of the work in this report is on the population distributions from four years of data rather than the individual HSE years. High BMI values for the 2017 population distributions are specifically considered and discussed in this chapter in Section 14.4.4.

14.3 Obtaining the population BMI distributions for 2017

14.3.1 Data selection and producing the data frequency distribution

The population BMI distributions for men and women for 2017 were produced from the 2015-2018 HSE data using the method set out in Chapter 3. The first step was to combine the four years of data together.

The sample sizes are shown in Table 14.1. The number of cases for the other years are shown in Chapter 3 in Table 3.1 and vary from 12432 to 23241 cases for men and from 15076 to 26958 cases for women. The number of cases for 2017 is therefore the smallest sample size for both men and women.

Table 14.1 Number of HSE cases used for 2015-2018.

Year	Men	Women	Total
2015	3004	3642	
2016	2851	3489	
2017	2842	3563	
2018	2948	3603	
"2017"	11645	14297	25942

The case weights for the whole dataset were then adjusted so that the age profile of the complete sample matches that of the 2011-2014 population (Section 3.3). The age categories in the HSE data for 2015-2018 had to be used in this instance, which are 18-19 and then five year age categories (20-24, 25-29, etc.) rather than the single years of age for the previous data.

The percentage of the case weights in each 0.5 BMI interval were then calculated to give a frequency distribution for the data (Section 3.4). The data distribution was then smoothed using the method described in the next section.

14.3.2 Smoothing the data with the Savitzky-Golay method

The Savitzky-Golay method (Savitzky and Golay, 1964; Steinier et al., 1972) was applied to the frequency distribution for the data to smooth the distribution (Section 3.5). As explained in Section 3.5.2, for each of the population distributions three alternatives for the smoothing weights were tried. These were the 11 point and 13 point smoothing versions, and double smoothing where the 11 point smoothing is applied first and then 13 point smoothing is applied to the 11 point smoothed values.

For the population BMI distributions for the earlier years in Chapter 4, the three alternative weightings give very similar results. The double smoothing generally produces a slightly smoother curve but a visual comparison indicated that the differences are extremely minor. Since there is very little difference, the 11 point smoothing was used for all the distributions in Chapter 4 as this was considered to give sufficient smoothing (Section 3.5.2).

For 2017, there is a bit more of a difference than for the earlier years for the distribution for men. The 11 point smoothing gives a peak of the distribution that is slightly irregular with a small bump of higher values between a BMI of 25 and 26 and lower values between 26 and 27 compared to a smooth curve. This is likely to just be due to the random effects of the sampling combined with the particular smoothing equation. The 13 point method smooths the peak out to some extent and the double smoothing makes the curve even smoother. The differences in the three curves are still pretty small, and away from the peak the curves are visually extremely close together. For the data for women, the curves for all three smoothing options are very close to each other throughout the whole curve.

The smoothed distribution is an estimate of the BMI distribution for the adult population of England. Since BMI is a simple equation of the physical quantities of height and weight, over a large population the BMI distribution would be expected to be a smooth curve. On this basis, the double smoothing method was considered to produce the most plausible shape. For this reason, the double smoothing was used to produce the 2017 distributions. This was applied for both the distributions for men and for women, although as mentioned above the three options make very little difference for the distribution for women.

The data distributions and the smoothed distributions are shown in Appendix 14.II. There are also charts comparing the double smoothing and the 11 point smoothing.

Any of the three smoothing approaches are considered suitable. However, it may be that the double smoothing approach is the best of these three methods for future BMI data analysis.

14.3.3 Comparison of the 11 point and double smoothing versions

As explained in Section 14.3.2, the double smoothing version was used for 2017 compared to the 11 point version for earlier years. The double smoothing version was chosen as it was considered to give a more plausible shape for the distribution for men. There is very little difference for the distribution for women. Since this is a slightly different version of the smoothing than for the earlier distributions, a check was made that the effect on the results was small. This was done by repeating most of the analysis using the 11 point smoothed distributions for 2017.

For the distributions for men, the differences between the results for the 11 point and double smoothed distributions can be summarised as follows:

For the basic statistics in Section 14.4.2, the mean, standard deviation and skew are exactly the same because the Savitzky-Golay method maintains these values from the source data distribution. The median is the same to 2 decimal places. The BMI category percentages differ by at most 0.1% (four of the categories differ by this when rounded to the nearest 0.1%). For the percentiles in Section 14.4.3, the BMI values differ by at most 0.05. The high BMI values in Section 14.4.4 are not affected as these come from the data rather than the smoothed distribution. The changes in the statistics from 1993 to 2017 in Section 14.4.5 naturally reflect the differences in the statistics for 2017 and so the increases in the mean and standard deviation are the same for the two smoothing versions and the changes in the category percentages differ by at most 0.1%. Similarly, the percentile increases differ by at most 0.05. The scaling model from 2013 to 2017 for the 11 point smoothing is very similar to that in Section 14.5.1 with the same starting point of 24.75 and a scaling factor of 1.044 compared to 1.048. For example, the BMI increase at a BMI of 30 is 0.23 compared to 0.25. The fitness value is slightly higher (worse) at 6.45 × 10^{-6} compared to 4.54 × 10^{-6}. The direct scaling model from 1993 to 2017 for the 11 point smoothing is also very similar to that in Section 14.5.2 with the same starting point of 20.25 and a scaling factor of 1.216 compared to 1.220. The BMI increase at a BMI of 30 is 2.11 compared to 2.15. The fitness value is slightly higher at 61.77 × 10^{-6} compared to 54.84 × 10^{-6}. The combined scaling model from 1993 to 2017 is inevitably also very similar since the only difference is in using the scaling from 2013 to 2017 and these vary little between the double smoothed and 11 point versions. For example, the BMI increase at a BMI value of 30 is 2.45 for the 11 point version compared to 2.48 for the double smoothed version in Section 14.5.2.

For the distributions for women, the differences between the results for the 11 point and double smoothed distributions can be summarised as follows:

For the basic statistics in Section 14.4.2, the mean, standard deviation and skew are exactly the same because the Savitzky-Golay method maintains these values from the source data distribution. The median differs by only 0.01 to 2 decimal places. The BMI category percentages differ by at most 0.1% (three of the categories differ by this when rounded to the nearest 0.1%). For the percentiles in Section 14.4.3, the BMI values differ by at most 0.03. The high BMI values in Section 14.4.4 are not affected as these come from the data rather than the smoothed distribution. The changes in the statistics from 1993 to 2017 in Section 14.4.5 naturally reflect the differences in the statistics for 2017 and so the increases in the mean and standard deviation are the same for the two smoothing versions and the changes in the category percentages differ by at most 0.1%. Similarly, the percentile increases differ by at most 0.03. The scaling model from 2013 to 2017 for the 11 point smoothing is very similar to that in Section 14.5.1 with a starting point of 23.00 compared to 22.75 and with the same scaling factor of 1.067. For example, the BMI increase at a BMI of 30 is 0.47 compared to 0.49. The fitness value is slightly higher (worse) at 7.21 × 10^{-6} compared to 5.84 × 10^{-6}. The direct scaling model from 1993 to 2017 for the 11 point smoothing is also very similar to that in Section 14.5.2 with the same starting point of 19.25

and a scaling factor of 1.255 compared to 1.259. The BMI increase at a BMI of 30 is 2.74 compared to 2.78. The fitness value is slightly higher at 21.34×10^{-6} compared to 17.15×10^{-6}. The combined scaling model from 1993 to 2017 is inevitably also very similar since the only difference is in using the scaling from 2013 to 2017 and these vary little between the double smoothed and 11 point versions. For example, the BMI increase at a BMI value of 30 is 2.90 for the 11 point compared to 2.91 for the double smoothed in Section 14.5.2.

The visual difference between the 11 point and double smoothed distributions is small. The values in this section confirm that there is only a very small difference in the statistical results. Either smoothing version would be suitable, but as explained in Section 14.3.2, the double smoothed version was chosen as giving a more plausible shape for the peak of the distribution for men. The results given in the following sections for the 2017 population BMI distributions use the double smoothed version.

14.4 Analysis of the BMI distributions for 2017

The BMI distributions obtained from the work described in Section 14.3 were analysed and compared with earlier years. This follows the approach and the statistical analysis set out in Chapter 4. Results are given for the distributions (Section 14.4.1), descriptive statistics (Section 14.4.2), percentiles (Section 14.4.3), high BMI values (Section 14.4.4), the overall difference between the distributions for 1993 and 2017 (Section 14.4.5), and the comparison of the distributions for men and women (Section 14.4.6). There are also some additional charts in Appendix 14.III.

14.4.1 BMI distributions

The distributions for 2017 for men and for women are shown along with those for the previous years in Figure 14.3 and Figure 14.4. These update Figures 4.1 and 4.5 in Chapter 4. As explained in Sections 3.1 and 4.1, the distributions are plotted as frequency polygons. The points are the mid-point of the interval as the x value and the percentage frequency in the interval as the y value. The y value is a density, but is expressed per 0.5 BMI interval on the charts here. If the interval widths are 0.5, which is the case for most of the charts here including Figures 14.3 and 14.4, then the y value for each point is the percentage frequency for the interval. The exceptions are the scaling models, such as those in Section 14.5, where some of the intervals are wider and the y values are adjusted accordingly as explained in Section 5.2.

In each chart, the start of the left tail of the 2017 distribution is still basically the same as all the previous distributions. For the rest of the distribution, the previous trend over time continues of the distribution becoming more stretched out to the right with a lower peak and a longer right tail. Visually the changes from 2013 to 2017 look reasonably large, particularly in Figure 14.4 for the distribution for women.

For men, Figures 14.III.1-4 in Appendix 14.III are the charts of the distributions, cumulative distributions, the differences from the previous distribution, and the differences from the previous cumulative distribution. The changes from 2013 to 2017 are very similar to those from 2005 to 2009, with the series for the changes for 2009 and for 2017 in Figures 14.III.3 and 14.III.4 being close together. The changes are larger than for the previous time interval from 2009 to 2013, but smaller than for the earliest three intervals (1993 to 1997, 1997 to 2001, 2001 to 2005). One way of assessing this is with the lowest trough point for the cumulative differences in Figure 14.III.4, which is -1.9% at a BMI value of 29.5. The full list of trough values for the changes from the previous distribution (as given up to 2013 in Section 4.2.2) are: 1997 -3.8%, 2001 -5.0%, 2005 -2.6%, 2009 -1.8%, 2013 -0.8%, 2017 -1.9%.

For women, the same charts are Figures 14.III.5-8 in Appendix 14.III. The changes from 2013 to 2017 are quite large and, from the depth of the trough in Figure 14.III.8, the changes are greater in magnitude than for the previous three intervals (2009 to 2013, 2005 to 2009, 2001 to 2005). The changes are not as big as for 1993 to 1997 or for 1997 to 2001. The changes also occur at higher BMI values than for the earlier distributions with the lowest trough point at a BMI value of 30.0 with a cumulative difference of -2.5%. The full set of trough values (values up to 2013 from Section 4.3.2) from 1997 to 2017 are: 1997 -4.1%, 2001 -3.7%, 2005 -1.6%, 2009 -1.1%, 2013 -0.9%, 2017 -2.5%. To some extent, the changes happening at higher BMI values reflects a different starting point since the BMI values for 2013 are higher than for the earlier distributions. From Figure 14.III.7, the greatest reductions in prevalence from 2013 to 2017 are between BMI values of about 23 to 27, with increases in prevalence above 30 (including quite high increases between 31 and 33, and between 36 and 40). There are even reasonable increases for BMI values above 45, reflected in the cumulative difference in Figure 14.III.8 still being quite a bit below 0 at the BMI value of 45. High BMI values are examined in Section 14.4.4.

Tables of the values for all the distributions in Figures 14.3 and 14.4 are given in Appendix 14.III in Tables 14.III.3 and 14.III.4.

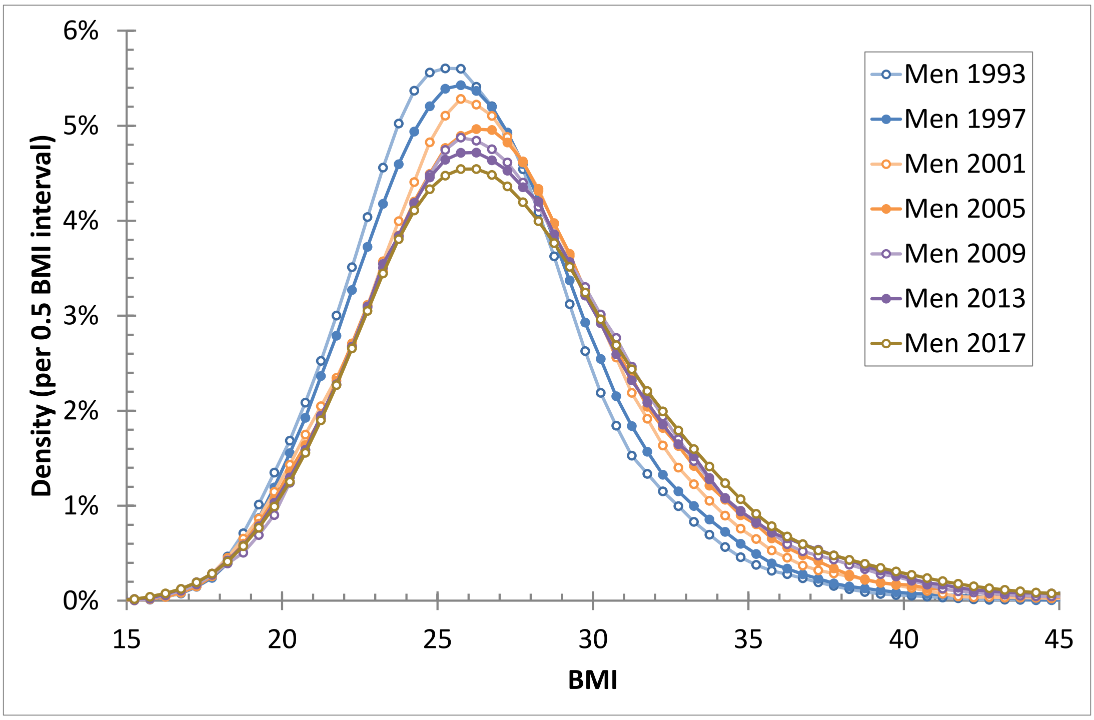


Figure 14.3 BMI population distributions for men for 1993 to 2017.

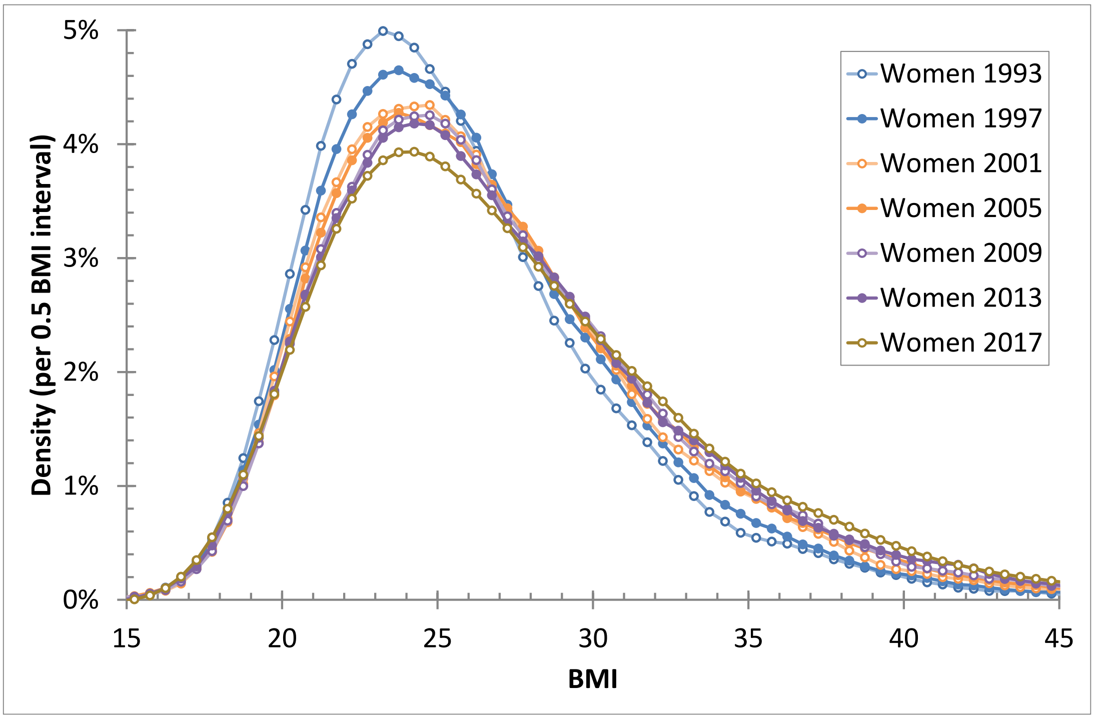


Figure 14.4 BMI population distributions for women for 1993 to 2017.

14.4.2 Descriptive statistics

The statistics for all the distributions are given below in Tables 14.2 and 14.3 (also in Appendix 14.III as Tables 14.III.1 and 14.III.2). These tables extend the Chapter 4 values given in Tables 4.1 and 4.2 to also include the values for the 2017 distributions. The statistics are discussed in the rest of this section which is split into a subsection for the mean, median, and standard deviation and a subsection for the category percentages.

Section 14.4.1 above has some analysis of the magnitude of the changes from 2013 to 2017 based on the charts of the distributions and charts of the differences between the distributions. These indicate that the changes for men are similar to those from 2005 to 2009 and the changes for women are quite large being greater than any successive distributions from 2001 to 2013 (but not as great as from 1993 to 1997 or 1997 to 2001). The statistics in Tables 14.2 and 14.3 and the discussions in this section consider specific aspects of the distribution such as the mean and each category percentage. There is some variation in the patterns, such as differences in the rates of change of the categories. However, the statistics and the discussions here are generally consistent with the comparisons in Section 14.4.1.

Table 14.2 Descriptive statistics for the population distributions for men.

	1993	1997	2001	2005	2009	2013	2017
Mean	26.05	26.41	26.90	27.19	27.39	27.42	27.63
Median	25.76	26.10	26.55	26.79	26.88	26.87	27.01
Modal interval mid-point	25.25	25.75	25.75	26.25	25.75	26.25	26.25
Standard deviation	3.87	4.00	4.29	4.48	4.67	4.84	5.04
Skewness	0.75	0.63	0.71	0.72	0.87	0.89	0.97
Underweight (BMI < 18.5)	1.1%	0.9%	1.0%	1.0%	1.0%	1.2%	1.1%
Healthy (18.5 ≤ BMI < 25.0)	40.4%	37.2%	32.8%	31.6%	30.9%	31.3%	30.7%
Overweight (25.0 ≤ BMI < 30.0)	44.7%	45.2%	45.4%	44.2%	43.1%	42.4%	41.1%
Obese I (30.0 ≤ BMI < 35.0)	11.6%	13.8%	16.6%	18.1%	18.8%	18.2%	19.4%
Obese II (35.0 ≤ BMI < 40.0)	1.9%	2.4%	3.4%	4.1%	4.8%	5.2%	5.5%
Severely obese (BMI ≥ 40.0)	0.3%	0.5%	0.7%	1.1%	1.3%	1.7%	2.2%

Table 14.3 Descriptive statistics for the population distributions for women.

	1993	1997	2001	2005	2009	2013	2017
Mean	25.76	26.18	26.65	26.88	27.03	27.13	27.47
Median	24.88	25.34	25.72	25.90	26.01	26.07	26.33
Modal interval mid-point	23.25	23.75	24.75	23.75	24.75	24.25	24.25
Standard deviation	4.95	5.04	5.38	5.54	5.63	5.77	6.13
Skewness	1.32	1.05	1.06	1.05	1.05	1.06	1.11
Underweight (BMI < 18.5)	2.1%	2.0%	1.7%	1.8%	1.7%	1.9%	2.0%
Healthy (18.5 ≤ BMI < 25.0)	49.0%	45.0%	42.2%	40.9%	40.0%	39.6%	38.2%
Overweight (25.0 ≤ BMI < 30.0)	32.0%	33.6%	33.5%	33.2%	33.1%	32.7%	31.6%
Obese I (30.0 ≤ BMI < 35.0)	11.7%	13.5%	14.7%	15.5%	15.9%	16.0%	16.8%
Obese II (35.0 ≤ BMI < 40.0)	3.8%	4.3%	5.5%	6.0%	6.2%	6.4%	7.3%
Severely obese (BMI ≥ 40.0)	1.4%	1.7%	2.3%	2.7%	3.0%	3.4%	4.1%

Mean, median and standard deviation

This subsection considers the values of the mean, median and standard deviation in Tables 14.2 and 14.3. A chart of the mean and median values is shown in Figure 14.5.

For men, the mean increases with each time period but with a faster rate of increase between 1993 and 2001 than since 2001. The increase in the mean from 2013 to 2017 of 0.21 is very similar to the increase from 2005 to 2009 of 0.20 and is larger than from 2009 to 2013 which is just 0.03. The pattern in the median is similar. Again, the increase from 2013 to 2017 of 0.14 is similar to that from 2005 to 2009 of 0.09, whereas the value hardly changes from 2009 to 2013 with a reduction of 0.01.

For women, the mean and median increase in a similar way to each other but with slightly larger increases for the mean (except for 1993 to 1997). The increases in the mean and median from 2013 to 2017 of 0.34 and 0.26 are larger than any of the successive increases from 2001 to 2013, but not as big as those from 1993 to 1997 (0.41 and 0.46) and from 1997 to 2001 (0.48 and 0.38). The increases in the mean and median from 2013 to 2017 are larger for women than for men. A chart of the changes in the mean and median between successive distributions is in Appendix 14.III in Figure 14.III.10.

One other aspect from the statistics in Tables 14.2 and 14.3 is that the standard deviation increases each year for both men and women, as the distributions get more stretched out to the right over time.

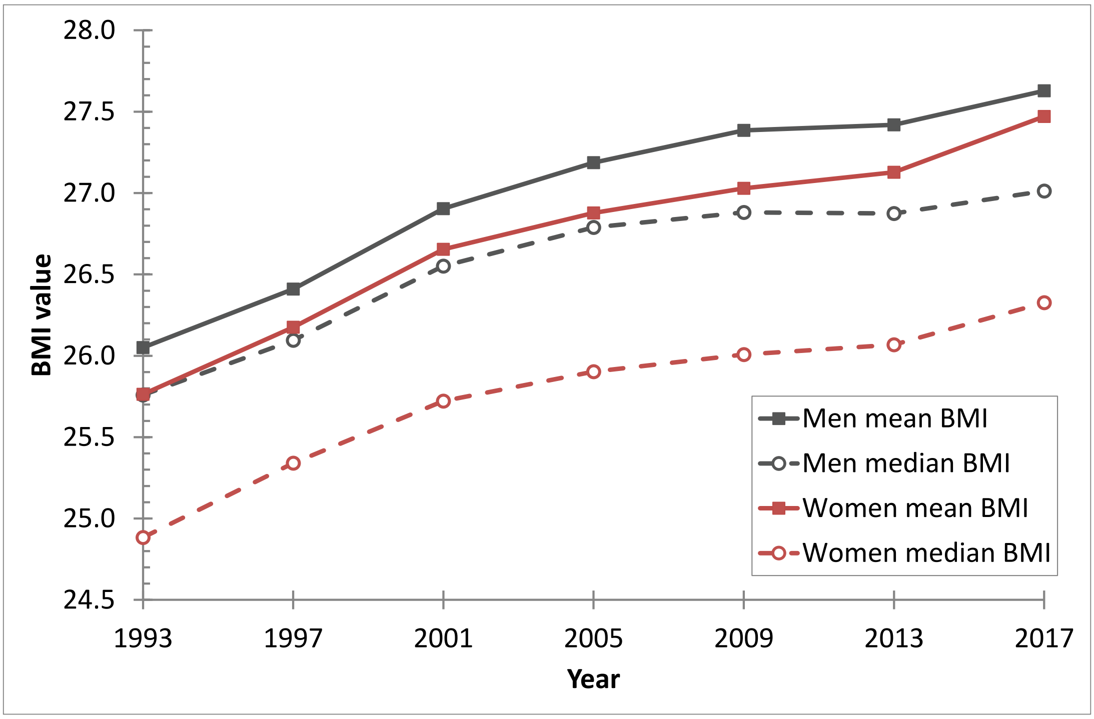


Figure 14.5 Mean and median values for the distributions.

Category percentages

The BMI category percentages in Tables 14.2 and 14.3 can also be compared over time. The differences between the percentages from one distribution to the next can be calculated and plotted. This was done in Chapter 4 in Figures 4.12 and 4.18. The changes for 2017 were added and the updated charts are Figures 14.6 and 14.7.

For men, the changes between successive distributions in Figure 14.6 again emphasise how similar the changes are for men in 2017 to those in 2009. The changes are smaller for 2013 and, as noted in Chapter 4, one unusual aspect for 2013 is that the healthy category percentage actually increases slightly. For 2017 the healthy percentage reduces and this reverses the increase for 2013 taking the healthy percentage to the lowest value for any distribution of 30.7% (slightly below the 30.9% of 2009). One notable aspect of the changes is that the increase in the severely obese category in 2017 of 0.5% is higher than for any of the previous years. The percentage in this category in 2017 is 2.2% compared to just 0.3% in 1993. The percentages increase for all the obese categories in 2017, and the combined percentage for obese I, obese II, and severely obese is 27.0% compared to 25.1% in 2013 and just 13.8% in 1993.

The changes in the category percentages for the successive distributions for women are shown in Figure 14.7. In 2017 the percentages reduce considerably for both the healthy and the overweight categories, with relatively big increases in the percentages for obese I, obese II, and severely obese. The reduction in the healthy category percentage is greater than for 2005, 2009 and 2013, but not as large as for 1993 and 1997. As for men, the increase in the severely obese percentage is the highest for any of the years, being 0.7%. This category has a percentage in 2017 of 4.1% compared to just 1.4% in 1993. The combined percentage for obese I, obese II, and severely obese is 28.3% compared to 25.8% in 2013 and just 16.9% in 1993.

Time series of the category percentages can also be produced. For example, Figure 14.8 shows the healthy category percentages and Figure 14.9 shows the severely obese category percentages. A full set of the charts is in Appendix 14.III, which has charts showing all the categories together in Figures 14.III.11 and 14.III.12, each of the individual categories in Figures 14.III.13-18, a chart of the total of the obese category percentages in Figure 14.III.19, and a chart of the total of the underweight and obese category percentages in Figure 14.III.20.

Figure 14.8 shows the faster rate of reduction in the healthy category percentages from 1993 to 2001 than since then. This is particularly the case for men where the value reduces by just 2.1% over 16 years from 32.8% in 2001 to 30.7% in 2017, compared to a reduction of 7.6% in the shorter time period of eight years from 1993 to 2001. As explained in Section 3.2, the 1993 distribution for both men and women is derived from larger sample sizes in 1993 and 1994 than in 1991 and 1992, and on this basis the time from the 1993 to 2001 distributions can actually be considered as more like 7½ years. Along with the charts and the median values this shows that the changes in the left side of the distribution for men have been quite small in recent years. For women, the healthy percentage reduces by more than for men from 2001 to 2017 at 4.1% but still less than the 6.7% over the shorter time from 1993 to 2001. The chart also illustrates clearly the much lower healthy percentage for men for each year. The difference for 2017 is 7.5% with the respective values being 30.7% for men and 38.2% for women.

The severely obese category percentages in Figure 14.9 increase fairly consistently over time for both men and women, and the percentages are higher for women than for men. As already mentioned in this section, the increases from 2013 to 2017 are actually the highest for both men and women between successive distributions. High BMI values are also examined in Section 14.4.4.

The differences between the distributions for men and women are considered in more detail in Section 14.4.6.

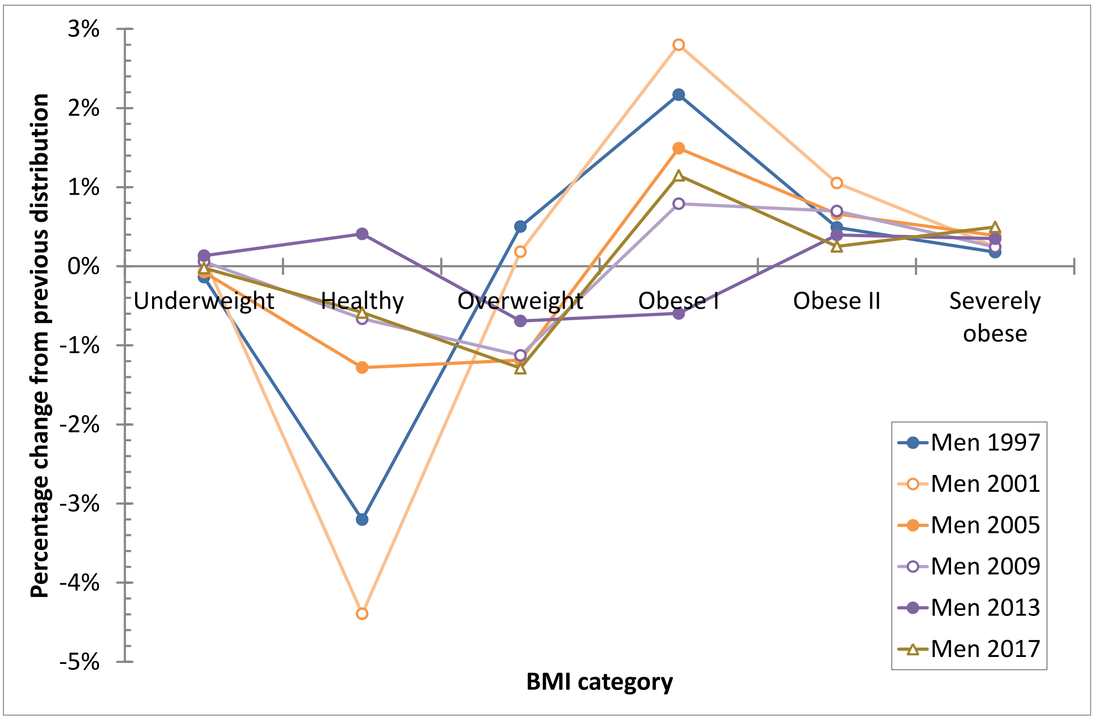


Figure 14.6 Changes in the BMI category percentages for the distributions for men.

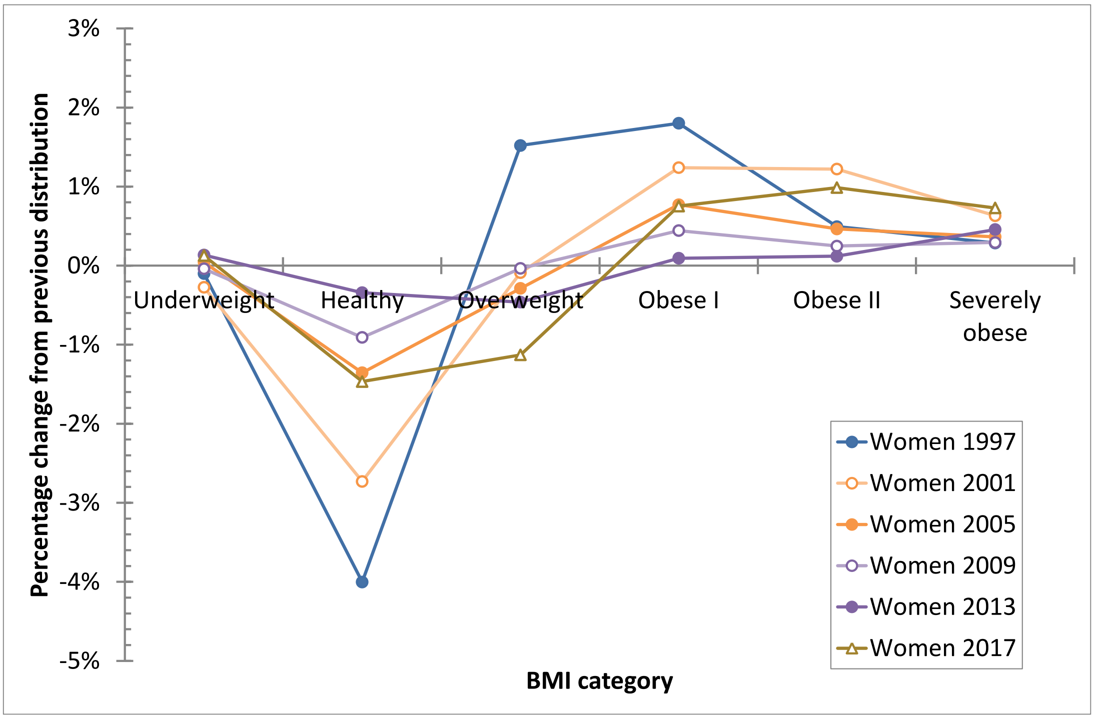


Figure 14.7 Changes in the BMI category percentages for the distributions for women.

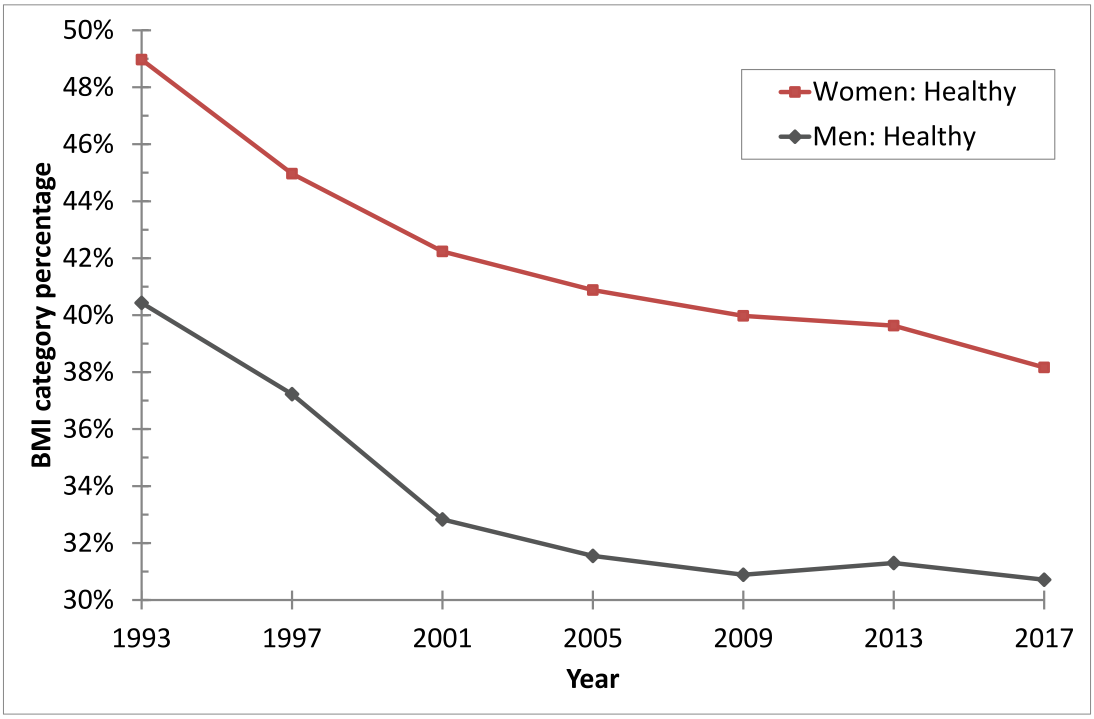


Figure 14.8 Healthy category percentages for the distributions for men and women.

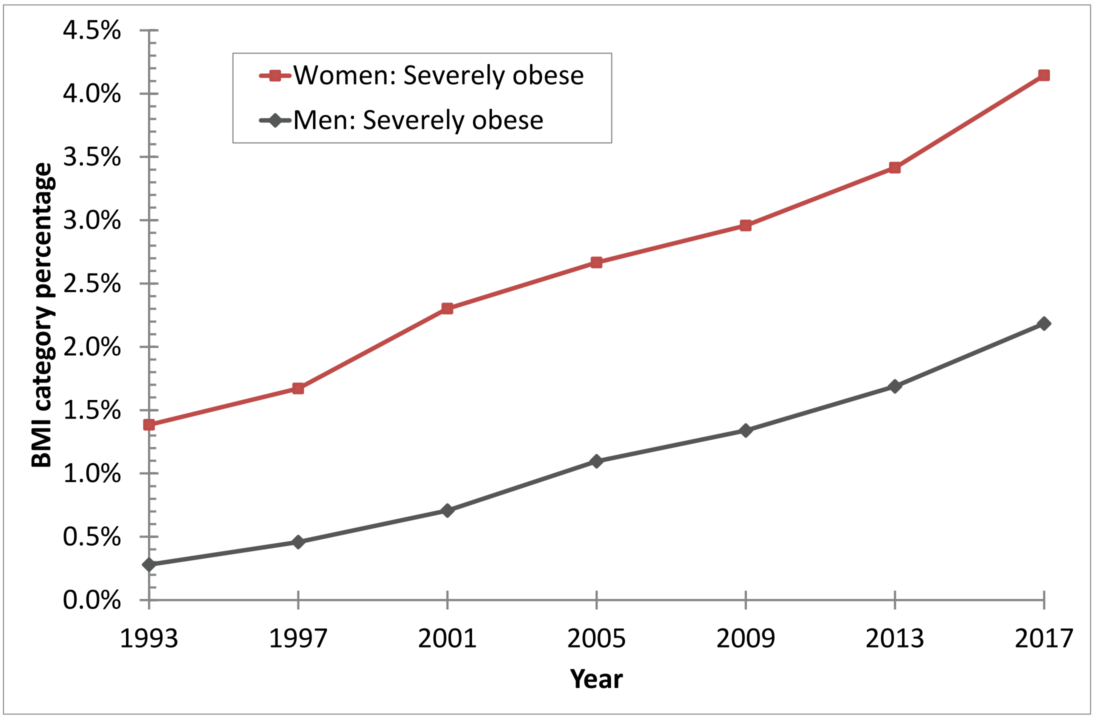


Figure 14.9 Severely obese category percentages for the distributions for men and women.

14.4.3 Percentiles

The percentile values were calculated for the 2017 distributions and the analysis described in Section 4.6 was extended to include these values. Figures 14.10 and 14.11 show the percentile increases compared to the previous distribution (adding the 2017 values to Figures 4.23 and 4.24 in Chapter 4). The values given are for each of the 10% percentiles along with 1%, 5%, 95%, and 99%.

For both men and women the changes for 2017 are approximately a linear increase starting from the x-axis. This indicates that the changes from the 2013 distribution to the 2017 distribution for men and for women are each approximately a linear scaling, as is also the case for the changes between each of the previous distributions. Modelling the changes as a linear scaling transformation is set out in Section 14.5.1.

The patterns in Figures 14.10 and 14.11 reinforce the comments in Sections 14.4.1 and 14.4.2 about the magnitude of the changes from 2013 to 2017. The percentile increases for men for 2017 are very close to those for 2009 indicating that the distribution has altered from the previous distribution in a very similar way. There is very little change in the percentile values up to a BMI of about 25 again indicating little change in the distribution up to that point. For women the percentile increases are quite large being generally the third largest behind 1997 and 2001.

One notable aspect in the percentile increases for women for 2017 is the large increase for the two points at the right end for 95% and 99%, showing that the highest BMI percentile values increase considerably and much more than for other recent distributions. These are the extreme values and the value for 99% is particularly uncertain as it depends on just a small number of high BMI values in the HSE sample data.

The increases in the percentile values for each distribution since 1993 are shown in Appendix 14.III in Figures 14.III.21 and 14.III.22, extending Figures 4.21 and 4.22 from Chapter 4 to include 2017. These are the total increases from 1993, and so these are equal to the cumulative sum of the values in Figures 14.10 and 14.11. As discussed in Section 4.6, the pattern is for each higher percentile to have a greater increase over time, which naturally follows from the patterns in Figures 14.10 and 14.11. The lower percentiles change very little whereas the higher percentiles have progressively larger increases from 1993 with each new distribution. This also corresponds to the observations already made about the distributions, that the start of the left tails stay much the same but the right tails stretch out more and more over time.

The charts in Figures 14.III.21 and 14.III.22 in Appendix 14.III show the percentile increases from 1993 for each of the distributions. However, the total percentile increases from 1993 to 2017 are of most interest in covering the widest time scale of the distributions, and these are presented and discussed further in Section 14.4.5.

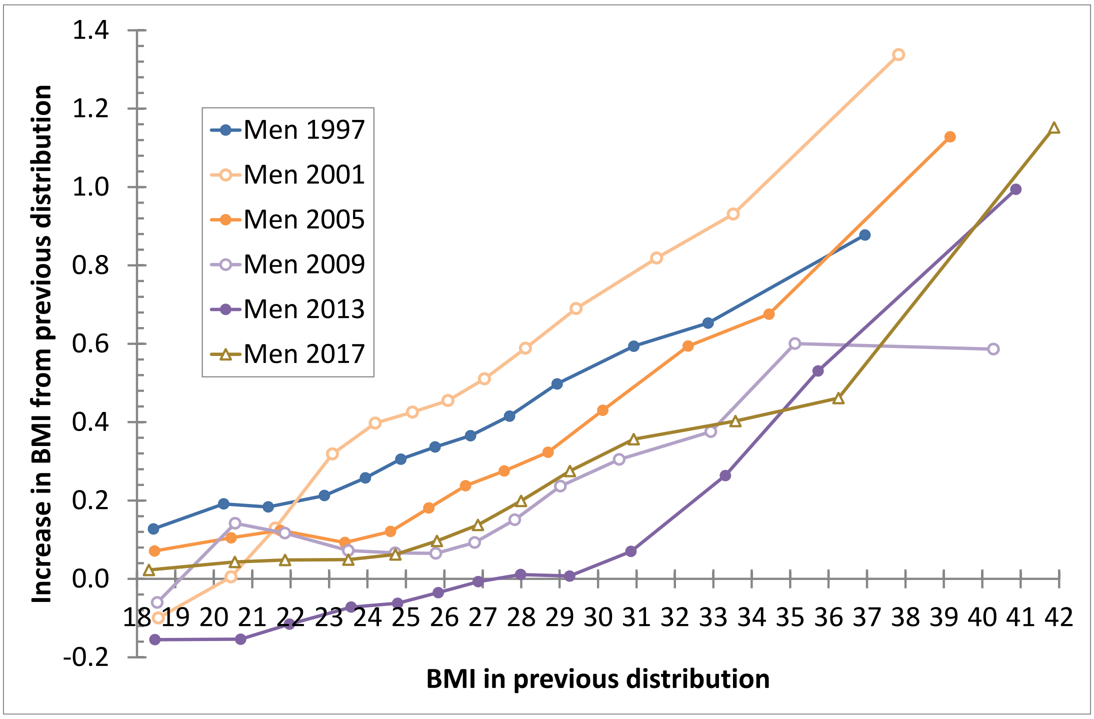


Figure 14.10 Increase in BMI against the previous BMI value for the percentiles for men.

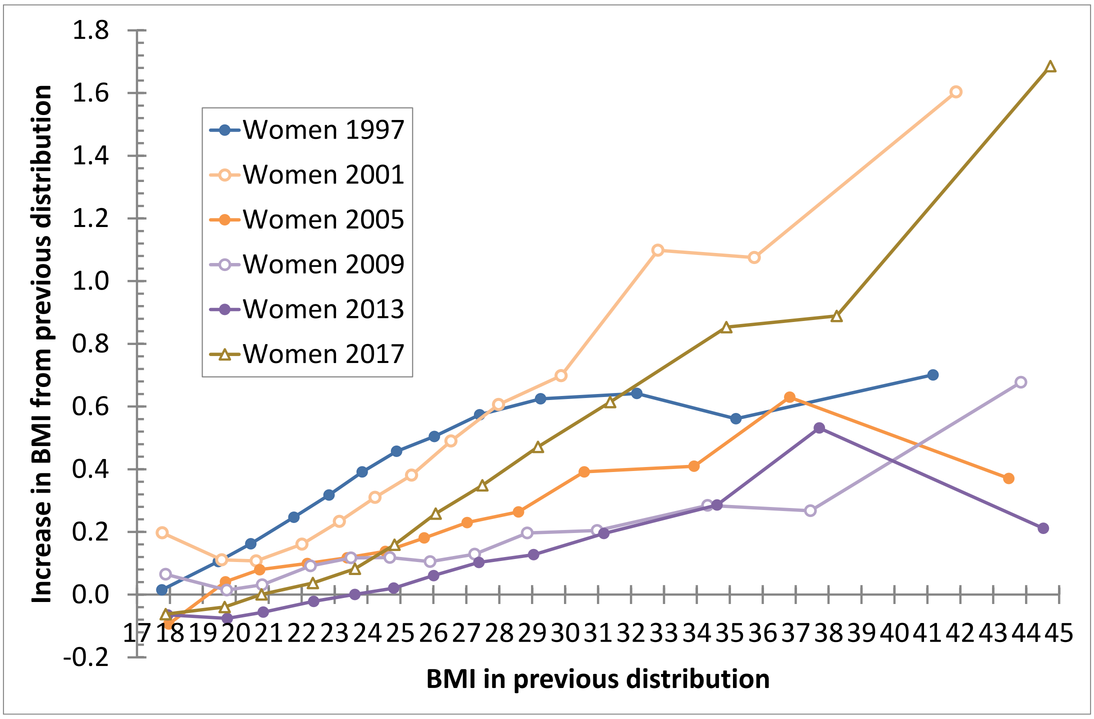


Figure 14.11 Increase in BMI against the previous BMI value for the percentiles for women.

14.4.4 High BMI values

Increases in high BMI values over time

Section 4.7 describes some work looking at high BMI values in the distributions. The cut-off values used were BMI of 38.6 and over for men, and BMI of 41.3 and over for women, and these were chosen so that the prevalence in 2013 is about 2.5%. This means that they are high values but still with a reasonable number of cases. The actual data (with the case weights) was used rather than the smoothed distributions so as to get the actual percentage frequencies. The percentages for each group of four years of data is shown in Chapter 4 in Figure 4.25 and are labelled by the distribution year (e.g., 2013 is the combined data for 2011-2014).

The percentage frequencies were calculated for the data used for 2017 and the updated chart is shown in Figure 14.12. The percentage of high BMI values increases by almost the same amount for both men and women, with the percentage frequencies for 2017 being 3.10% and 3.11% respectively (compared to 2.51% and 2.53% in 2013). Hence, for the 2015-2018 data the respective cut-off values of 38.6 and 41.3 still give very similar prevalence for men and women.

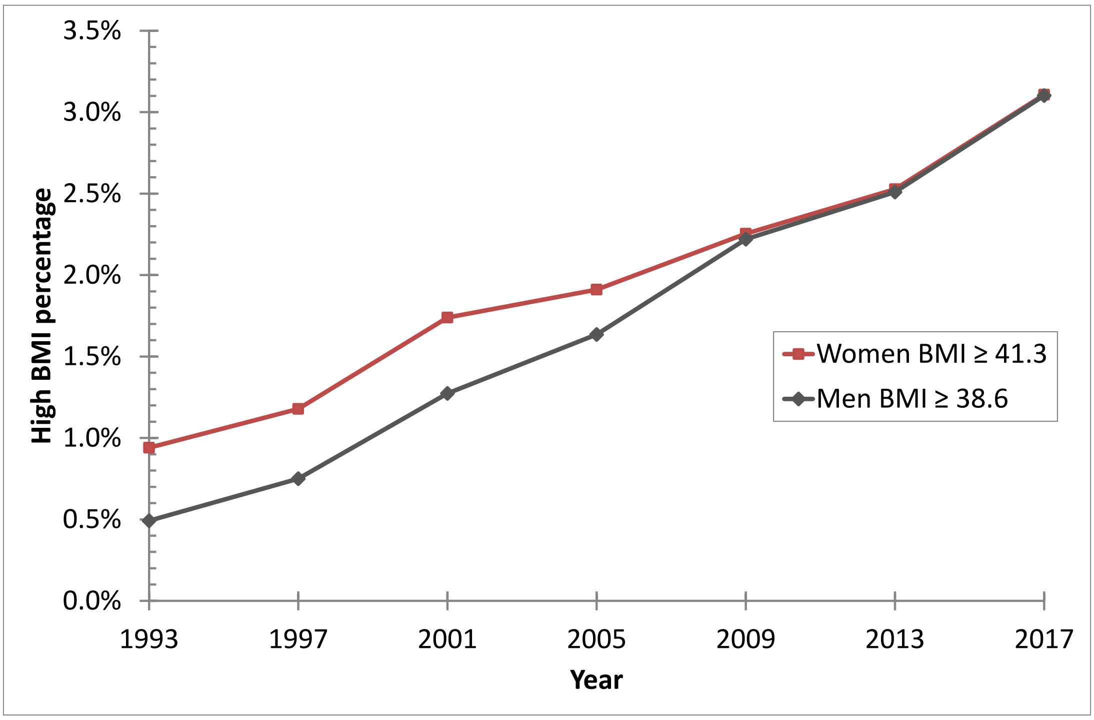


Figure 14.12 Percentage of high BMI values in the groups of four year data.

Figure 14.12 shows that the percentage increases roughly linearly for men and for women over the time period from 1993 to 2017. These are fairly small sample sizes and there is some variation from year to year. The increases compared to the previous year for men are: 1997 0.26%, 2001 0.52%, 2005 0.36%, 2009 0.58%, 2013 0.29%, 2017 0.59%. The increases compared to the previous year for women are: 1997 0.24%, 2001 0.56%, 2005 0.17%, 2009 0.34%, 2013 0.27%, 2017 0.58%. Hence, the increases for 2017 are actually the highest for both men and women, although only slightly greater than the next largest in each case.

The increases in the percentages in these categories must get lower at some point in the future and it will be interesting to see the level of increases in the next few years of data. In the very long run a continuation of a linear increase would eventually end up with 100% of people in these categories, and this will presumably not be the case in reality. For sure, the percentage cannot increase above 100%!

Prevalence of high BMI values by age group

Also considered in Section 4.7 in Chapter 4 is the prevalence of these high BMI values by age categories. The age categories used are 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+. The same analysis was done for the 2017 data, and the extended charts of age category prevalence of these high BMI values are shown in Appendix 14.III in Figure 14.III.23 for the data for men and Figure 14.III.24 for the data for women.

Some sampling variability would be expected with the small number of cases for the age groups but generally the prevalence of these high BMI values increases in 2017 for most age groups and the age group pattern is similar to previous years. Over time, the increase in high BMI values has occurred fairly consistently across the age groups. With the small sample sizes, there is some variation between the values for the distributions for different years, but generally the high BMI values are most common for ages 40s, 50s and 60s. The prevalence is still reasonably high for the younger age groups, particularly for the data for women.

14.4.5 Overall change from 1993 to 2017

Comparison of the distributions for 1993 and 2017

The 1993 and 2017 distributions can be compared to see the extent of the changes across the timescale of the HSE. The distributions are shown in Figure 14.13, and for both men and women there is a very large difference between 1993 and 2017.

For both men and women, the end of the left tail of the 2017 distribution is basically the same as the distribution for 1993. Then, in each case, the 2017 distribution is much more stretched out to the right than for 1993, with a lower peak that is further to the right, and a longer right tail. The differences from 1993 to 2017 are of course similar to those between 1993 and 2013 analysed in Section 4.8, but slightly greater as a result of the further increases in BMI from 2013 to 2017.

The differences between the 1993 and 2017 distributions can be plotted and this is done for both men and women in Appendix 14.III in Figure 14.III.25. The differences are very similar for men and women. For men, the percentages are lower for 2017 compared to 1993 up until a BMI value of 28.5, where the curves for 1993 and 2017 cross, with the total difference being 14.4%. In other words, compared to 1993, in the 2017 distribution 14.4% fewer men have a BMI < 28.5, and 14.4% more men have a BMI ≥ 28.5. For women the curves cross at a BMI of 27.5 with the total difference being 12.7%. Hence, 12.7% fewer women have a BMI < 27.5, and 12.7% more women have a BMI ≥ 27.5.

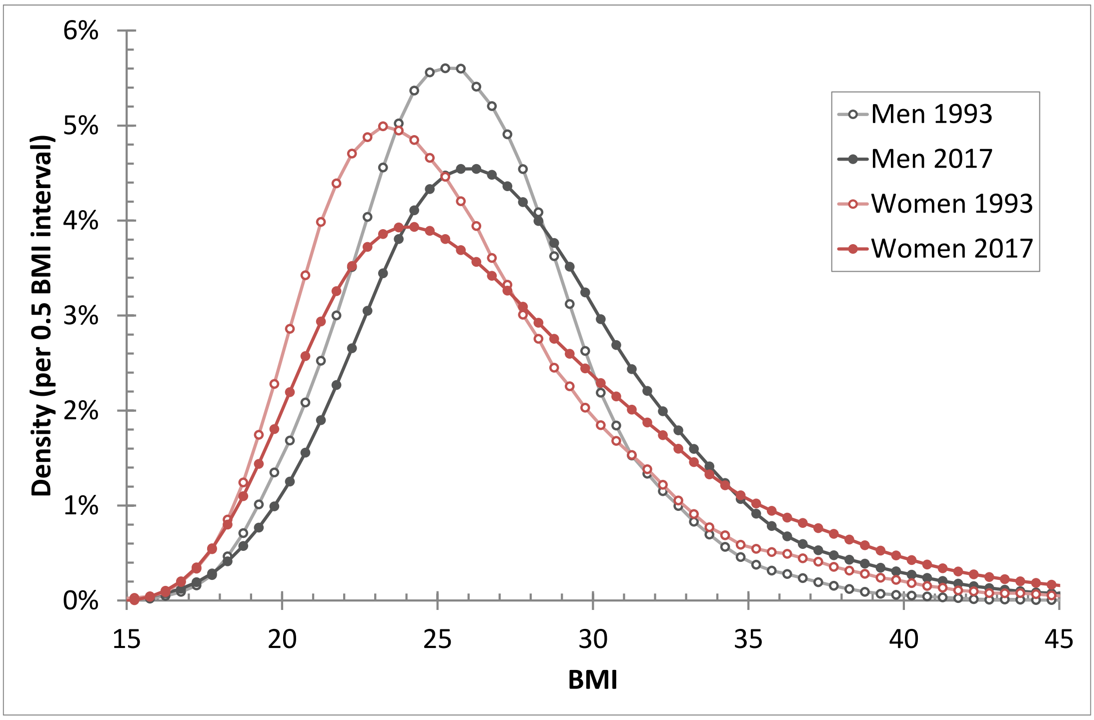


Figure 14.13 BMI distributions for men and women in 1993 and 2017.

Changes in the statistics from 1993 to 2017

The descriptive statistics for 1993 and 2017 are included in the Tables 14.2 and 14.3 in Section 14.4.2 earlier in this chapter, which show the statistics for all the distributions. The changes in the statistics from 1993 to 2017 are given below in Table 14.4. Bar charts of the category percentages are in Appendix 14.III (Figure 14.III.27 and Figure 14.III.28).

For men, the mean increases from 1993 to 2017 by 1.58 (from 26.05 to 27.63), the healthy category percentage reduces by 9.7% (from 40.4% to 30.7%), and the total for the three obese categories increases by 13.3% (from 13.8% to 27.0%). The underweight category percentage changes little and so the total increase for all the overweight and obese categories is very similar to the reduction in the healthy category percentage, being an increase of 9.6% (from 58.5% to 68.2%).

For women, the mean increases from 1993 to 2017 by 1.71 (from 25.76 to 27.47), the healthy category percentage reduces by 10.8% (from 49.0% to 38.2%), and the total for the three obese categories increases by 11.4% (from 16.9% to 28.3%). Again, the underweight category percentage changes little and so the total increase for all the overweight and obese categories is very similar to the reduction in the healthy category percentage, being an increase of 10.9% (from 48.9% to 59.8%).

Along with the visual difference in the distributions evident in Figure 14.13, Table 14.4 gives further information on the considerable extent to which BMI increases and obesity becomes more prevalent over the timescale of the HSE data.

Table 14.4 Change in the descriptive statistics for the population distributions from 1993 to 2017.

	Men	Women
Mean	1.58	1.71
Median	1.25	1.44
Modal interval mid-point	1.00	1.00
Standard deviation	1.17	1.17
Skewness	0.22	-0.21
Underweight (BMI < 18.5)	0.1%	-0.1%
Healthy (18.5 ≤ BMI < 25.0)	-9.7%	-10.8%
Overweight (25.0 ≤ BMI < 30.0)	-3.6%	-0.5%
Obese I (30.0 ≤ BMI < 35.0)	7.8%	5.1%
Obese II (35.0 ≤ BMI < 40.0)	3.5%	3.5%
Severely obese (BMI ≥ 40.0)	1.9%	2.8%

Increases in the percentiles from 1993 to 2017

The percentile increases from 1993 to 2017 were also calculated and are shown in Figure 14.14. As usual, the percentile values calculated are for each 10% along with 1%, 5%, 95%, and 99%.

For both men and women the values in Figure 14.14 show an approximately linear pattern increasing from the x-axis and this indicates that the change from 1993 to 2017 is approximately a linear scaling transformation. The changes are very similar for men and women even though the distributions for each year differ considerably. The percentile increases are naturally similar to those from 1993 to 2013 in Figure 4.33 in Section 4.8, but slightly greater with the further increases in BMI from 2013 to 2017. As discussed in Section 4.8.4, these results extend those of Wardle and Boniface (2008) who calculated and plotted the percentile values and the increases over time using the HSE data for 1993-1994 and 2002-2003.

The general gradient of the series for men and women in Figure 14.14 is quite steep and the higher percentiles increase by large amounts. For example, the 90th percentile value for men increases by 3.05 from 30.93 to 33.98, and for women increases by 3.57 from 32.18 to 35.75. The 95th and 99th percentiles increase by 3.85 and 6.08 for men, and by 3.95 and 5.25 for women. The lower percentiles with 1993 values in the healthy range have much smaller BMI increases of less than 1.5. The biggest of the calculated percentiles values in the healthy range are the 40th percentile for men with a 1993 value of 24.87 and an increase of 1.04, and the 50th percentile for women with a 1993 value of 24.88 and an increase of 1.44.

As with the other analysis in the report, these values shows that the changes in the obesity factors over time have tended to affect those with a higher BMI much more than those with a lower BMI. The pattern being close to a straight line implies that the magnitude of the overall effect of the changes in the obesity factors on average has an approximately linear relationship with BMI.

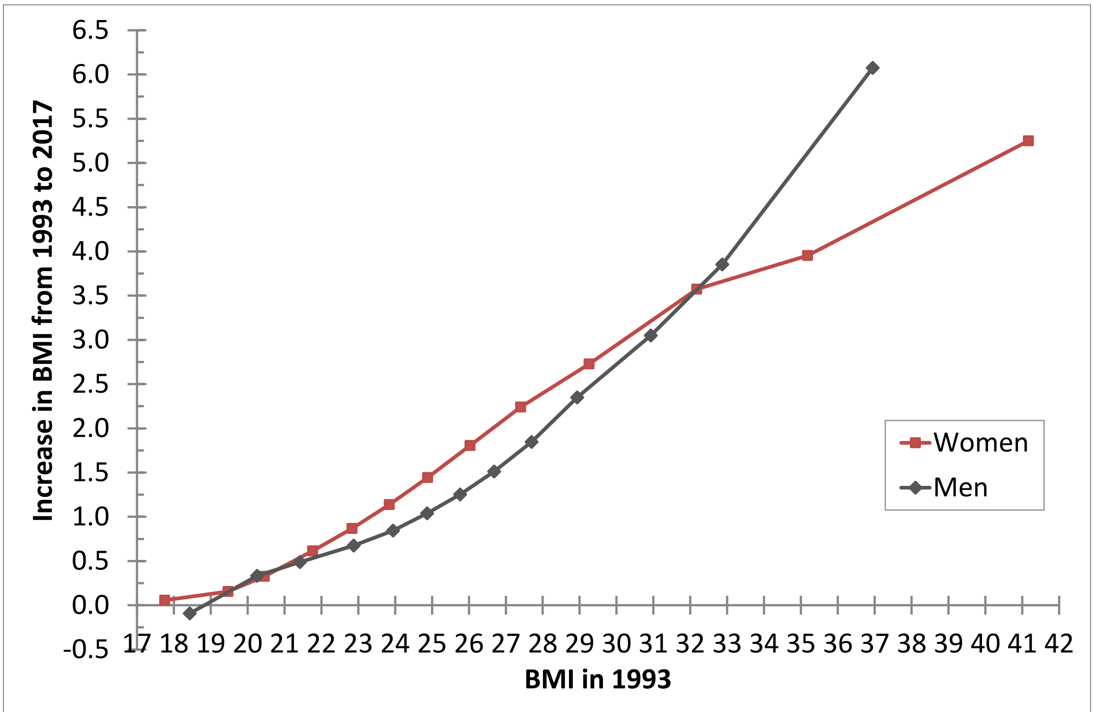


Figure 14.14 Percentile increases from 1993 to 2017 for men and women.

Weight values for the BMI changes from 1993 to 2017

As explained in Chapter 4 in Section 4.10, BMI values and BMI increases can be converted into weight values easily for a given height. Using units of kg and m, the weight is obtained by multiplying the BMI value by height squared. As in Section 4.10, this was done for the changes since 1993 in the mean and the increases in the percentiles. Here it is the changes to 2017 whereas in Section 4.10 it is the changes to 2013. Mean adult height values were used of 1.75 m (5 ft 8.9 in) for men and 1.62 m for women (5 ft 3.8 in), taken from the SACN (2012) values in Table 13.2 in Section 13.3.

The increases in the mean BMI from 1993 to 2017 of 1.578 for men and 1.708 for women correspond to increases in weight using mean height of $1.578 \times 1.75^2 = 4.83$ kg (10.7 lbs) for men and $1.708 \times 1.62^2 = 4.48$ kg (9.9 lbs) for women. The BMI values and BMI increases for the percentiles in Figure 14.14 can also be converted into weight values in the same way. Figure 14.15 shows the values in kg and Figure 14.16 shows the values in stone (st) and pounds (lbs), where 1 stone = 14 lbs. The values for the 50th, 70th, and 90th percentiles are given in Tables 14.5 and 14.6.

Weight is a more familiar variable than BMI and so these values help in seeing the magnitude of the changes and they emphasise further that the increases from 1993 to 2017 are substantial. For example, the weight increase values for the 90th percentile for men and women in Tables 14.5 and 14.6 are each about 9.3 kg (or about 1½ stone). Potentially, some analysis could also be done directly on the weight values given in the HSE data.

The changes can also be expressed as percentages, which are the same values whether the variable is BMI or weight. For example, the increases for the 50th, 70th, 90th percentiles are 4.9%, 6.7%, 9.9% for men and 5.8%, 8.2%, 11.1% for women. A chart of the percentage changes is given in Figure 14.III.26 in Appendix 14.III.

Table 14.5 Percentile increases for men from 1993 to 2017 as weights (using average height).

Percentile	1993 BMI	BMI increase	1993 weight	Weight increase
50%	25.76	1.25	78.89 kg (12 st 6 lbs)	3.84 kg (8.5 lbs)
70%	27.70	1.85	84.82 kg (13 st 5 lbs)	5.66 kg (12.5 lbs)
90%	30.93	3.05	94.72 kg (14 st 13 lbs)	9.34 kg (20.6 lbs)

Table 14.6 Percentile increases for women from 1993 to 2017 as weights (using average height).

Percentile	1993 BMI	BMI increase	1993 weight	Weight increase
50%	24.88	1.44	65.30 kg (10 st 4 lbs)	3.79 kg (8.4 lbs)
70%	27.41	2.24	71.92 kg (11 st 5 lbs)	5.88 kg (13.0 lbs)
90%	32.18	3.57	84.44 kg (13 st 4 lbs)	9.38 kg (20.7 lbs)

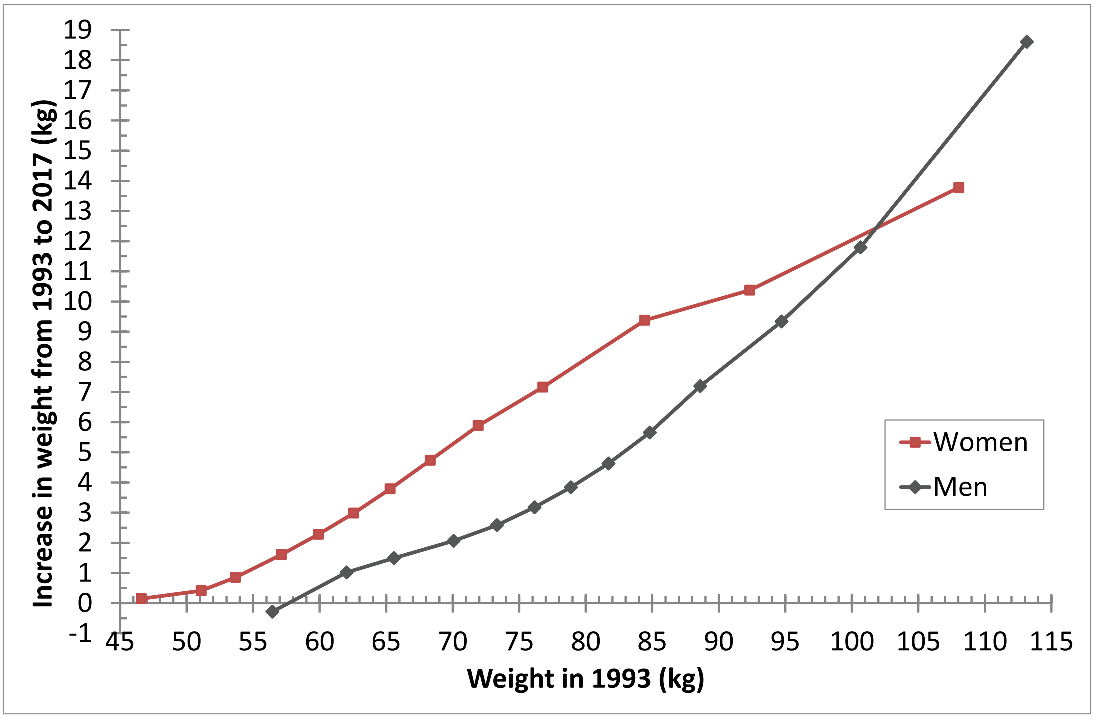


Figure 14.15 Percentile increases in kg, using average height, from 1993 to 2017.

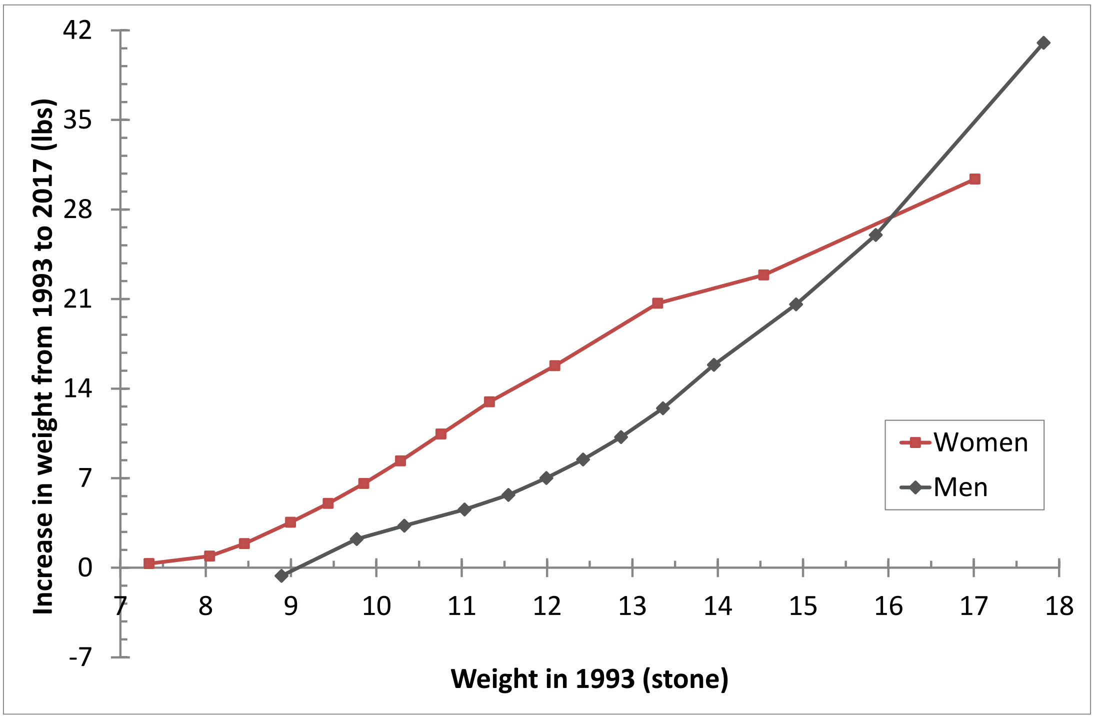


Figure 14.16 Percentile increases in stone and lbs, using average height, from 1993 to 2017.

14.4.6 Comparison of the BMI distributions for men and women

The distributions for men and women can also be compared with each other, as was done in Section 4.9. Looking at Figure 14.13 the shapes are clearly very different whether looking at 1993 or 2017. The distribution for women has a lower peak that is further to the left. The distribution is also more skewed with a longer right tail. The effect is that the prevalence is higher for women for lower BMI values up to about the peak of the distribution for women. Then the prevalence is higher for men around the peak of the distribution for men. Then at the end of the right tail the prevalence is higher for women. These intervals and the differences are given in Table 14.7. Positive frequencies mean that the value for women is higher, negative that the value for women is lower. The region is labelled in brackets by whether the values for women or men are higher (W or M). Hence, for 2017 there are 8.9% more women than men with BMI < 24.0, 12.8% fewer with 24.0 ≤ BMI < 34.5, and 3.9% more with BMI ≥ 34.5. A chart of the differences is also shown in Appendix 14.III in Figure 14.III.29.

Table 14.7 Intervals over which the frequency values for women are higher or lower than for men.

Region	1993 interval	1993 frequency	2017 interval	2017 frequency
Left tail (W)	BMI < 23.5	11.1%	BMI < 24.0	8.9%
Middle (M)	23.5 ≤ BMI < 31.0	-14.7%	24.0 ≤ BMI < 34.5	-12.8%
End of right tail (W)	BMI ≥ 31.0	3.6%	BMI ≥ 34.5	3.9%

The differences in the shapes of the distributions for men and women mean that the category percentages differ considerably. As discussed in Section 4.9.2, the differences between the distributions for men and women are very consistent across all the years. Taking the values for the seven distributions from 1993 to 2017, the mean differences for each category for women relative to men (women percentage – men percentage) are as follows (with the values for just the 2017 distribution given in the square brackets):

underweight 0.9% [0.9%], healthy 8.6% [7.5%], overweight -10.9% [-9.6%]
obese I -1.8% [-2.6%], obese II 1.8% [1.9%], severely obese 1.5% [2.0%].

As discussed in Chapter 4 in Sections 4.9.2 and 4.9.3, the differences mean that there is a lower percentage of men in the healthy BMI category (by between 7.5% and 9.4% for the different years) and a higher percentage overall in the overweight and obese categories. Specifically, for the 2017 distributions the healthy category percentages are 30.7% for men and 38.3% for women (Tables 14.2 and 14.3). The total percentage overweight or obese is 68.2% for men and 59.8% for women, with the difference being an 8.3% higher percentage for men. In this sense, the general issue of being overweight or obese is more widespread for men.

On the other hand, the prevalence of the highest BMI values is greater for women with about 3% or 4% higher frequency in the total of the obese II and severely obese categories for the different years. For the 2017 distributions the total percentage of obese II and severely obese is 7.6% for men and 11.5% for women, with the difference being 3.9% higher for women. Hence the narrower issue of very high BMI values is more prevalent for women.

Obviously, for both men and women there are considerable numbers in each of the overweight and obese categories and so these are just differences in relative prevalence.

14.5 Modelling the changes in the distributions with linear scaling transformations

14.5.1 Scaling models from 2013 to 2017

This section describes the application of the methodology in Chapter 5 to model the changes over time for the 2017 distributions as linear scaling transformations. The approach used, as described in Chapter 5, is to find the linear scaling that best transforms the initial distribution to the target distribution. The parameters that are altered are the scaling starting point and the scaling factor. The best transformation is defined as the one that minimises the fitness score, which is the sum of the squared differences up to a BMI value of 50 (Section 5.3). For the final model the scaling starting point BMI value is rounded to the nearest 0.25 and the scaling factor is rounded to 3 decimal places.

This approach was applied to model the changes in the distributions from 2013 to 2017. The scaling model distributions are shown in Figure 14.17. The parameter values and fitness score values are given in Table 14.8, which adds these results to those found for the earlier distributions that are given in Chapter 5 in Table 5.1. The models match the 2017 distributions well and the fitness values in Table 14.8 for 2017 are some of the lowest (best) values across all the models.

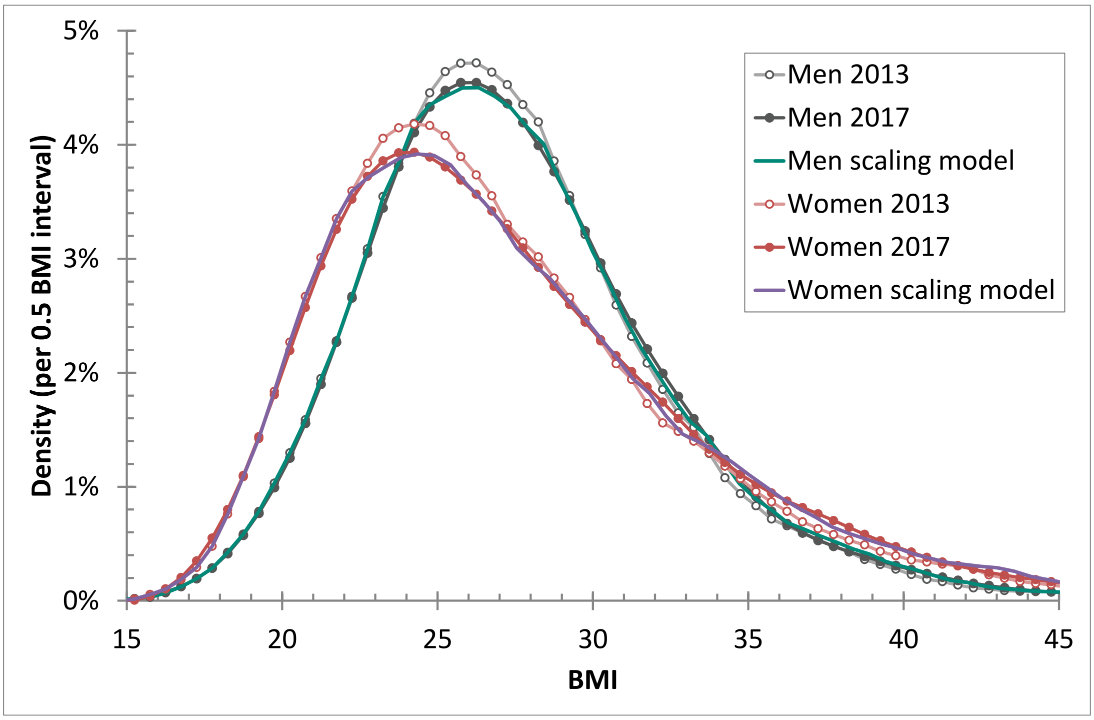


Figure 14.17 Scaling models for men and women for 2013 to 2017.

As discussed in Section 5.7, the scaling models can also be assessed by how the statistics compare with those of the actual distributions. Since the models start from the previous actual, the best comparison is to look at the changes from the previous actual. For example, for men the mean and median BMI of the 2017 scaling model increase by 0.16 and 0.10 from the actual for 2013, compared to increases of 0.21 and 0.14 for the actual 2017 distribution. For women the increase in mean and median for the scaling model is 0.32 and 0.22, compared to 0.34 and 0.26 for the actual. Hence, these are similar, particularly for the distributions for women. The changes in the BMI category percentages can also be compared and this is done in the charts in Appendix 14.IV. Again, the values are similar for the model and the actual.

The main conclusion is therefore that, as for the changes between the previous consecutive distributions, the changes in BMI from 2013 to 2017 have the general pattern of a linear scaling and can be modelled well using a linear scaling transformation.

Regarding the nature of the scaling models, for both men and women the models for 2017 have a high start point and also a high scaling factor. The scaling functions can be understood best by plotting the BMI increases. These are shown in Figures 14.18 and 14.19 (adding the 2017 function in each case to Figures 5.6 and 5.7). As would be expected these are similar to the percentiles in Figures 14.10 and 14.11.

The high start points again give an indication that the distributions in 2017 have changed little from 2013 in most of the left tail. This is evident in the previous analysis in this chapter such as the charts of the distributions in Figures 14.3 and 14.4 where for both men and women the curves for 2013 and 2017 only diverge to any noticeable extent just before the peaks of the distribution.

The high scaling factors mean that the distributions have changed quite a lot for higher BMI values. Consider a BMI of 30, which is a fairly high value but still with high prevalence. For men, the scaling model BMI increases for a BMI of 30 for the six time intervals are 0.51, 0.72, 0.38, 0.20, 0.10, 0.25 respectively. For women the increases are 0.79, 0.66, 0.29, 0.20, 0.17, 0.49. Hence, the increases are quite high in 2017 particularly for the data for women.

The patterns in Figures 14.18 and 14.19 reflect the earlier discussions in this chapter. The scaling model for men for 2013 to 2017 is very similar to that for 2005 to 2009. The scaling model for women has large increases in BMI which are not as quite as much as for 1993 to 1997 and 1997 to 2001, but bigger than the other models (apart from for very low BMI values).

Table 14.8 Final values for the scaling models.

Period	1993 to 1997	1997 to 2001	2001 to 2005	2005 to 2009	2009 to 2013	2013 to 2017
Men						
Start point	18.25	18.25	23.25	26.25	25.25	24.75
Scaling factor	1.043	1.061	1.056	1.052	1.021	1.048
Fitness value	3.73×10^{-6}	12.15×10^{-6}	6.88×10^{-6}	7.85×10^{-6}	9.06×10^{-6}	4.54×10^{-6}
Women						
Start point	17.50	20.25	18.00	17.25	22.75	22.75
Scaling factor	1.063	1.068	1.024	1.016	1.023	1.067
Fitness value	5.85×10^{-6}	8.79×10^{-6}	5.95×10^{-6}	7.69×10^{-6}	5.91×10^{-6}	5.84×10^{-6}

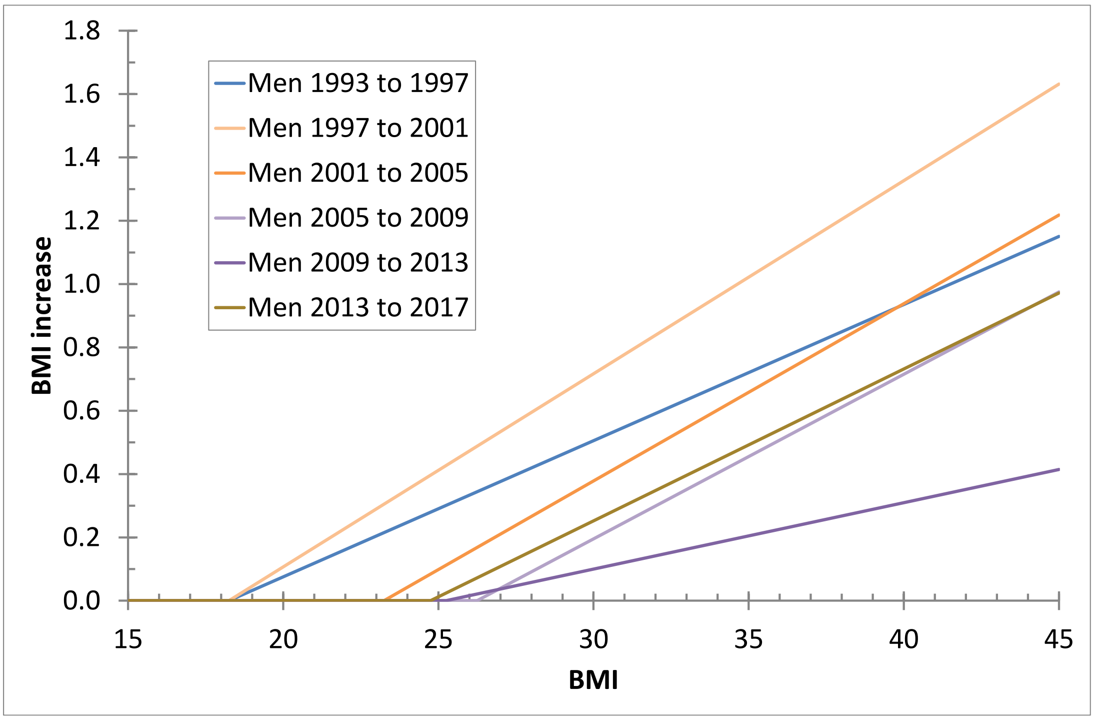


Figure 14.18 BMI increases for the scaling models for men.

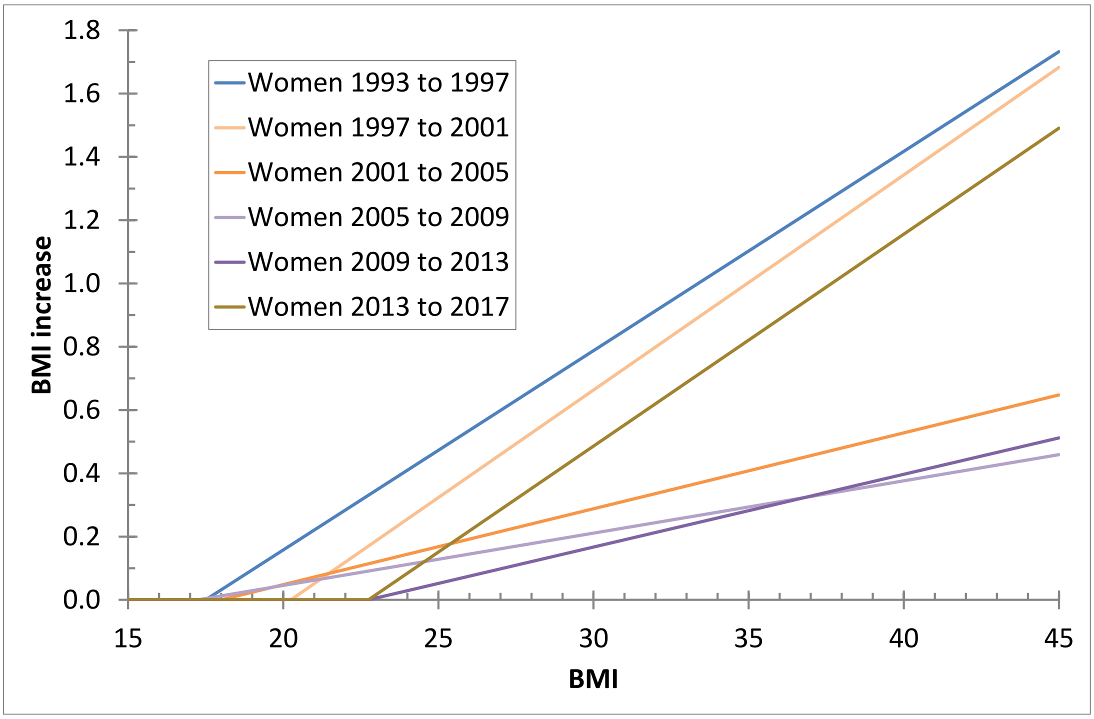


Figure 14.19 BMI increases for the scaling models for women.

14.5.2 Scaling models from 1993 to 2017

As discussed in Section 5.8, scaling models can be applied to the distributions across the whole time period of the HSE. That section considers the 1993 and 2013 distributions, but now that the 2017 distributions are available the changes from 1993 to 2017 can be modelled.

Section 5.8 sets out two main ways of using the scaling model approach. One is a combined model, which aggregates the scaling models derived for the successive distributions. The other is to use a direct scaling model.

The combined model requires taking each of the scaling models given in Table 14.8 and joining them together. As described in Section 5.8.1, this is done by taking the 0.5 BMI interval endpoints for the 1993 distribution and then applying each of the scaling functions in turn to adjust them to 2017 values. The resulting models for men and women are shown in Figure 14.20. The fitness values for the combined models are 22.29×10^{-6} for the model for men and 10.98×10^{-6} for the model for women. The changes are very large from the 1993 to the 2017 distributions and, given this, the models do well in matching the 2017 distributions. The models capture most of the changes from 1993 to 2017.

In some places the combined model curves are not particularly smooth and do not quite follow the shape of the 2017 distribution. In particular, the model for women has a fairly sharp angle at a BMI of about 22.75 due to the scaling models for 2013 and 2017 both having this as their start point. There are some small differences between the model and the actual curves around the peaks of the distribution for both men and women.

The BMI increases for the combined models are shown in Figure 14.21. The percentile increase values (from Figure 14.14) are also shown. As would be expected, the model functions are close to the percentile values. Although the values in Figure 14.21 go up to a BMI value of 45, most values in the 1993 distributions are less than 35 (Figure 14.20) and so these are the most important values for the transformation functions. The 99th percentile increase values (the right end point in the series) differ a little from the functions but these are sensitive to sampling variation and are much less important for the functions in matching the 2017 distributions.

The second main modelling approach is to apply a single direct scaling model (Section 5.8.2). This involves the same methodology used for modelling the changes between successive distributions. The scaling start point and scaling factor are chosen using Excel solver to give the best fitness value against the target distribution. This was done for scaling the 1993 distributions to model the 2017 distributions. The direct scaling model for men has a scaling start point of 20.25, a scaling factor of 1.220, and a fitness value of 54.84×10^{-6}. The direct scaling model for women has a scaling start point of 19.25, a scaling factor of 1.259, and a fitness value of 17.15×10^{-6}. The fitness values are a bit higher than for the combined models, and also a little higher than for the direct scaling models from 1993 to 2013 in Section 5.8.2. However, the models still fit pretty well. They are shown in Appendix 14.V. As was the case for the models for 1993 to 2013, both models have a reasonably high scaling start point and before that point they just use the 1993 values. This creates a noticeable bump in the curves, particularly for the model for men.

As described in Section 5.8.3, other models were also implemented for 1993 to 2013. These are a direct scaling model with a 2 piece linear pattern, and using a binomial probability distribution for the BMI increases with either a linear function for the mean of the binomial or a 2 piece linear function. These were applied for 1993 to 2017 and the results for all the models are given in Appendix 14.V. All the models are able to match the 2017 distributions well. A 2 piece linear pattern inevitably gives a better fitness value than a single linear pattern since it is a more flexible function and includes the single linear pattern within its range of options. The model with the lowest fitness values for both men and women is the 2 piece binomial model.

Appendix 14.V includes a chart of the BMI increases for all the models. They are all similar to each other and close to the pattern of the percentiles, especially up to a BMI of 35 where most 1993 values are. Hence, the main conclusion is that this type of pattern, of an essentially linear pattern of BMI increases, is able to model the changes from 1993 to 2017 extremely well.

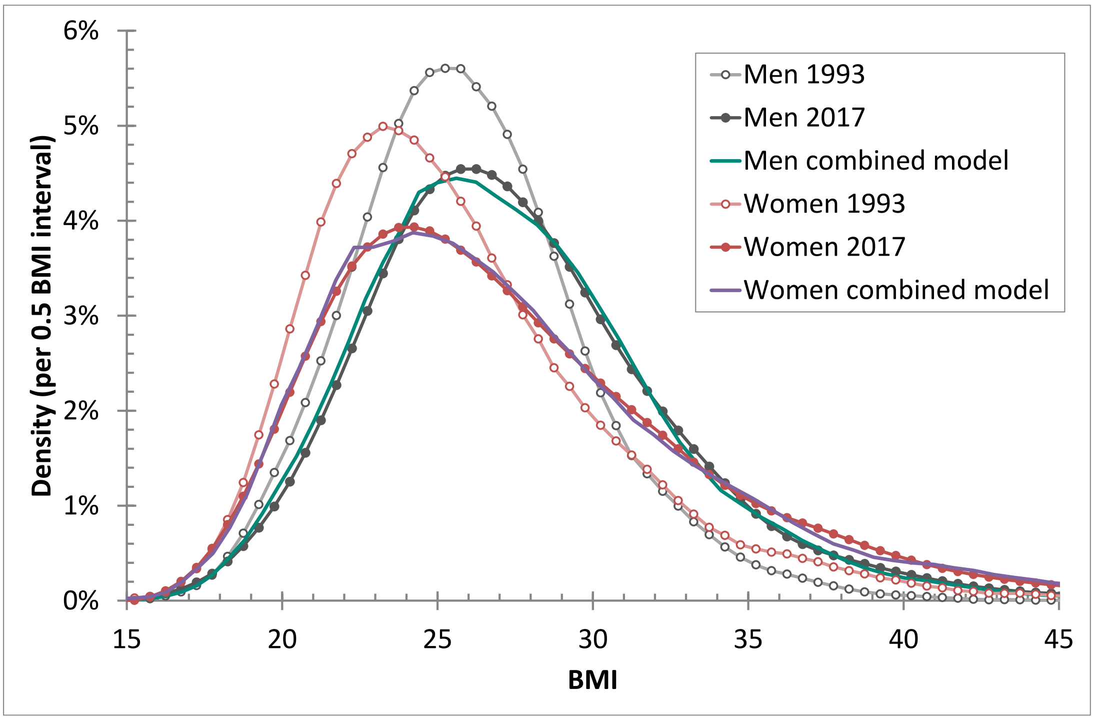


Figure 14.20 Combined scaling models for 1993 to 2017 for men and women.

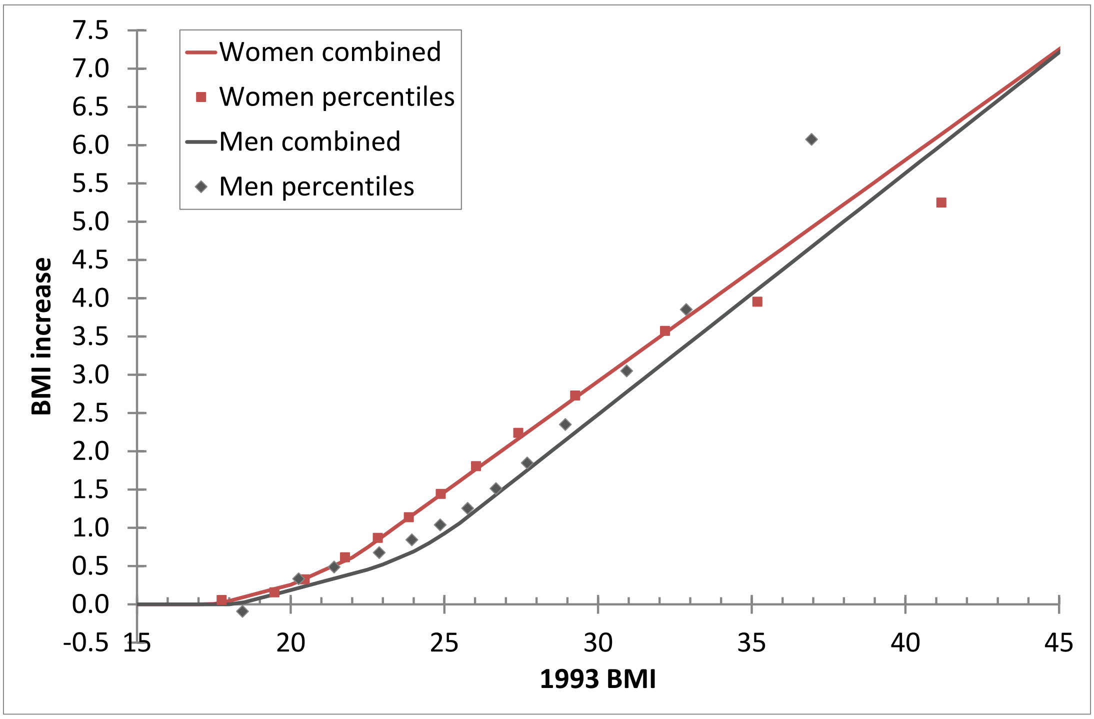


Figure 14.21 BMI increases for combined scaling models for 1993 to 2017 for men and women.

14.6 Interpretation of the 2017 distribution results

The results from the analysis and modelling of the 2017 distributions show that the patterns are similar to those identified from the analysis of previous distributions. The changes from the 2013 distribution to the 2017 distribution and the overall changes from 1993 to 2017 can be modelled well as linear scaling transformations. The changes from 2013 to 2017 are fairly large and there are particularly big increases in high BMI values.

As discussed in Chapter 6, the overall population at different times is assumed to be essentially the same in terms of physiology and genetics. Therefore, changes in the population BMI are considered to be due mainly to the population living in a different time period and therefore experiencing different conditions. A wide range of lifestyle factors that might affect BMI are discussed in Section 6.3.

One implication of the results here, as for the previous distributions, is that the changes in lifestyle factors over time have affected the BMI of those with a higher BMI much more than those with a lower BMI.

14.7 Variation of BMI with age

Chapters 7-10 look at how BMI varies with age. The analysis in Chapter 7 investigates mean BMI and then Chapters 8-10 look at the BMI distributions for different ages.

The results in Chapter 7 for mean BMI are calculated for single years of age. The main analysis in that chapter takes into account birth year by considering birth cohorts, and it is important to do this to avoid the confounding effect of living in different times with different conditions. The results are that BMI tends to increase with age but with the rate of increase getting less with age. Older cohorts are at lower levels, presumably because of the effects of experiencing different conditions through growing up in earlier times. A suggested aging pattern is presented in Chapter 7, with Figures 7.5 and 7.6 showing the cohort values for the whole HSE data for 1991-2014 along with the aging model curve. The aging model represents the hypothesised trajectory of the youngest cohorts in the future if the current conditions stay the same. The indication is also that older cohorts are tending to move towards the aging model because of the effect of the current environment and the current factors that affect obesity.

The HSE data for 2015-2018 does not give age in single years and so the main analysis given in Chapter 7 cannot be extended with the new data. Whilst it would be good to see the patterns in the most recent data this would only give a relatively small amount of extra data. The analysis that can be done is more limited but some work was carried out on the new data and the age effect. Section 14.7.1 considers mean BMI and age for the HSE 2015-2018 data, with some work on the distributions for age groups in Section 14.7.2. This gives some updates on BMI and age for the latest data but the most relevant and detailed analysis on the relationship between age and BMI is still that in Chapters 7-10.

14.7.1 Mean BMI and age

For the HSE 2015-2018 data, the mean BMI was calculated for each of the HSE age groups (18-19, 20-24, 25-29, etc.). For comparison, this was also done for the same age groups for the older data in groups of eight HSE years. All the data used was the actual data with the case weights for the population distributions as described in Chapter 3. As discussed in Section 7.2, the case weights could be adjusted in various ways and the analysis in Chapter 7 adjusted the weights so that the mean case weighting for each year of age within each HSE year equals 1. However, various other weightings were tried for some of the Chapter 7 analysis, including using the Chapter 3 weightings, with very little difference in the results. The same adjustment cannot be made for the 2015-2018 data as the year of age is not given. It was therefore considered easiest just to use the case weights calculated for the population distributions.

As part of the process for the case weights, the population distribution analysis adjusts the case weights of the data for each group of four years to match the population age profile for each year of age (Section 3.3). The 2015-2018 data has wider age groups and so can only be matched to the population age profile for these age groups (Section 14.3.1). The mix of cases within each of the 2015-

2018 HSE age groups is unknown. This means that the proportions for each year of age could be different between the 2015-2018 data and the data for earlier years. However, given the good sampling methodology of the HSE it is probably only slightly different and it is considered that the effect is likely to be very small.

The resulting mean BMI values are shown in Figures 14.22 and 14.23. The mean values are plotted at the mid-value of the integer ages for the age groups, which is consistent with the other charts in the report (age group 18-19 plotted at 18.5, age group 20-24 plotted at 22, etc.).

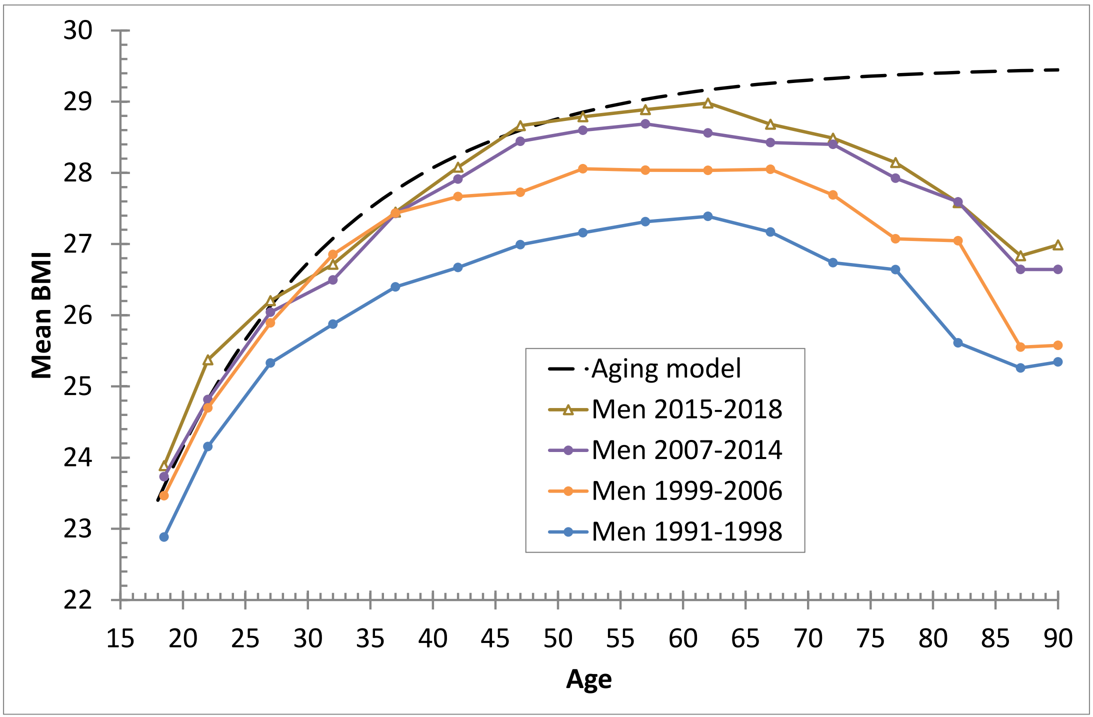


Figure 14.22 Mean BMI for the age group data for men.

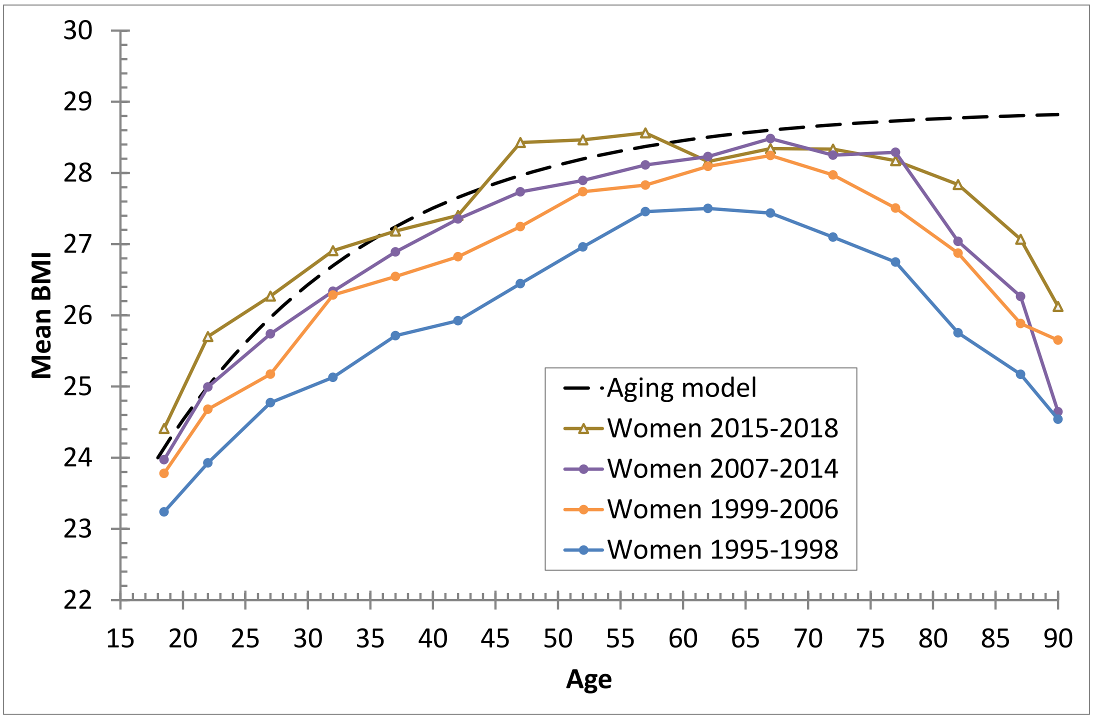


Figure 14.23 Mean BMI for the age group data for women.

It is important to note that the values in Figures 14.22 and 14.23 are simply the relationship between age and mean BMI. This is a combination of the effect of age and also the effect of being born at different times and therefore living through different conditions. The charts do not show the true relationship with age because of the confounding effect of birth year. As discussed at the start of Section 14.7, this is factor is taken account of in Chapter 7, with the cohort analysis, in looking at the relationship with age in narrow birth cohorts. This cannot be done with the new data. Instead, the interest here is just in seeing the changes in the latest data compared to the previous data and also how the latest values compare with the aging model curve particularly for the younger ages.

Figures 14.22 and 14.23 show that for both men and women the mean BMI for 2015-2018 increases fairly consistently across the ages compared to the previous data for 2007-2014. The sample sizes are quite small and for the 2015-2018 data are about 1000 for most age groups, and so there is inevitably some irregularity from sampling variability. From the detailed analysis in Chapter 7, the mean BMI values reduce for older ages due to older cohorts being at lower levels, rather than because BMI actually reduces with age. The cohort analysis in Chapter 7 showed that BMI tends to always increase with age within the cohorts, hence the shape of the aging model.

The differences in 2015-2018 compared to 2007-2014 are plotted in Figure 14.24 for age groups up to 75-79. Age groups above that for 2015-2018 have small sample sizes of less than 500. Age group 18-19 also has small sample sizes being just two age years of data with 236 cases for men and 234 cases for women. One noticeable point is that the increase is high for both men and women for age group 20-24. This is a concern in terms of what the future long term trajectory is for the younger ages and suggests that conditions in the last few years of the HSE data might have led to further increases in BMI for these younger ages.

Looking at Figures 14.22 and 14.23, the 2015-2018 mean BMI data is quite close to the aging model curve for all ages up to the 60-64 age group. However, the values for age group 20-24 are quite a bit above the aging model curve for both men and women. One difference for the data for women compared to men is that quite a few age groups are above the aging model. This may mean that the aging model is a little optimistic and that the long term future BMI may be worse than the model, particularly for women. It will be interesting to compare new data to the model in the future as the data becomes available.

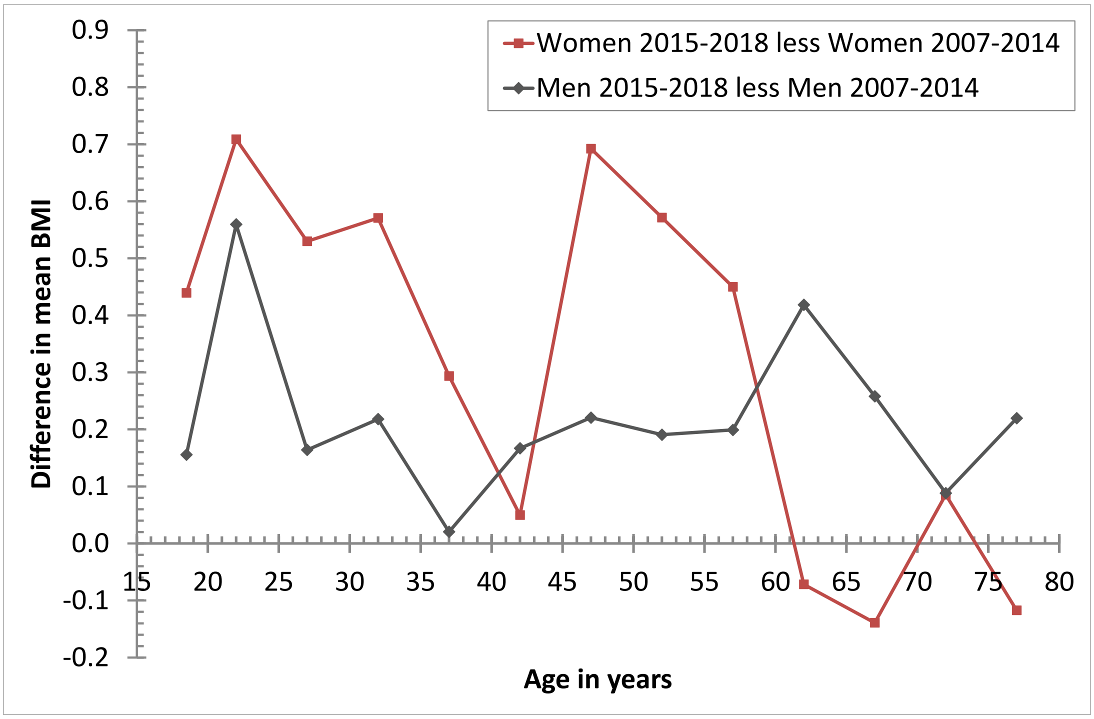


Figure 14.24 Difference in mean BMI of 2015-2018 compared to 2007-2014 for men and women.

14.7.2 Age group distributions

Distributions for each age group for 2015-2018

As already mentioned, the HSE 2015-2018 data does not have age in single years and so distributions for single years of age cannot be produced. In any case, the amount of data would be very small being only four years of data and with each year having relatively small sample sizes. The single year distributions given in Chapters 8-10 for the most recent time period use data for the wide range of HSE years of 1999-2014 to give reasonable sample sizes.

What can be done is to extend some of the age group distribution analysis from Sections 8.2-8.4 and Sections 9.2-9.4. This calculates BMI distributions for age group categories of 18-24, 25-34, 35-44, etc. for 1991-1998, 1999-2006, and 2007-2014. These same age groups can also be used for 2015-2018 by combining the age categories given in the 2015-2018 data (e.g., ages 18-19 and 20-24, ages 25-29 and 30-34).

Given the increases in mean BMI for 2015-2018 that are presented in Section 14.7.1, the main interest is in looking in more detail as to how BMI has increased over time in the age groups, particularly in the 18-24 age group. As discussed in Section 14.7.1, the case weightings used for the 2015-2018 data are those for producing the full population BMI distributions. For consistency, the age group distributions here for 1991-1998, 1999-2006, and 2007-2014 also use the case weightings for the population BMI distributions (explained in Chapter 3). This means that they differ very slightly from the distributions in Chapters 8 and 9 which use adjusted weightings. The age group distributions were produced in the same way as in Chapters 8 and 9 by calculating a distribution for the data in 0.5 BMI intervals for the particular age group and then smoothing this with the 13 point Savitzky-Golay method.

The approach for comparing the age group distributions for the time periods described in Chapters 8 and 9 was extended to include the distributions for 2015-2018. This included comparing the distributions, the descriptive statistics, the category percentages, and the percentiles. Only selected results are presented here.

The distributions for each age group are shown for men in Figure 14.25 and for women in Figure 14.26. With there just being four years of data for 2015-2018 the samples sizes are quite small and so some irregularities are expected in the distributions. However, the general patterns as to how BMI varies visually between the age groups are very similar to the previous time intervals.

Looking at Figure 14.25, the age group distributions for men are all a similar shape. As age increases through the youngest four age groups the main change is for the distribution to shift right along the x-axis, although the peak heights for 35-44 and 45-54 are a bit lower than for the youngest two age groups. From Figure 14.26 there is more change in shape for the distributions for women with age 18-24 having a much higher peak than the other distributions.

Further analysis and discussion of how the age group distributions change with time and with age follows in the rest of this section. Various other results are shown in Appendix 14.VI including descriptive statistics and charts.

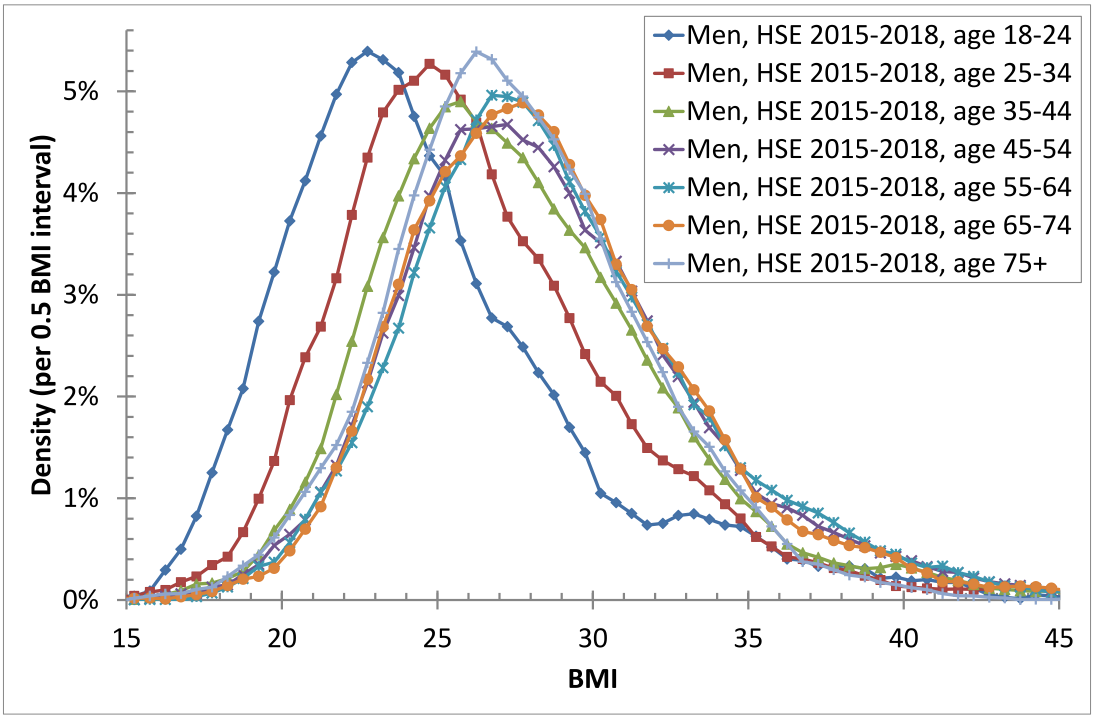


Figure 14.25 Age group BMI distributions for 2015-2018 for men.

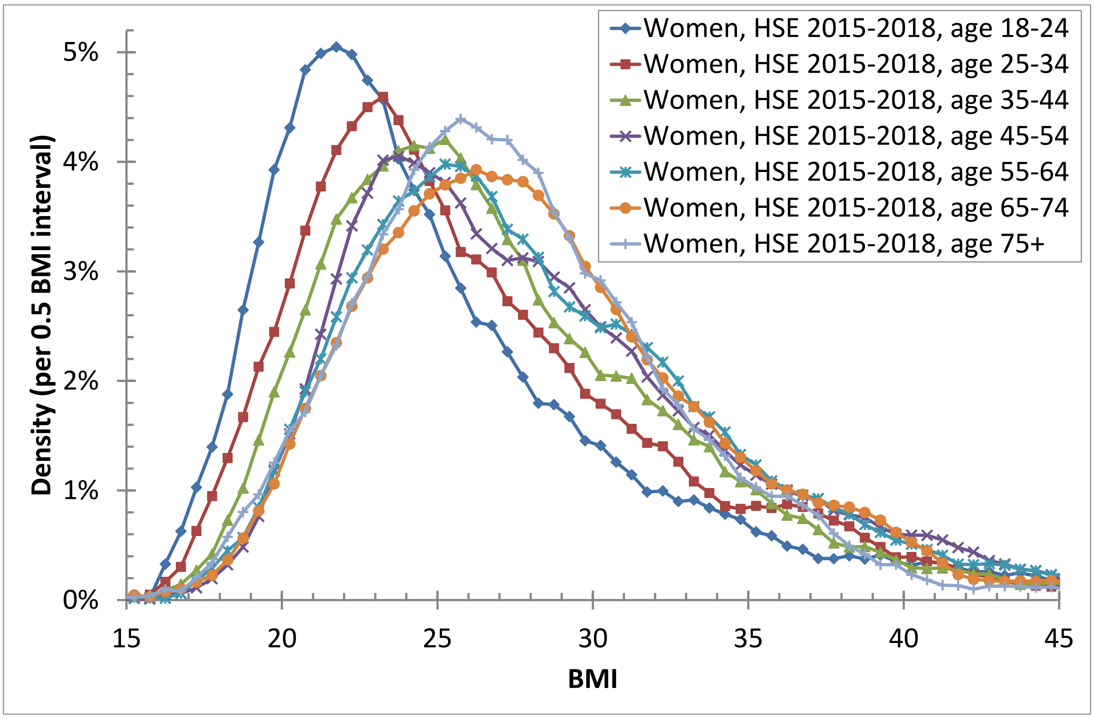


Figure 14.26 Age group BMI distributions for 2015-2018 for women.

Changes in the age group BMI distributions with time

The way that the age group distributions change over time was examined by comparing 2015-2018 with 2007-2014 for each age group. The increases in the mean and median for each age group are shown in Appendix 14.VI in Figure 14.VI.2. Naturally the increases in the mean for the age group distributions show a similar pattern to changes in the mean for the different ages in Figure 14.24, although with wider age ranges for the categories used for the distributions.

For men the highest increase in the mean is for the 18-24 age group with an increase of 0.46. The increases for the other age groups vary between 0.10 and 0.31. The increases in the median are similar for each age group being between 0.06 and 0.18, and are much smaller than the increases in the mean (except for age group 75+). This indicates that, as seen before in the earlier analysis, the increases for 2015-2018 are mainly in the higher percentiles of the data.

For women the increases in the mean values are more variable, but are particularly high for age groups 18-24, 25-34, 45-54, and 75+ being between 0.55 and 0.65. The increases in the median values are also variable but less than the increases in the mean for each age group. For age group 18-24 the increase in the mean of 0.64 is much larger than the increase in the median of 0.20, again indicating that most of the increase is in the higher percentiles.

The pattern of increases was investigated in more detail by plotting the percentile increases (each 10% percentile along with 1%, 5%, 95%, 99%) and the results for men and women are shown in Figures 14.27 and 14.28. These figures also include the percentile increases for the whole population which were obtained by producing population distributions for men and for women for the 2007-2014 HSE data using the usual methodology and comparing the percentile values with those for the 2017 population in Section 14.4. Given the small sample sizes for 2015-2018, linear scaling models were not fitted and just the percentiles used to indicate the nature of the changes over time.

The percentile increases for men in Figure 14.27 show considerable similarity between the age groups. Most age groups are quite close to the percentile increases of the whole population which has a roughly linear increase from a BMI value of about 25. This is approximately the 30th percentile value (this point for whole population is a BMI value of 24.76 and an increase of 0.03). There is very little change for the percentiles below this. Hence, the increase over time for the population has the pattern approximately of a linear scaling from quite a high starting point. The age groups are mostly similar to this. The increases in mean BMI from 2007-2014 to 2015-2018 are therefore mainly due to considerable increases in the higher percentile values. There is some variability between the age groups for men. The most noticeable difference is age group 18-24, where the percentile increases for BMI values above 25 are much higher than for the other age groups. Age 20-24 was identified from the mean values as having the highest increase in BMI (Figure 14.24). Hence, the general pattern for age 18-24 is having greater BMI increases than the other age groups particularly in the higher percentiles.

The percentile increases for the age groups for women are shown in Figure 14.28. There is quite a lot of variability here with some irregularities in the patterns. There is also quite a lot of difference between the age groups. The highest increases for the most common BMI values between 20 and 35 are for the two youngest age groups of 18-24 and 25-34 (apart from a few points on the 75+ age group for low BMI values). Age 35-44 has quite low increases, but the increases are high again for age 45-54. The small sample sizes mean that there will be differences in the patterns due to sampling variation. The patterns mostly look approximately like linear scaling transformations. The two youngest age groups have some increases even at fairly low BMI values and so a scaling model for these age groups would probably have quite a low starting point. This implies some increases even for those in the healthy BMI category along with the general pattern of higher increases for higher BMI values. The 90th percentile values have an increase for age 18-24 of 1.66 from a BMI value of 32.15, and for age 25-34 an increase of 1.77 from 33.71. These are large increases and emphasise the significant changes for these young age groups.

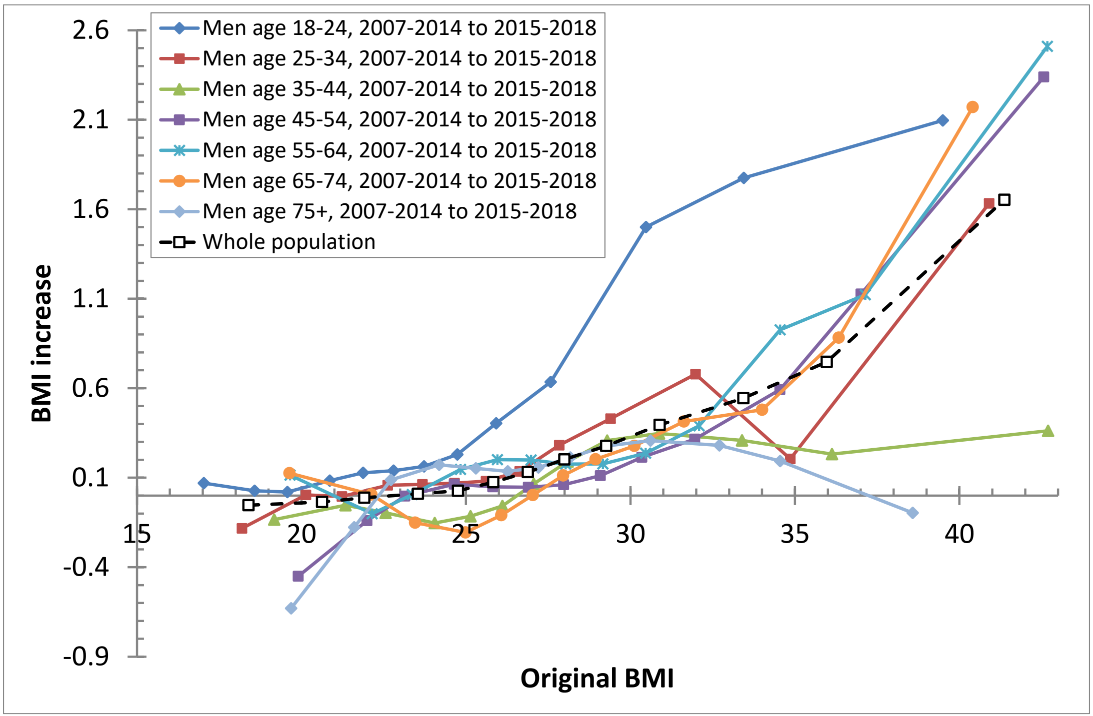


Figure 14.27 Percentile increases for 2015-2018 from 2007-2014 for age groups for men.

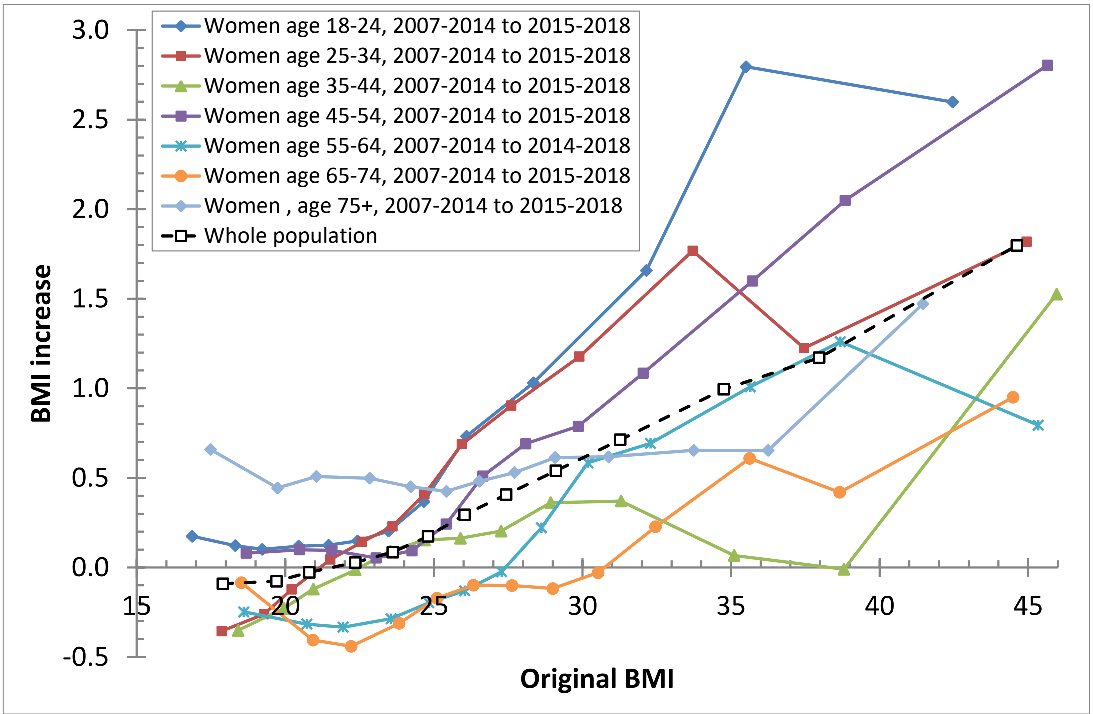


Figure 14.28 Percentile increases for 2015-2018 from 2007-2014 for age groups for women.

As seen before throughout this report, for both men and women this pattern of linear scaling increases will stretch out the distribution by extending the right tail. One noticeable effect is to increase the standard deviation, as can be seen in Appendix 14.VI in Figures 14.VI.3 and 14.VI.4 which plot the standard deviation values for each age group and each time period. The standard deviation values are fairly similar for each age group, although older age groups tend to have a smaller standard deviation. Taking the average of the standard deviation values for age groups up to 55-64, the average values for each of the four HSE time intervals are: men 3.79, 4.31, 4.65, 5.00; women 4.89, 5.42, 5.66, 6.21. So, a considerable increase in 2015-2018, particularly for the data for women. The change in 2015-2018 is also over a shorter time interval with this just being four years of data.

Appendix 14.VI also plots the changes in the category percentages from 2007-2014 to 2015-2018 in Figures 14.VI.7 and 14.VI.9. There is quite a lot of variation between the age groups. However, combining the healthy and overweight categories, and the three obese categories, gives a much more consistent picture (Figures 14.VI.8 and 14.VI.10) with total reductions in healthy and overweight for all age groups for men of between 1.4% and 2.5%, and for most age groups for women between 2.3% and 3.5%. For women, age group 18-24 has the largest change with a greater reduction than this of 5.0%.

Overall, the changes in the age group distributions from 2007-2014 to 2015-2018 are consistent with the changes over time for the other populations and age group distributions presented in this report with patterns similar to a linear scaling, albeit with some irregularities from the small sample sizes.

Changes in the age group BMI distributions with age

The way that BMI changes with age can be seen most clearly and in most detail with the single year of age BMI distributions in Sections 8.5 and 9.5, and in Chapter 10. This cannot be done for the HSE data for 2015-2018 as age is given only in wider categories. However, the differences in the age group distributions can be examined and compared with the results in Chapters 8 and 9.

The patterns of the age group BMI distributions for 2015-2018 in Figures 14.25 and 14.26 can be compared to the distributions for 2007-2014 in Figures 8.3 and 9.3. The percentile changes were also calculated from one age group to the next and the results are shown in Appendix 14.VI in Figures 14.VI.11 and 14.VI.12.

For men, the patterns of the way the age groups change as age increases are very similar to that for 2007-2014 (Section 8.4). This is not surprising given the similarity between the age groups of the changes from 2007-2014 to 2015-2018 in the percentiles in Figure 14.27. The age group distributions all have a similar shape and a fairly similar peak height. The main increases are between the three youngest age groups. The percentile increases (Figure 14.VI.11) are a roughly horizontal shape and are very much similar to those for 2007-2014 in Figure 8.10. This all indicates that, as for 2007-2014, the changes in BMI between the age groups are mostly a shift (or translation) of the distribution along the x-axis.

For women, the patterns are again similar to 2007-2014. The distribution with the highest peak is age 18-24 with the peak reducing for 25-34 and then again for 35-44. The left tail does shift along the x-axis for these three distributions but the distributions also seem to get more stretched out and so a bit longer in the right tail. The percentile increases (Figure 14.VI.12) have more irregularities than for 2007-2014 presumably due to the smaller sample sizes. The patterns are fairly horizontal but with trends of the BMI increases getting larger as BMI increases for age group 18-24 from percentile 1% to 70%, for age group 25-34 from percentile 1% to 40%, and for age group 35-44 from percentile 30% to 90%. Hence, as for 2007-2014, the main changes with age for women appear to be a mixture of a shift along the x-axis and a linear scaling change in shape (Section 9.4).

Overall, the general way that BMI changes between the age groups is consistent with that for 2007-2014 in Chapters 8 and 9. However, the analysis that can be done for 2015-2018 is limited by the wider age categories in the data. The single year of age distributions in Section 8.5, Section 9.5, and Chapter 10, that are based on one year of age and a wide range of HSE years, therefore give a better and more detailed picture of how the BMI distributions change as age increases.

14.8 Future scenarios: comparison with the current trend and low BMI scenarios

14.8.1 Nature of the current trend scenario

The current trend scenario is set out in Chapter 12. This applies the general scenario modelling approach in Chapter 11 where distributions based on the lognormal are generated for each year of age and are combined into a whole population distribution. The main parameters required for each year of age are the minimum, mean, and standard deviation.

The current trend scenario in Chapter 12 aims to model a situation where the current conditions continue, with "current conditions" meaning the conditions in the last few years of the HSE data used here. Hence, this assumes that BMI follows the general trend and patterns identified in the recent data. It uses the aging model curves from Chapter 7 for the mean values (based on data for 1991-2014). The standard deviation values are based on the data for 2011-2014 and the minimum values on the age group distributions for 2007-2014. Hence it is based on the most recent data up to 2014.

The aging model curves in Chapter 7 were chosen to generally follow the actual mean values for the youngest cohorts and then the shape of the values for the older cohorts. The curves are mostly above the actual values for the older ages. The hypothesis with the aging model curve is that that the youngest ages will follow this curve as they get older if recent conditions stay the same. The older ages will also tend to move towards it. With the current trend scenario being based on the aging model curve the hypothesis is that the population BMI distributions will move towards the current trend scenario distributions over time unless conditions change.

The 2017 distributions were therefore compared with the current trend model to see how close they have got to it. The results are given in Section 14.8.2 for men and in Section 14.8.3 for women.

14.8.2 Current trend scenario and the 2017 distribution for men

The BMI distribution for men for 2017 is compared with the current trend model in Figure 14.29, and this also shows the 2013 distribution. The changes in the BMI categories for the current trend model compared to the 2017 and 2013 distributions are shown in Figure 14.30. These are updated versions of Figures 12.3 and 12.4 in Chapter 12 that compare the current trend model with the 2013 distribution.

The previous analysis in this chapter shows an increase in BMI from 2013 to 2017. As a result, the 2017 distribution has got closer than the 2013 distribution to the current trend model. For example, the mean values for 2013 and 2017 are 27.42 and 27.63 compared to 27.87 for the current trend model. From Figure 14.29 the peak height of the 2017 distribution is nearer than 2013 to the height of the current trend model, and it is also closer in parts of the right tail.

The effects on the category percentages are shown in Figure 14.30. The difference in the healthy percentage of the model compared to the population is smaller for 2017 than for 2013 with the current trend model being 2.4% lower (28.3% for the model compared to 30.7% for 2017) rather than 3.0% lower (with 2013 being 31.3%). Similarly, the increases in obese I and obese II for the current trend model are lower compared to 2017 than compared to 2013.

The mean BMI for the age groups in Section 14.7.1 shows that the 2015-2018 values are higher than for 2007-2014 and generally closer to the aging model (Figure 14.22). Hence, the distribution has moved closer to the current trend model, although the healthy percentage for the 2017 distribution is still quite a bit higher than for the model.

One noticeable point is that the severely obese percentage for 2017 is higher than for the current trend model by 0.2% and so at the end of the right tail the 2017 distribution has actually gone slightly beyond the current trend model. This reflects the increases in prevalence of high BMI values seen in 2017 (Section 14.4.4). As discussed in Section 14.7.2, one result of this is an increase in the standard deviation for the age groups. The current trend model uses a standard deviation of 4.7 for all ages based on 2011-2014 data. However, the standard deviation for the age groups in the 2015-2018 data is about 5.0 on average for ages up to 64 (Section 14.7.2). The current trend model could be updated by increasing the standard deviation and this is done in Section 14.8.4.

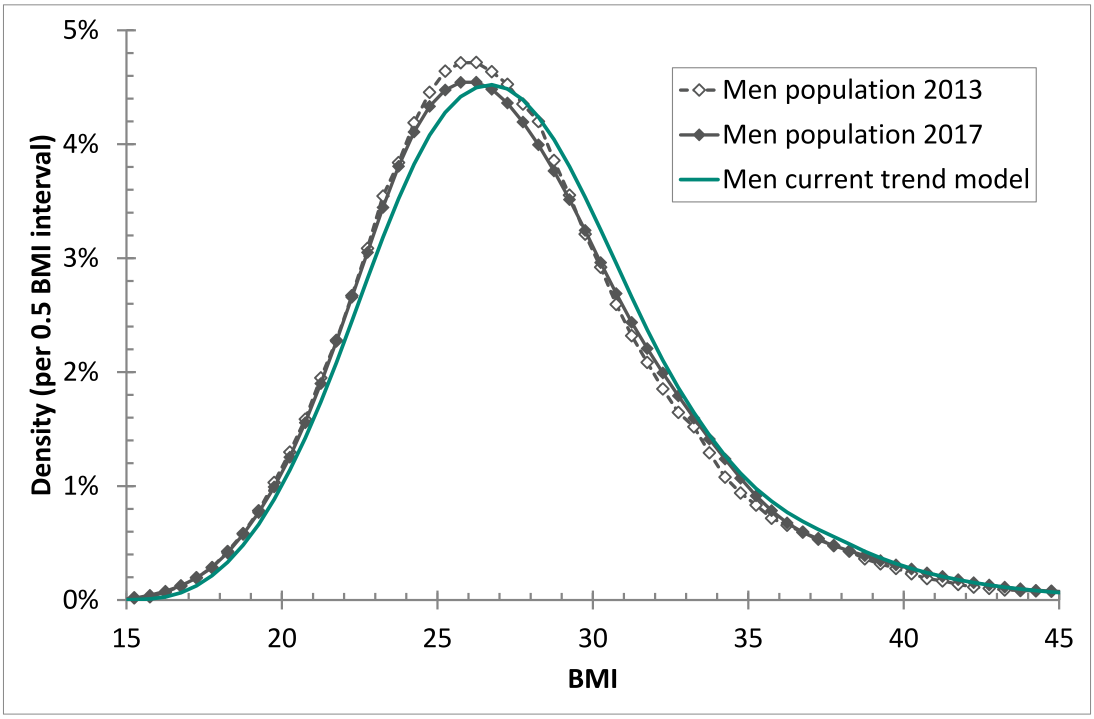


Figure 14.29 BMI distributions for the 2017 population and the current trend scenario for men.

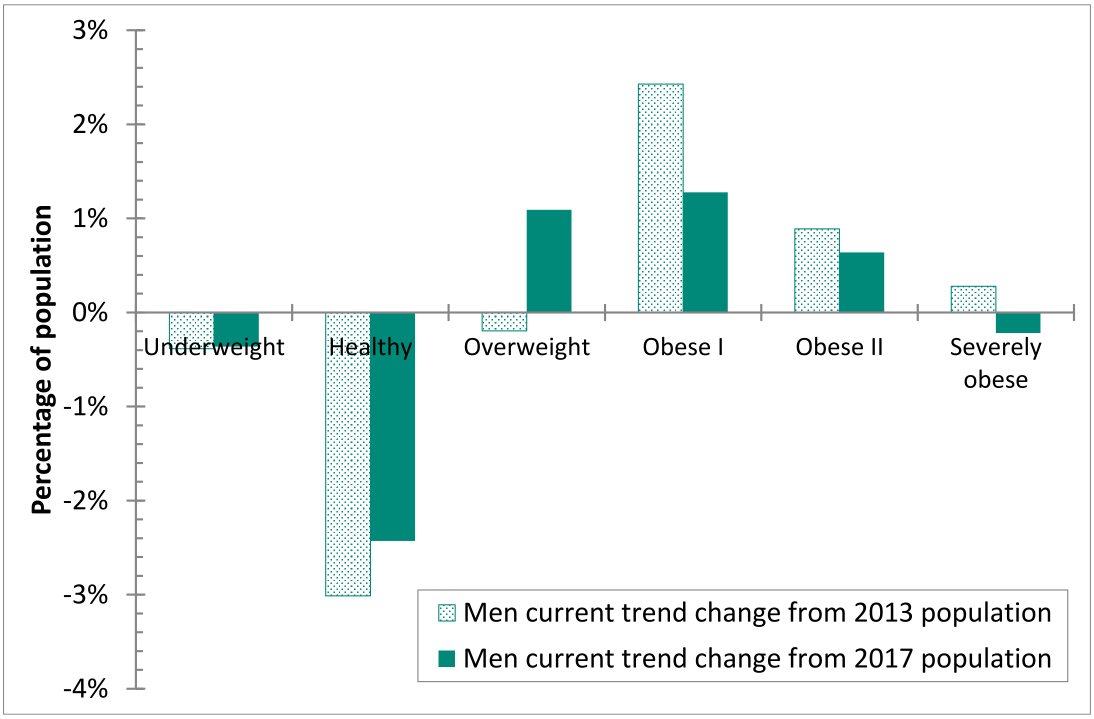


Figure 14.30 Change in category percentages for current trend scenario from 2017 for men.

14.8.3 Current trend scenario and the 2017 distribution for women

The comparison of the current trend model for women with the 2017 and 2013 populations is in Figure 14.31, with the changes in the category percentages shown in Figure 14.32.

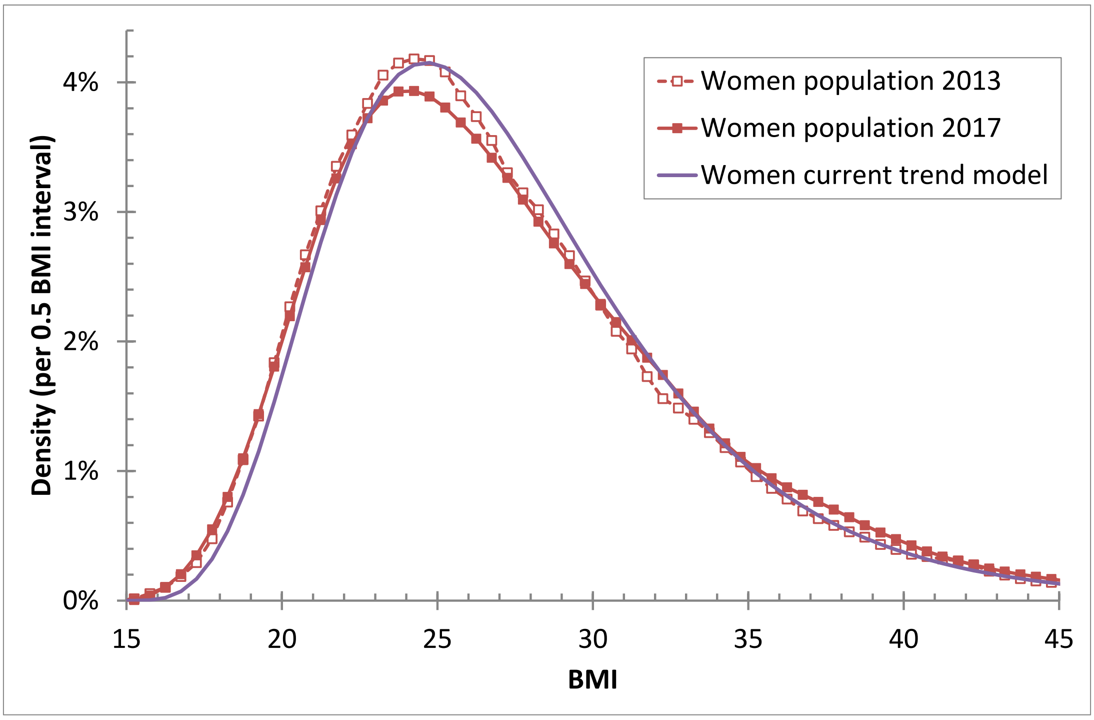


Figure 14.31 BMI distributions for the 2017 population and the current trend scenario for women.

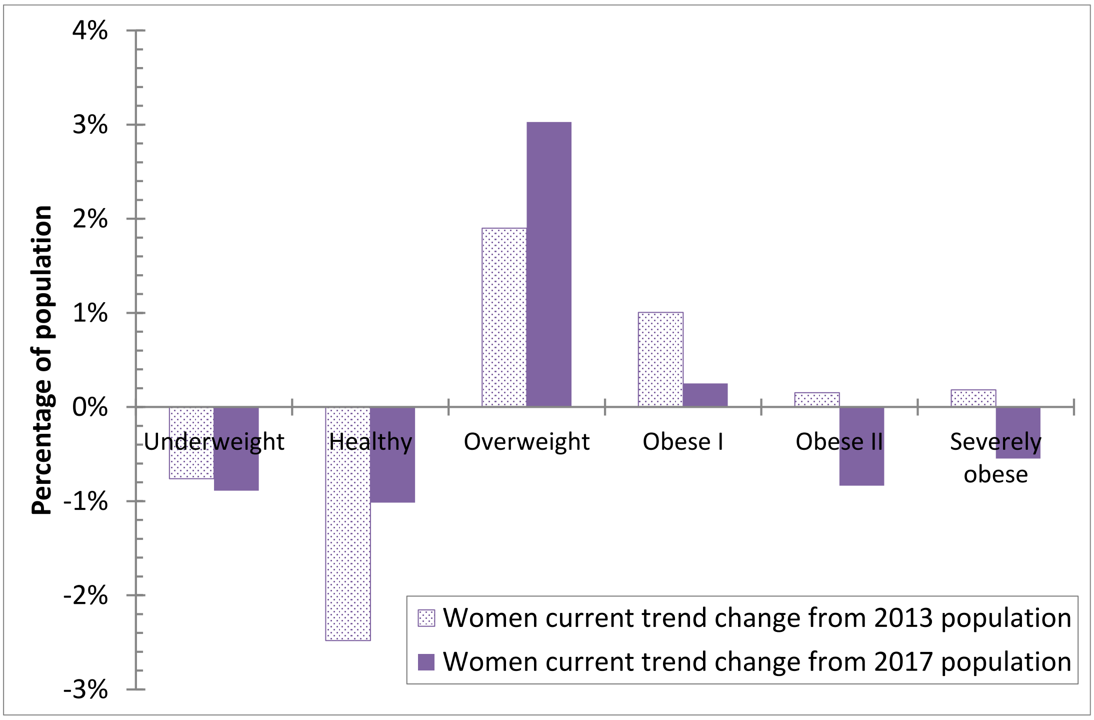


Figure 14.32 Change in category percentages for current trend scenario from 2017 for women.

As discussed earlier in this chapter, there is a considerable increase in BMI from 2013 to 2017 for the distribution for women. As a result, the 2017 distribution has got much closer than the 2013 distribution to the current trend model. The mean for the current trend model is 27.50 compared to 27.47 for 2017 and 27.13 for 2013. Therefore, the mean for 2017 is almost at the level of the current trend model.

From Figure 14.31, the shape of the 2017 distribution is different to the current trend model distribution with a lower peak and a longer right tail. It is also higher at the start of the left tail. The result, as shown in Figure 14.32, is that the current trend model has considerably more in the overweight category and slightly more in obese I, but less in the other categories. The healthy category percentage is only 1.0% lower for the model (37.15% for the model compared to 38.16% for 2017).

For high BMI values, the 2017 distribution has higher prevalence than the current trend model in obese II and severely obese (7.3% and 4.1% for 2017 compared to 6.5% and 3.6% for the model). This is similar to what has happened for men in the way that the 2017 population exceeds the model in the right tail, but has happened even more so in the data for women.

From Figure 14.23 the mean values for several of the age groups are higher than the aging model values and so this is part of the reason why the population distribution has gone past the current trend model in the prevalence of high BMI values. However, there has also been a considerable increase in high BMI values in the population which has increased the standard deviation for the age groups. As noted in Section 14.7.2, the average standard deviation for age groups up to age 64 is 6.2 whereas the standard deviation used in the current trend model based on the 2011-2014 data is only 5.7. As for the model for men, the current trend model could be updated by increasing the standard deviation and this is done in Section 14.8.4.

14.8.4 Revised current trend scenario

As discussed in Sections 14.8.2 and 14.8.3, the 2017 distributions exceed the current trend models for high BMI values at the end of the right tails. The prevalence of high BMI values has increased considerably from 2013 to 2017. This is one effect of BMI increasing as a linear scaling pattern. One result of this is to increase the standard deviation of the BMI distributions for the population as a whole (Tables 14.2 and 14.3) and for the age groups. The current trend model can be updated with standard deviation values that reflect the latest actual values.

Therefore, revised current trend models were generated with larger standard deviations based on the average values in Section 14.7.2 from the age group distributions up to ages 55-64 using the 2015-2018 data. These are standard deviation values of 5.0 for men and 6.2 for women. The resulting distributions are shown in Figure 14.33.

In Figure 14.33, for men, the revised current trend model has a similar left tail, a lower peak, and a higher right tail compared to the 2017 distribution. For women, the model distribution and the 2017 distribution are very close to each other. The very slight differences are the model being a bit higher at the very start of the left tail and in the right tail just after the peak, and then a bit lower in the middle of the right tail.

The effect of the larger standard deviation on the model distribution is to lower the peak of the distribution and extend both tails. The mean of the distribution stays the same because the same values are used for the mean values for each age. Therefore, the difference between the mean values of the models and the 2017 distributions remains the same.

The category percentages for the revised current trend model for men (with the 2017 values in brackets) are: underweight 1.0% (1.1%), healthy 29.5% (30.7%), overweight 40.4% (41.1%), obese I 20.2% (19.4%), obese II 6.4% (5.5%), severely obese 2.5% (2.2%). The category percentages for the revised current trend model for women (with the 2017 values in brackets) are: underweight 1.7% (2.0%), healthy 38.6% (38.2%), overweight 32.2% (31.6%), obese I 16.3% (16.8%), obese II 6.8% (7.3%), severely obese 4.5% (4.1%).

Extending the right tail of the distribution increases the prevalence of high BMI values. This results in higher percentages for obese II and severely obese compared to the original current trend models for both men and women. As discussed above, the 2017 distributions have gone beyond the original current trend models in the right tails with higher percentages for severely obese for both men and women, and for obese II for women. In the revised model, the percentage for severely obese is higher for the models than for 2017 for both men and women (although the obese II percentage is still lower than 2017 for the distribution for women). This aspect is more realistic than the original model for a scenario of continuing current trends.

Extending the left tail of the distribution increases the prevalence of low BMI values and has the effect for this scenario of increasing the healthy category percentage. This reduces the amount by which the healthy category percentage is less for the model than for 2017, and in fact the percentage is slightly higher than 2017 for the distribution for women.

One use of the current trend model, as discussed in Chapter 12, is as a benchmark or comparison to assess future data. Making the comparison with the 2017 populations has highlighted some of the changes that have happened from 2013 to 2017. These include increases in the prevalence of high BMI values, giving more variability in the distributions and larger standard deviation values.

By updating the standard deviation, the revised scenario is a more up-to-date estimate of the BMI distributions under the scenario assumptions of conditions staying as they were in the recent HSE years considered here. Other updates or changes could also be applied. For example, using even higher standard deviation values to reflect BMI keeping increasing for each age as a linear scaling. There could also be higher mean values, particularly for the model for women since the 2017 population mean value is already at about the same level as the model.

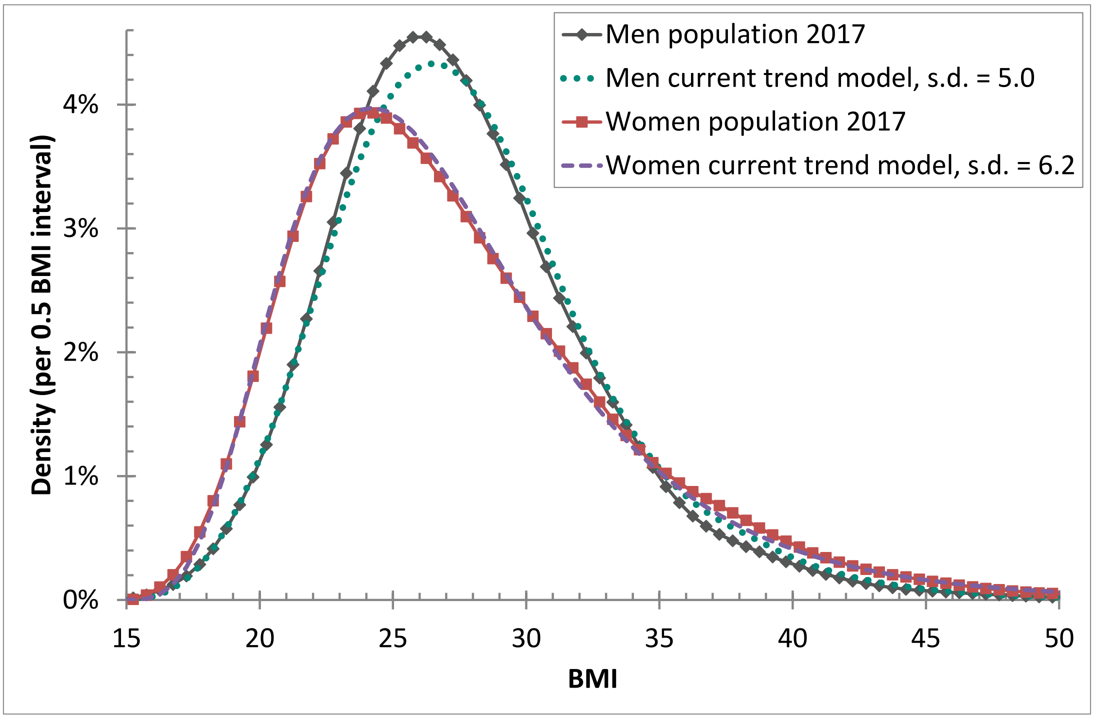


Figure 14.33 BMI distributions for the 2017 populations and revised current trend scenarios [Revised scenarios have larger standard deviation values].

14.8.5 Low BMI scenario

The low BMI scenario in Sections 12.1.2, 12.3, and 12.5 uses a much lower BMI aging profile to give distributions that are similar to 1993. It therefore represents one way of getting back to the BMI levels of the early 1990s. The parameters and age distributions could be used as targets to aims for.

The low BMI scenario is not generated from current data and so is not altered by the new HSE data available for 2015-2018. Therefore, the low BMI distributions stay the same as presented in Sections 12.3 and 12.5, with charts of the distributions given in Figures 12.6 and 12.15. The only update to be made in the results is in the comparison between the low BMI distributions and the latest BMI distributions. In Chapter 12, the comparison was with the 2013 distributions but the differences can be updated using the 2017 distributions. As discussed in Chapter 12, the low BMI distributions are very similar to the 1993 distributions and so the differences between the distributions for low BMI and 2013 are consequently similar to the differences between the distributions for 1993 and 2013.

The low BMI scenario statistics can be compared to those of the 2017 distributions to update the results given in Sections 12.3 and 12.5. For men, the low BMI distribution mean is 1.49 lower than for 2017 and the category differences compared to 2017 are: underweight -0.1%, healthy 9.1%, overweight 3.4%, obese I -7.1%, obese II -3.3%, severely obese -1.9%. For women, the low BMI distribution mean is 1.43 lower than for 2017 and the category differences compared to 2017 are: underweight -0.9%, healthy 9.7%, overweight 1.5%, obese I -4.4%, obese II -3.5%, severely obese -2.4%. Naturally, these statistics values are very similar to the differences between the 1993 and 2017 distributions given in Table 14.4 in Section 14.4.5.

If it is possible for the population to follow the low BMI scenario in the future then the improvement in BMI would be considerable. For example, the healthy category percentage would be 9.1% higher for men and 9.7% higher for women compared to the 2017 distributions, and there would be considerable reductions in each of the obese category percentages.

14.9 Calories

Chapter 13 sets out how to convert BMI changes into calories using energy balance equations from the literature. In that chapter calories are calculated for the changes in BMI from 1993 to 2013 and from age 18 to 24. The values from 1993 to 2013 can be extended to 2017 using the results set out in this chapter. As was done for the work presented in Chapter 13, this involves taking the BMI increases for the combined scaling models and multiplying the increases by the calorie values derived from the Henry (2005) equations. There are three levels for the physical activity level multiplier (PAL). For the median PAL, the calorie values per unit BMI are 56.91 kcal / day for men and 34.99 kcal / day for women (Section 13.5.1). The combined scaling model values are therefore just multiplied by these numbers. Similarly for the other PAL levels using the values in Table 13.4 in Chapter 13.

The calorie values for the BMI increases from 1993 to 2017 using this approach are in Figure 14.34 (which updates Figure 13.1 in Chapter 13). As discussed in Chapter 13, these are estimated equilibrium amounts at the population level. They are not applicable for individuals as values may vary considerably at an individual level. As with any modelling there are some uncertainties, and there are also different models, as discussed in Sections 13.6 and 13.7.

With the increases in BMI from 2013 to 2017 the calorie amounts are obviously slightly higher than in Figure 13.1, particularly for high BMI values. As noted in Section 13.5.1, most people have a BMI of less than 35 in 1993 and the calorie amounts for this range of BMI values are quite small. For example, for a BMI of 30 the calorie increase for the median PAL is 141 kcal / day for men and 102 kcal / day for women. As discussed in Section 13.5.1, these calorie increases are similar to the increase in daily calories for the U.K. data for retail supply in Section 6.3.3 (data given on Roser, Ritchie, and Rosado, 2013 using source data from the United Nations Food and Agricultural Organization). The average calories for this data increases by 170 kcal / day from 1985-1996 to 1997-2019.

The chart shows the increases since 1993, but this also gives an indication of the reduction in average calories at a population level that might be needed to get back to 1993 BMI levels. These are equilibrium values and so are long term differences, rather than short term changes.

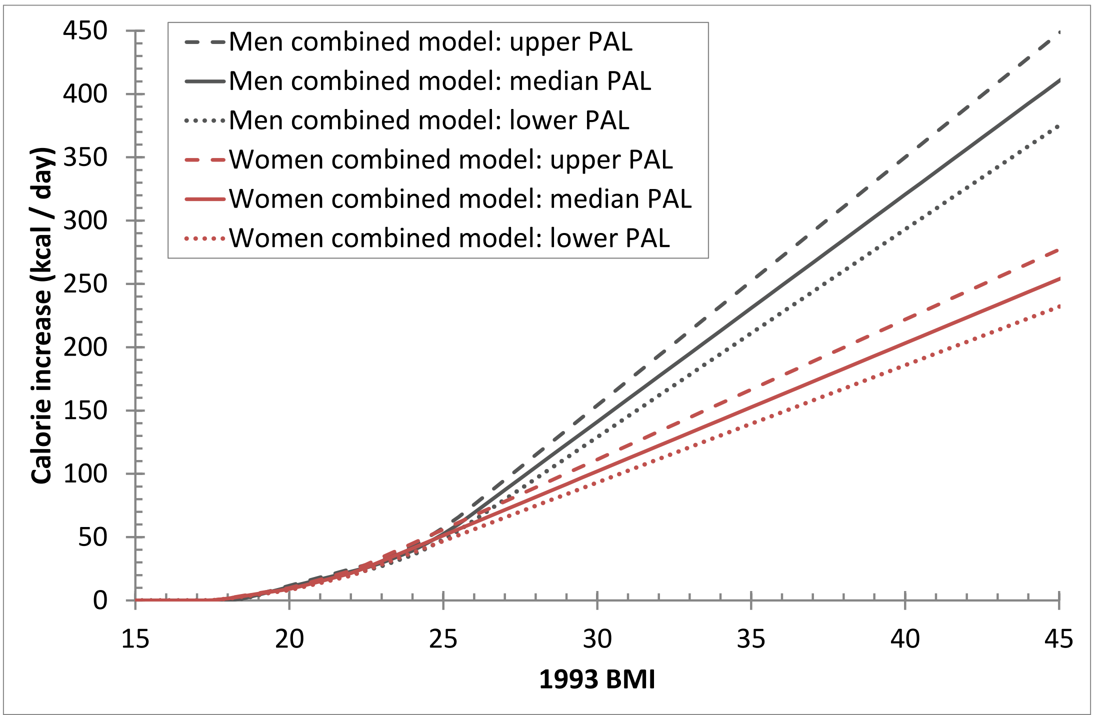


Figure 14.34 Calorie values for the BMI increases from 1993 to 2017.

14.10 Summary

This chapter presents the results from the latest HSE data for 2015-2018 that has enabled the 2017 population BMI distributions to be produced. This has allowed the results from the previous chapters in this report to be updated.

The distributions and the pattern of the results are similar to those from the earlier chapters. BMI has increased from 2013 to 2017 in approximately a linear scaling pattern with higher BMI values tending to increase more than lower BMI values. The increase in BMI from 2013 to 2017 is greater than from 2009 to 2013 and is higher for women than for men. The increases in the age groups over time are also approximately a linear scaling pattern although with some irregularities with the small sample sizes for the data for 2015-2018. There is a relatively high increase in BMI for the youngest age groups which is a concern as to the future BMI trajectory for these age groups.

Overall, the results and the general patterns are consistent with those from the earlier data and so the main findings and conclusions from the previous chapters are still applicable for the newer data.

Chapter 15. Examples of other variables: deprivation and height

Key points from this chapter:

- This chapter looks at how BMI varies for different categories of deprivation and for different categories of height. The reason for this analysis is partly due to interest in the relationship of BMI with these variables and partly illustrative to show how the methodology in this report can be applied to look at other variables.
- The same approach as in the rest of the report was applied to produce estimated population BMI distributions for each category of the variables using four years of HSE data. The smoothing version used was the double smoothing. This works well in smoothing the data even with the fairly small sample sizes here.
- For deprivation, the HSE data since 2001 categorises cases by 5 levels of deprivation. Distributions were produced for each deprivation category for 2003 (using 2001-2004 data) and for 2017 (using 2015-2018 data).
- For 2017, BMI is higher for higher levels of deprivation (i.e., living in more deprived areas) with a larger mean and a greater prevalence of obese values of BMI. The BMI differences between the deprivation categories are much greater for women than for men. Comparing the distributions for the most deprived and least deprived categories for 2017, for men the mean BMI is 0.88 higher and the healthy category percentage 3.7% lower, for women the mean BMI is 2.42 higher and the healthy percentage 15.2% lower. These differences for women are very large and, by comparison, are greater than the differences over time for the population between 1993 and 2017 in Chapter 14.
- For 2003, there is not much difference between the distributions for men for the deprivation categories. The mean values are all about the same. The distribution for the most deprived category actually has the highest healthy category percentage, although also the highest severely obese category percentage. For women, there are substantial differences in the distributions with the most deprived category having a mean BMI that is 1.35 higher than for the least deprived category and a healthy category percentage that is 10.2% lower. These are substantial differences but not as great as for 2017.
- As found elsewhere in the report, the changes in the percentile values over time from 2003 to 2017 for each of the deprivation categories are approximately a linear pattern with higher increases for higher values of BMI. This indicates that the changes in the distributions over time are approximately a linear scaling transformation. The increases tend to be greater for the higher levels of deprivation and hence the differences between the deprivation categories have increased over time.
- For height, the cases were split into three categories and distributions produced for 1993 and 2017. BMI tends to be slightly lower as height increases, with the differences being greater for women than for men. The changes over time from 1993 to 2017 for each category are again approximately a linear scaling pattern.

15.1 Relationship with other variables

This report has mainly looked at how BMI varies over time and with different ages. There are many other variables that might be associated with differences in BMI. The same methods and analyses applied in this report can be used to investigate other variables. This chapter shows two examples by considering deprivation and height.

Deprivation is a factor that is of general interest in obesity research and is sometimes reported in statistics, such as in the HSE data tables spreadsheet (Section 1.2.5). In some years of the HSE data, each case has a variable for a deprivation category and so BMI can be analysed by these categories and distributions produced and compared. The analysis on this is presented in Section 15.2.

Height is part of the BMI equation. The idea of including it in the equation is to take into account the relationship between height and weight so that BMI reflects body type in terms of whether someone has a relatively low or high weight given their height. It is therefore interesting to see whether the BMI distribution varies by different categories of height. It is, however, difficult to draw clear inferences from this as any differences between the height categories could be due to genuine relationships between height and body type, or could be due to some limitations in how BMI corrects for height. The analysis on this is in Section 15.3.

Part of the interest in the work in this chapter is to look at how BMI is associated with each of these two variables. In addition, the chapter is also illustrative to show how the general approach in this report can be applied to look at any variables of interest, given suitable data.

15.2 Deprivation

Section 15.2 sets out the analysis on the deprivation categories. The HSE data has a variable for deprivation with five categories. This is explained in Section 15.2.1, along with some of the statistics that are given in the data tables published with the HSE.

Distributions were produced for each deprivation category with the method used throughout this report, and the approach taken is set out in Section 5.2.2.

The HSE data for 2015-2018 was used to generate distributions for each deprivation category for 2017 and the results are in Section 15.2.3. The deprivation variable is included in the HSE from 2001 onwards and distributions were also produced for 2003 using the first four years of data for 2001-2004. The results for 2003 are in Section 15.2.4. Then Section 15.2.5 analyses how BMI has changed over time from 2003 to 2017 for the different deprivation categories by looking at the percentile values.

Section 5.2.6 examines the magnitude of the differences between the distributions for the deprivation categories by comparison with the changes in the whole population over the timescale of the HSE.

15.2.1 Deprivation variable and HSE statistics

The HSE data has a variable for deprivation that gives a quintile value from 1 to 5 for each case, using an "index of multiple deprivation" (IMD). The variable is called QIMD. The documentation that is included in the download with the HSE 2018 data (HSE 2018 Dataset Documentation Variable List) explains on page 19 that a deprivation value is produced as "a composite index of relative deprivation at small area level, based on seven domains of deprivation: income; employment; health deprivation and disability; education, skills and training; barriers to housing and services; crime and disorder; and living environment." The index values are calculated for "Super Output Areas" and are divided into five quintiles numbered from 1 (least deprived) to 5 (most deprived). Each case in the HSE is assigned the deprivation value for the "Super Output Area" that they live in using their postcode. The five deprivation categories are labelled from least to most deprived as D1 to D5 in this chapter.

The data tables spreadsheet of statistics provided with the HSE and discussed in Section 1.2.5 includes some statistics of BMI by deprivation. This is in Table 5 of the data tables spreadsheet for the HSE 2018 [41]. This gives the mean and category percentages for men and for women for each of the five IMD quintiles. The values are for ages 16 and over and the spreadsheet says that the data has been "age-standardised". These statistics indicate that BMI tends to increase with increasing deprivation. The prevalence values given for all the BMI obese categories (BMI ≥ 30 for ages 16+, age-standardised) for the five deprivation categories from least deprived to most deprived are men: 20.4%, 24.9%, 25.0%, 25.5%, 34.6%; women: 20.6%, 26.4%, 30.9%, 32.2%, 36.5%. Hence, there is a big difference in obesity prevalence across the deprivation categories. The Public Health England (PHE) 2020 obesity slide set [42] (Section 1.2.5) includes a chart of these obesity prevalence values on slide 15.

The mean values in the data tables spreadsheet for 2018 for the five deprivation categories are men: 26.7, 27.5, 27.3, 27.4, 28.2; women: 26.2, 27.2, 27.9, 27.8, 28.6. Again, there are big differences between D1 and D5, particularly for the data for women.

The sample sizes are quite small for single years when dividing the data up into the deprivation categories. For example, for the 2018 HSE data tables the number of cases (adults age 16+) in each deprivation category are between 553 and 668 for men and between 588 and 722 for women. This means that there is some uncertainty and variability from year to year. For example, the data tables for the 2019 HSE (Table 6 of the 2019 HSE statistics spreadsheet [43]). have the obesity prevalence values as men: 21.9%, 26.3%, 27.3%, 28.6%, 30.2%; women 22.4%, 24.9%, 30.5%, 29.2%, 39.5%. Therefore, there are quite big differences with the 2018 values, such as D5 being 4.4% lower for men and 3.0% higher for women. This is likely to be due mainly to sampling variation. Larger samples can obviously be obtained by combining years together, and this is done in the analysis set out in this chapter.

The same methodology used before in this report can be applied to give more detailed information than the BMI category percentages by deriving probability distributions. Exactly the same general process as earlier in this report was applied. As for the full populations in Chapters 3 and 4, and in Chapter 14, four years of data were used. The deprivation variable only seems to be available in the HSE data from 2001 onwards and so only data since then can be considered. The latest four years of data considered in this report for 2015-2018 were used to produce distributions for 2017. The earliest four years of data that has the deprivation variable for 2001-2004 were used to produce distributions for 2003. More details of the specific process used are in Section 15.2.2.

[41] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2018/health-survey-for-england-2018-data-tables Excel file link: Health Survey for England, 2018: Overweight and obesity in adults and children data tables

[42] https://www.gov.uk/government/publications/adult-obesity-patterns-and-trends web page link for "PHE obesity adult slide set: England 2020"

[43] https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2019/health-survey-for-england-2019-data-tables Excel file link: Health Survey for England, 2019: Overweight and obesity in adults and children data tables

15.2.2 Method used for the deprivation BMI distributions

As explained in the previous section, BMI distributions by deprivation category were produced for 2017 using the 2015-2018 data, and for 2003 using the 2001-2004 data. The same methodology as in Chapters 3, 4, and 14 was applied. For each deprivation category, valid data for each year was extracted, the case weights adjusted to have a mean of 1, four years of data combined, case weights adjusted to match the standard 2011-2014 population age profile, data combined into 0.5 BMI intervals to give a data distribution, and the data distribution smoothed using the Savitzky-Golay method.

Age profile and adjusting for age

One aspect of the deprivation data is that the age profile is quite different for the different deprivation categories. For example, for the data for men, comparing the D1 (least deprived) and D5 (most deprived) quintiles there is a large difference. Using the HSE case weights for the combined 2015-2018 data (with each year adjusted to have a mean of 1), D1 has lower percentages than D5 for each age category up to age 45-49, and higher percentages for each age category 50-54 and over (except for 90+ being very slightly less). In other words, D1 has more people aged 50 and over, and D5 has more people aged 18-49. The total percentage values for age 18-49 and age 50+ are: D1 46.6%, 53.4%; D5: 65.2%, 34.8%. Therefore, the difference is substantial being 18.6%. There are particularly big differences for age categories 25-29 and 30-34 with a combined total of 10.7% higher percentage for D5 compared to D1. The overall population percentages for 2011-2014 used in this report (Section 3.3) are between the values of D1 and D5 for most age categories with age 18-49 56.8%, age 50+ 43.2%. Hence D1 (least deprived) has more older people than the population age profile and D5 (most deprived) has more younger people. The other three deprivation categories are between D1 and D5 in sequence with the percentages for age 18-49 for all the deprivation categories being: D1 46.6%, D2 50.0%, D3 53.6%, D4 59.0%, D5 65.2%.

The pattern is similar for the data for women. The slight difference is that the percentages are higher for D5 up to only age 40-44. The values for age 18-44 and 45+ are D1: 34.7%, 65.3%; D5: 55.1%, 44.9%; 2011-2014 population: 44.9%, 55.1%. The difference between D1 and D5 is 20.5%. The percentages for age 18-44 for all the categories are: D1 34.7%, D2 38.1%, D3 43.9%, D4 49.7%, D5 55.1%.

As shown in earlier chapters BMI tends to increase with age and so it is important to adjust for age. As in the previous analysis in this report, the data was adjusted to a standard age profile of the 2011-2014 population (Section 3.3). This is to remove the confounding effect of age and therefore to assess better the difference due to deprivation only. The data for 2015-2018 has wide age categories, mostly of five years, and so these had to be used for the age adjustment, as for the 2017 population (Section 14.3.1). The data for 2001-2004 has age in single years and so the data age profile could be matched to the standard 2011-2014 age profile by single years. The adjustment for age does mean that the actual BMI distributions and statistics in the deprivation areas will be a bit different as these will reflect the actual age profiles in the deprivation areas.

Smoothing version

The different Savitzky-Golay smoothing approaches of 11 point, 13 point, and double smoothing were considered for producing the distributions. The double smoothing approach smooths the data first with the 11 point smoothing, and then applies further smoothing to these values using the 13 point smoothing. These were compared in detail for the 2017 distributions and are discussed in Chapter 14 in Sections 14.3.2 and 14.3.3, and the results for 2017 are very similar for the different options. The double smoothing was chosen for 2017 in Chapter 14 to give a slightly smoother distribution. This was considered more plausible and realistic for 2017 given that these are distributions of a simple equation of physical quantities for a large population.

The spreadsheets for the BMI distributions for the deprivation categories produce each of the three smoothing versions, as is also the case for the population distribution spreadsheets. The three versions give curves that are visually quite similar. The sample sizes here are small since the data is split up into the five deprivation categories. As a result, there are greater irregularities in the curves than for the whole population distributions. These are most pronounced for the 11 point and 13 point smoothing. The double smoothed distributions are noticeably quite a bit smoother both around the peak and in the tails. The statistics for the three versions are very similar, with the mean, standard deviation, and skew being identical because the smoothing method maintains these values as the same as the data distribution. The BMI category percentages are often the same to 1 decimal place and differ by at most 0.2% between the three versions.

As for the 2017 population distributions in Chapter 14, it was decided to use the double smoothing version for all the deprivation category BMI distributions. This was preferred as it gives a smoother distribution, which is considered more plausible for BMI for a population. The double smoothing method works well in producing a smooth distribution that follows the general shape of the data, even with the small sample sizes of this data. It may therefore also be a very good option for any future analysis.

15.2.3 Results by deprivation category for 2017

The BMI distributions for men and for women by deprivation category for 2017 are shown in Figures 15.1 and 15.2. The distribution for the whole population for 2017 from Chapter 14 is also shown in each chart as the dashed line. The statistics are in Tables 15.1 and 15.2, and a chart of the mean and median values is given in Figure 15.I.1 in Appendix 15.I.

The distributions and the statistics for both men and women show BMI increasing for the higher deprivation categories. The statistics are generally consistent with and similar to the values quoted above in Section 15.2.1 for 2018 and 2019 from the HSE statistics spreadsheets for the percentages for all obese. The data here is for larger samples being four years, and is for ages 18+ rather than 16+. The results for 2017 are discussed here. Further comments on the magnitude of the changes between the categories for both 2003 and 2017 are in Section 15.2.6.

Results for men for 2017

In the chart of the distributions for men in Figure 15.1 the peak of the distributions gets lower with each deprivation category and both tails tend to get a bit higher. There is not much difference in the x-axis BMI value of the position of the peaks. Around the peak of the distributions D1 and D2 look quite similar but then the peaks are quite a bit lower for D3, D4, and D5. In the tails of the distribution, D5 is visually higher than the others at the ends of both tails.

Comparing the distributions for D5 and D1, the peak of D5 is a lot lower than for D1 and both tails are higher. The D5 and D1 distributions cross at BMI values of 20 and 29, with D5 having 1.0% higher frequency for BMI < 20, 9.0% lower for 20 ≤ BMI < 29, and 8.0% higher for BMI ≥ 29. Hence, the D5 category has a slightly higher prevalence than D1 of underweight BMI values and BMI values at the low end of the healthy range, as well as a higher prevalence of high BMI values.

From the statistics in Table 15.1, the mean increases slightly from one category to the next from D1 to D5. Comparing D5 and D1, the mean for D5 is 0.88 higher than for D1. The difference in the healthy category percentages is fairly small with D5 at 28.7% being 3.5% lower than D1 at 32.2%. The other three deprivation categories have healthy percentage values between the values for D1 and D5. The D5 distribution also has a lower overweight category percentage than D1 but higher percentages for each of the obese categories by a total of 7.6%. For the very high BMI values, all deprivation categories have some prevalence of severely obese BMI values, but D5 is quite a bit higher than the others with 3.6% compared to between 1.4% and 2.2% for D1-D4.

Results for women for 2017

The distributions for women in Figure 15.2 have considerable differences. The start of the left tails are similar to each other but with each successive deprivation quintile the peak gets lower and the distribution gets much more stretched out to the right with a higher right tail. There do seem to be some irregularities of bumps and troughs in the tails with the fairly small sample sizes.

Comparing the distributions for D5 and D1, D5 is slightly above D1 up to a BMI of 17.5, is higher for BMI above 27.5, and is lower in between. For these intervals, D5 has 0.3% higher frequency for BMI < 17.5, 17.7% lower for 17.5 ≤ BMI < 27.5, and 17.4% higher for BMI ≥ 27.5. These differences are large and are much greater than the corresponding differences for the distributions for men.

From the statistics in Table 15.2, the mean for D5 is 2.42 higher than for D1 and the healthy category percentage is 15.2% lower. The total for the obese categories is 15.6% higher. Again, these are very large differences and are much greater than the differences between D1 and D5 for men.

The severely obese category percentage increases with deprivation quintile from 2.3% for D1 to 6.7% for D5. The value for D5 is a lot higher than even the value for D4 of 4.3%. As can be seen from Figure 15.2, the tails of the distributions are still reasonably high at a BMI of 45. The percentages for the D1-D5 categories for BMI ≥ 45 are 0.7%, 1.5%, 1.5%, 1.3%, 2.1%, and so again the D5 percentage is quite a bit higher than the others.

Effect of the age profile adjustment

As mentioned at the start of this section the age profile varies a lot between the sample data for the different deprivation quintiles. The distributions weight the data using a standard age profile and this is so that the distributions produced are an estimate of the effect of the deprivation category whilst controlling for age. Actual BMI distributions and category percentages in the different deprivation areas will reflect the actual age profile. This was investigated by calculating some statistics for the data using the HSE case weights (standardised to a mean weight of 1 for each year) but without adjusting for the age profile, and for the data weighting each case equally. The differences are fairly small and some results are given in the next two paragraphs.

For men using the HSE case weights, the means and total obese percentages are D1: 27.49, 25.0%; D2: 27.64, 25.9%; D3: 27.62, 27.0%; D4: 27.67, 27.5%; D5: 28.00, 30.3%. Compared to the results in Table 15.1, BMI is slightly higher for D1 and is slightly lower for D5 and so the differences between them are a bit smaller. This is expected given the age profiles compared to the standard age profile used, with D1 having more in the older age groups which tend to have higher BMI and D5 having more in the younger age groups which tend to have lower BMI. However, there is still a noticeable difference from D1 to D5 with, for these statistics, the mean BMI 0.51 higher and 5.3% more in all the obese categories. Using equal weights is probably a less suitable approach since the HSE case weights are designed to take into account non-responses (Section 2.1.1). The statistics for equal weights give higher BMI values for all quintiles for these statistics but with a similar difference between D1 and D5 (e.g., 5.4% more in the obese categories).

For women using the HSE case weights, the means and total obese percentages are D1: 26.58, 22.2%; D2: 27.16, 25.5%; D3: 27.59, 29.0%; D4: 28.13, 31.3%; D5: 28.66, 36.4%. This is similar to the situation for the data for men with BMI slightly higher for D1 and lower for D5 compared to the statistics in Table 15.2. The difference of D5 compared to D1 is 2.08 for the mean and 14.2% for the all obese percentage. So, slightly less than in Table 15.2 but still very substantial. The calculations for equal weights again give higher BMI values for these statistics with a difference of 14.9% between D1 and D5 for the all obese percentage.

Hence, based on these results, the higher deprivation quintile areas will have considerably higher actual BMI levels even with the effect of the lower age profile. The differences in the actual distributions will probably be slightly less than for the distributions shown in Figures 15.1 and 15.2 which are adjusted to a standard age profile.

Variation of BMI with age

Calculations were done of the mean BMI values for each of the HSE data age categories (18-19, 20-24, 25-29, etc.). Charts of the results are given and discussed in Appendix 15.I (Figures 15.I.2 and 15.I.3). The results follow the patterns found for the distributions with the mean values for each successive deprivation category generally being at a higher level, and the differences for the deprivation categories for women being greater than for men. However, the sample sizes are very small and so there is considerable uncertainty with the values. The differences between the mean values for D5 and D1 are fairly consistent across the ages for both men and women, albeit with considerable fluctuations arising from the small sample sizes (Figure 15.I.4 in Appendix 15.I).

Further analysis could be done as future work to produce distributions for each of the deprivation categories for age groups or for single years of age. It would probably be best to use a wider range of HSE data to give larger sample sizes. The distributions for single years of age in Chapters 8-10 use the HSE data for 1999-2014. The data for 2001-2014 could be used for the deprivation categories, although the sample sizes would still be small when dividing the data between the five categories. Alternatively, age group distributions could be produced, as in Chapters 7, 8, and 14 for age groups such as 18-24, 25-34, 35-44, etc. Again, a wide range of HSE data could be used for this.

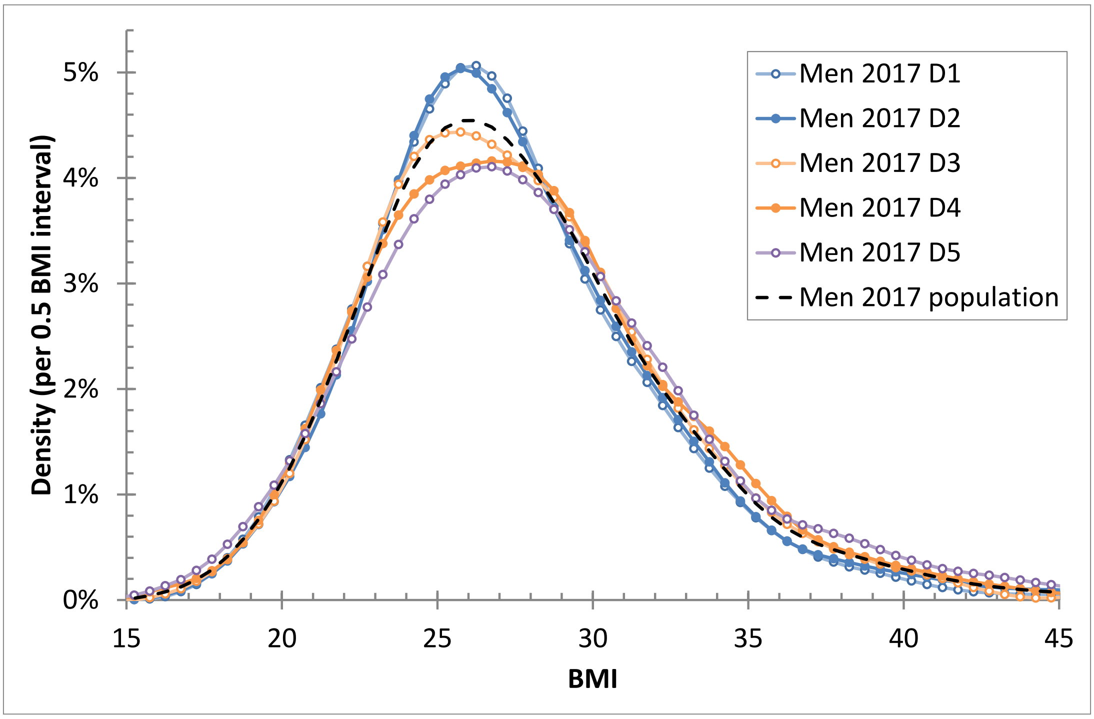


Figure 15.1 BMI distributions for the deprivation quintiles for men for 2017.

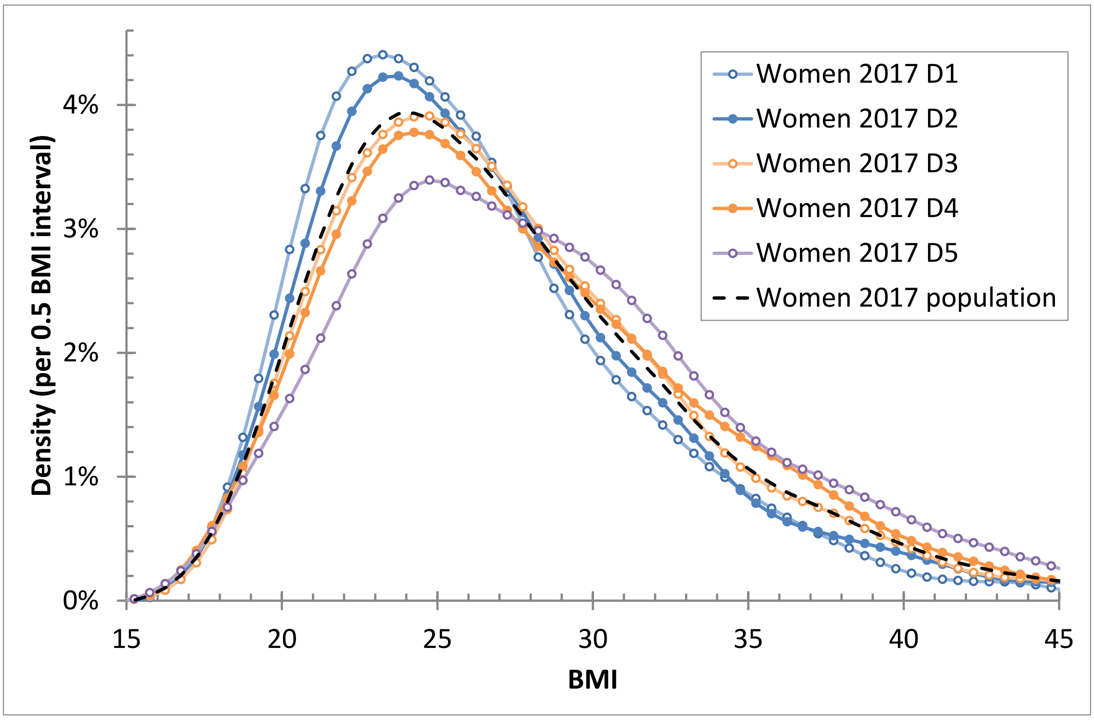


Figure 15.2 BMI distributions for the deprivation quintiles for women for 2017.

Table 15.1 Descriptive statistics for the 2017 distributions for deprivation quintiles for men.

	D1	D2	D3	D4	D5	D5-D1
Number of cases	2352	2458	2452	2205	2178	
Mean	27.23	27.50	27.56	27.77	28.10	0.88
Median	26.68	26.83	27.04	27.25	27.43	0.74
Modal interval mid-point	26.25	25.75	25.75	26.75	26.75	0.50
Standard deviation	4.68	4.88	4.77	5.08	5.65	0.97
Skewness	1.01	1.07	0.69	0.76	1.06	0.05
Underweight (BMI < 18.5)	0.9%	0.9%	1.0%	1.2%	1.7%	0.7%
Healthy (18.5 ≤ BMI < 25.0)	32.2%	30.9%	31.1%	30.2%	28.7%	-3.5%
Overweight (25.0 ≤ BMI < 30.0)	43.4%	43.1%	40.7%	39.7%	38.6%	-4.8%
Obese I (30.0 ≤ BMI < 35.0)	17.7%	18.4%	20.0%	20.5%	20.8%	3.1%
Obese II (35.0 ≤ BMI < 40.0)	4.3%	4.5%	5.6%	6.1%	6.6%	2.3%
Severely obese (BMI ≥ 40.0)	1.4%	2.2%	1.5%	2.2%	3.6%	2.2%
All obese (BMI ≥ 30.0)	23.4%	25.1%	27.2%	28.8%	31.0%	7.6%

Table 15.2 Descriptive statistics for the 2017 distributions for deprivation quintiles for women.

	D1	D2	D3	D4	D5	D5-D1
Number of cases	2945	2947	2876	2772	2757	
Mean	26.36	26.97	27.54	27.82	28.78	2.42
Median	25.32	25.80	26.45	26.71	27.74	2.42
Modal interval mid-point	23.25	23.75	24.75	24.25	24.75	1.50
Standard deviation	5.50	6.02	6.06	6.20	6.70	1.20
Skewness	1.20	1.39	1.08	0.94	0.91	-0.28
Underweight (BMI < 18.5)	2.1%	2.0%	1.8%	2.2%	2.1%	0.1%
Healthy (18.5 ≤ BMI < 25.0)	45.3%	41.8%	37.2%	35.7%	30.1%	-15.2%
Overweight (25.0 ≤ BMI < 30.0)	31.3%	31.8%	32.3%	30.9%	30.8%	-0.5%
Obese I (30.0 ≤ BMI < 35.0)	13.8%	15.1%	17.3%	18.1%	20.4%	6.6%
Obese II (35.0 ≤ BMI < 40.0)	5.2%	5.6%	7.2%	8.9%	9.8%	4.6%
Severely obese (BMI ≥ 40.0)	2.3%	3.8%	4.0%	4.3%	6.7%	4.3%
All obese (BMI ≥ 30.0)	21.3%	24.4%	28.6%	31.3%	36.9%	15.6%

15.2.4 Results by deprivation category for 2003

The previous section gave the results for the deprivation categories for 2017 using the most recent data. BMI distributions can also be produced for earlier points in time. Then, the changes in the BMI distributions over time can be examined and compared. The deprivation variable only seems to be available in the HSE data from 2001 onwards and so only data since then can be considered. BMI distributions were produced for 2003 using the first four years available of 2001-2004. The results are described in this section. Section 15.2.5 then looks at the changes from the 2003 to 2017 distributions using percentiles.

The approach for producing the distributions was the same as described in the Section 15.2.2. One slight difference for this data compared to that for 2015-2018 is that the age for each case is given in single years and so the age profile adjustment for single years was used (as in Chapters 3 and 4). The HSE years for 2001 and 2002 do not have the interview case weight and so each case in these years just has a case weight of 1 (prior to the age adjustment). The combination of years is different to the four year groupings of earlier chapters for the population distributions (which used 1999-2002 and 2003-2006). The estimated deprivation category population distributions here for the 2001-2004 data are referred to as the 2003 populations consistent with the labelling used before in this report. The distributions for men and for women are shown in Figures 15.3 and 15.4, with the statistics in Tables 15.3 and 15.4.

The population distributions for 2003 are also included as the dashed lines in Figures 15.3 and 15.4. These were produced in the usual way from the 2001-2004 HSE data, with the smoothing option chosen being the Savitzky-Golay double smoothing. The three smoothing options are discussed for the 2017 data in Section 14.3. As for 2017, there is very little difference between the results for the three smoothing options for 2003 but the double smoothing gives a slightly smoother distribution.

Results for men for 2003

For men, the distributions in Figure 15.3 each have the peak at a similar BMI value but the higher deprivation quintiles have lower peaks and slightly wider tails. It is noticeable that the tail for D5 is higher than D1 and the other deprivation categories in the end of the left tail as well as the right tail.

For men there is actually a negligible difference in the mean and median between D1 and D5 with differences of just 0.02 and -0.08. The peak is also at about the same BMI value. The main difference in the distributions is that D5 has a lower peak and higher tails for both left and right tails, and hence a higher standard deviation. The curves for D1 and D5 cross at BMI values of 23.0 and 30.5, with D5 having 3.9% higher frequency for BMI < 23.0, 6.6% lower for 23.0 ≤ BMI < 30.5, and 2.6% higher for BMI ≥ 30.5 (almost the same category as all obese).

Hence, D5 has a higher percentage than D1 for the underweight category. Also, the result of the greater prevalence up to a BMI of 23.0 is that D5 actually has a higher percentage than D1 for the healthy BMI category as well. This is different to the situation for 2017. Indeed, from Table 15.3, D1 (least deprived) has the lowest healthy category percentage and D5 (most deprived) the highest healthy category percentage, although the differences are quite small being 2.6%.

However, D5 has a lower percentage than D1 for the overweight category and then higher percentages for each of the obese categories. Again, the differences are small with the all obese percentage for D5 being just 2.6% higher than for D1 (22.7% compared to 20.1%). D5 does have almost double the severely obese percentage compared to D1, although the values are small (1.1% compared to 0.6%).

The charts and statistics all indicate the much greater difference in the level of obesity between D5 and D1 in 2017 compared to 2003. For example, the mean is only 0.02 higher in 2003 but 0.88 higher in 2017. The total obese percentage is 2.6% higher in 2003 but 7.6% higher in 2017. The severely obese percentage is 0.6% higher in 2003 compared to 2.2% higher in 2017.

Results for women for 2003

For women, the distributions for D2, D3, and D4 for 2003 in Figure 15.4 are close together but D1 and D5 are quite different. As for 2017 all the distributions have a similar end of the left tail. Going from D1 to D2 / D3 / D4 to D5, the rest of the distribution gets more stretched out to the right with the peak of the distribution getting lower and further to the right, and the right tail mostly being higher and more stretched out to the right.

The curves for D1 and D5 cross at BMI values of 18.5 and 27.0. Category D5 has 0.4% higher frequency than D1 for BMI < 18.5, then 11.3% lower for 18.5 ≤ BMI < 27.0, and 10.9% higher for BMI ≥ 27.0. These are similar crossing points to 2017 (BMI values of 17.5 and 27.5) but 2017 has greater differences including D5 being 17.4% higher for BMI ≥ 27.5 (Section 15.2.3).

For the statistics in Table 15.4 the values for D2 and D3 are very similar and D4 has only slightly higher BMI levels with a mean that is slightly higher and healthy category percentage a bit lower. Comparing D5 and D1, the mean BMI of D5 is 1.35 higher, the healthy category percentage is 10.2% lower and there are 9.3% more in the all obese category. The severely obese percentage is 2.5% higher at 3.7% compared to 1.3%.

These are much greater differences between D5 and D1 than for men in 2003 where, as explained above, there is a negligible difference in the mean and quite small differences in the category percentages (with D5 actually having a slightly higher healthy category percentage, although also a slightly higher all obese percentage).

The differences for women between D5 and D1 in 2003 are substantial but not as great as for women in 2017 (Section 15.2.3). For example, the difference in the mean for 2017 is 2.42, the healthy percentage is 15.2% lower, the all obese percentage 15.6% higher, and the severely obese percentage 4.3% higher.

The next section looks at the changes from 2003 to 2017 for each of the deprivation categories, which gives more information on the changes over time and hence the different levels of difference between D5 and D1 in 2003 and 2017.

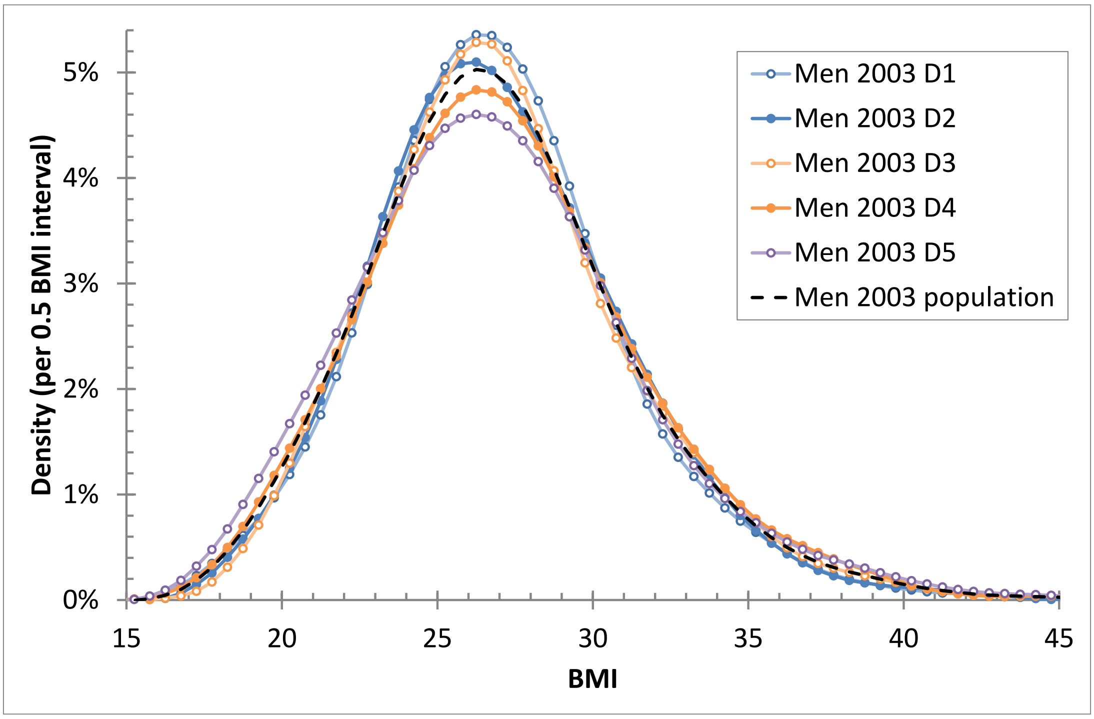


Figure 15.3 BMI distributions for the deprivation quintiles for men for 2003.

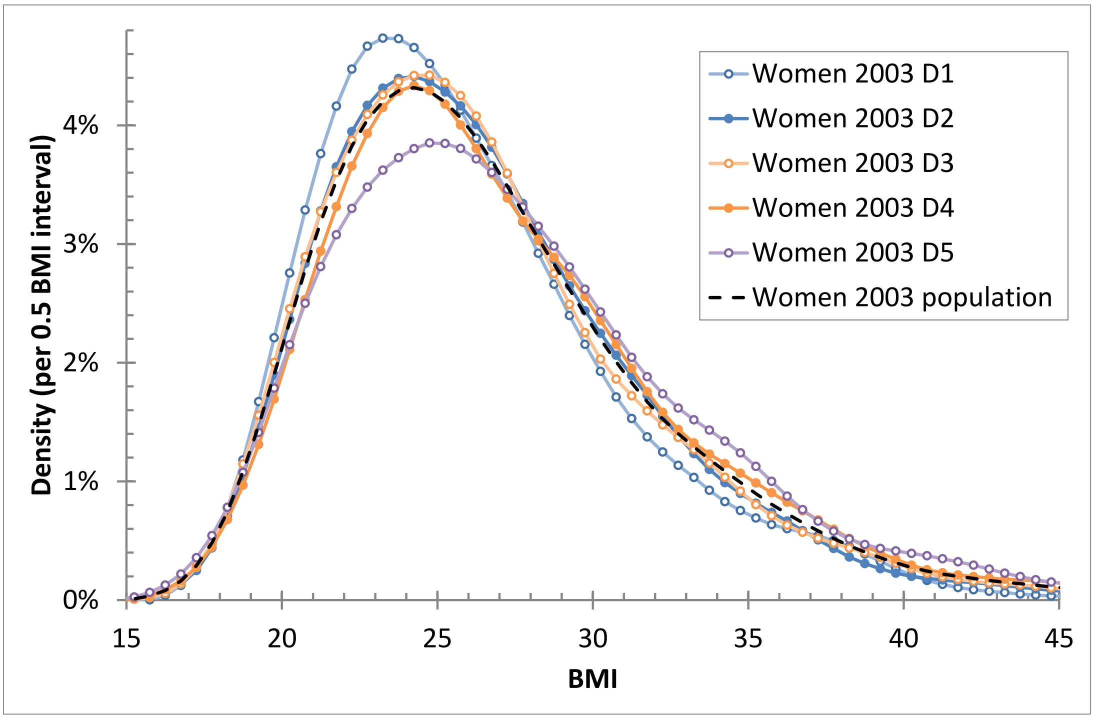


Figure 15.4 BMI distributions for the deprivation quintiles for women for 2003.

Table 15.3 Descriptive statistics for the 2003 distributions for deprivation quintiles for men.

	D1	D2	D3	D4	D5	D5-D1
Number of cases	3340	3061	3225	3454	3672	
Mean	26.95	27.04	27.14	27.17	26.97	0.02
Median	26.71	26.69	26.74	26.81	26.62	-0.08
Modal interval mid-point	26.25	26.25	26.25	26.25	26.25	0.00
Standard deviation	4.12	4.22	4.33	4.55	4.70	0.58
Skewness	0.55	0.70	0.82	0.69	0.72	0.17
Underweight (BMI < 18.5)	1.3%	0.8%	0.6%	1.3%	1.8%	0.5%
Healthy (18.5 ≤ BMI < 25.0)	30.9%	32.1%	31.5%	31.5%	33.5%	2.6%
Overweight (25.0 ≤ BMI < 30.0)	47.8%	45.1%	46.0%	43.6%	42.1%	-5.7%
Obese I (30.0 ≤ BMI < 35.0)	16.4%	18.1%	17.4%	18.3%	17.2%	0.8%
Obese II (35.0 ≤ BMI < 40.0)	3.1%	3.1%	3.7%	4.4%	4.3%	1.2%
Severely obese (BMI ≥ 40.0)	0.6%	0.7%	0.8%	0.9%	1.1%	0.6%
All obese (BMI ≥ 30.0)	20.1%	22.0%	22.0%	23.6%	22.7%	2.6%

Table 15.4 Descriptive statistics for the 2003 distributions for deprivation quintiles for women.

	D1	D2	D3	D4	D5	D5-D1
Number of cases	3896	3601	3882	4249	4512	
Mean	26.06	26.57	26.59	27.13	27.42	1.35
Median	25.17	25.73	25.68	26.09	26.49	1.32
Modal interval mid-point	23.25	24.25	24.75	24.25	24.75	1.50
Standard deviation	5.01	5.19	5.33	5.66	5.83	0.82
Skewness	1.07	1.01	1.06	1.11	0.90	-0.18
Underweight (BMI < 18.5)	1.7%	1.7%	1.8%	1.6%	2.1%	0.4%
Healthy (18.5 ≤ BMI < 25.0)	46.8%	42.1%	42.3%	39.5%	36.6%	-10.2%
Overweight (25.0 ≤ BMI < 30.0)	32.8%	34.2%	34.0%	33.4%	33.3%	0.5%
Obese I (30.0 ≤ BMI < 35.0)	12.5%	15.1%	14.4%	16.0%	17.5%	5.0%
Obese II (35.0 ≤ BMI < 40.0)	5.0%	4.9%	5.2%	6.4%	6.8%	1.9%
Severely obese (BMI ≥ 40.0)	1.3%	2.0%	2.3%	3.0%	3.7%	2.5%
All obese (BMI ≥ 30.0)	18.7%	22.0%	21.9%	25.5%	28.0%	9.3%

15.2.5 Changes over time from 2003 to 2017 for the deprivation categories

The BMI distributions increase from 2003 to 2017 as would be expected from the results in earlier chapters in this report. The distributions and statistics can be compared from the charts and statistics in Sections 15.2.3 and 15.2.4. For example, for men the increases in the mean and the all obese percentage from 2003 to 2017 are D1: 0.28, 3.3%; D2: 0.46, 3.2%; D3: 0.43, 5.2%: D4: 0.60, 5.2%; D5: 1.13, 8.3%. For women the corresponding increases are D1: 0.30, 2.6%; D2: 0.40, 2.5%; D3: 0.95, 6.7%; D4: 0.69, 5.8%; D5: 1.36, 8.9%. So, there are increases from 2003 to 2017 for all quintiles with the highest increases being for D5 for both men and women.

The changes over time were investigated, as before in this report, by looking at the percentiles for the distributions. Figures 15.5 and 15.6 show the percentile increases for each of the 10% percentiles along with 5% and 95%. The results for both men and women are approximately a straight line from the x-axis, indicating that the changes over time are approximately a linear scaling transformation, as for the results in previous chapters. The charts also show the percentile increases for the whole population from the population distributions for 2003 and 2017.

The 1% and 99% values are not included in the charts so as to show the main part of the curves more clearly. These values are more uncertain and there are some irregularities with the small sample sizes, although they mostly continue the general trend of the other values. Charts including these values are in Appendix 15.I (Figure 15.I.5 and Figure 15.I.6).

The chart for men in Figure 15.5 has D5 with the highest increases. Then, D3 and D4 are similar to each other at a lower level. Then, D2 and D1 are similar and slightly lower. The main increases for D1 to D4 start at quite a high BMI value, with little change for BMI values below 25. Hence, the distribution has the main changes only above 25, which is approximately above the 40th percentile. Again, this is similar to what was found in the earlier chapters for recent changes in the full population distributions. The D5 percentiles also have a roughly linear increase above 25 but at a higher level. Below that the increases are small, all being about 0.5 with not much of an increasing slope. This implies a small element of the distribution shifting along the x-axis. For very low BMI values, an increase could potentially be a beneficial change in reducing the proportion in the underweight category.

In the chart for women in Figure 15.6, all the percentile series have a linear scaling type of pattern with approximately linear increases from the x-axis with similar gradients. D5 has the highest increases starting from a BMI of about 19. Then, D3 actually has slightly greater increases than D4 although the differences between them are small. The increases for D1 and D2 are similar and at a lower level and only start at a BMI of about 25. Hence D5 has increases in BMI values above 19 whereas D1 only has increases above a BMI of about 25. For a BMI of 30, the increase for D5 is about 1.8 compared to about 0.6 for D1.

From the previous chapters BMI increases more in the 1990s than in recent years. However, the deprivation variable in the HSE only goes back to 2001 and so this is the only time interval that can be compared. It would be interesting to look at the changes back to the start of the HSE and also how much difference there is between the quintiles in the early 1990s.

Overall, the increases over time for all the deprivation categories have approximately a linear pattern with higher increases for higher BMI values, implying that the changes are mainly a linear scaling transformation. The D5 category has the highest increases, and these are quite a bit larger than for D1. Hence the amount by which BMI and obesity levels for D5 exceed those for D1 will increase from 2003 to 2017, and this is what was found in the results in Sections 15.2.3 and 15.2.4.

From the discussion in Chapter 6, the interpretation of the changes in BMI distributions over time is that it is due to living in a different time period and as a result experiencing different conditions and having different lifestyles. The implication is that, as for the population analysis in earlier chapters, the effect of the different conditions has been greater for those with a higher BMI and this applies for all deprivation categories. Also, the effect of the changes has been generally greater in the higher deprivation quintile areas, particularly D5 compared to D1.

Work on this could be extended in various ways. These include modelling the changes as linear scaling transformations, looking at distributions for each four years of data, using eight years of data to increase sample sizes, and splitting the data into different age groups.

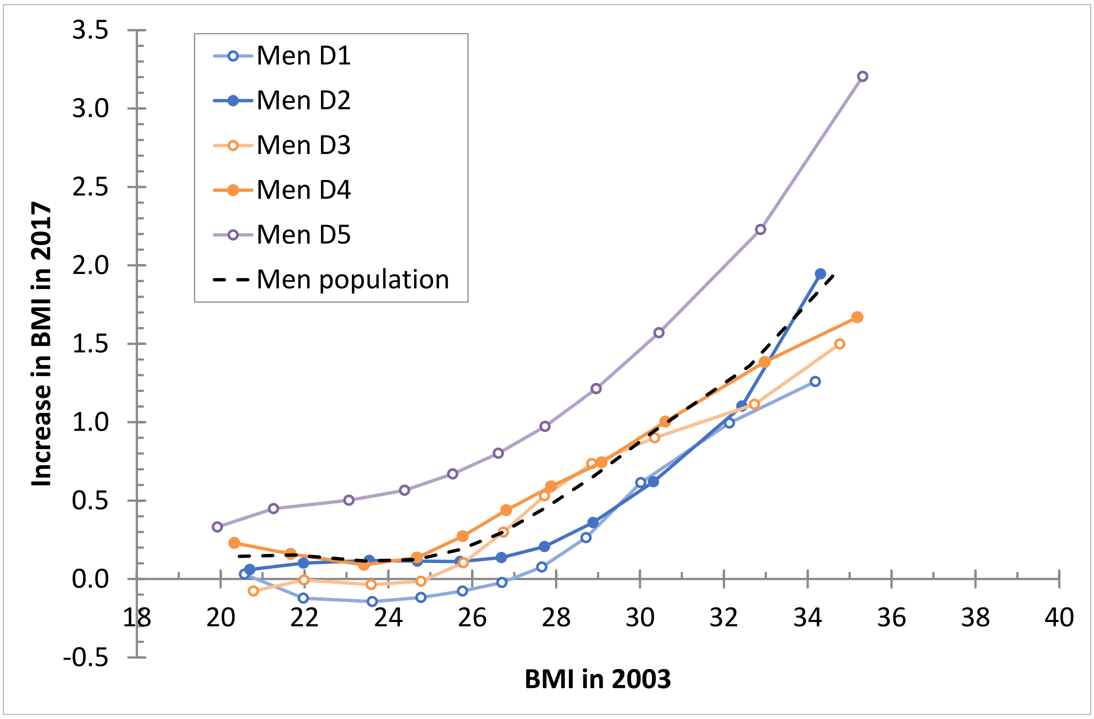


Figure 15.5 Percentile increases for deprivation BMI distributions for men from 2003 to 2017.

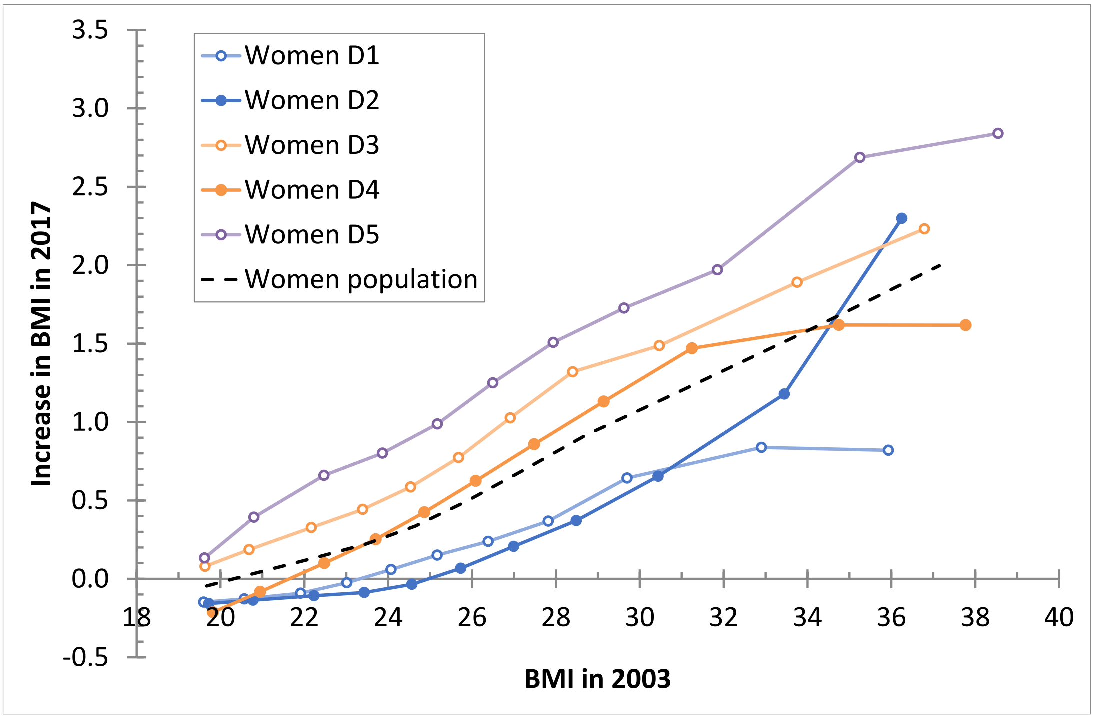


Figure 15.6 Percentile increases for deprivation BMI distributions for women from 2003 to 2017.

15.2.6 Magnitude of the changes between the deprivation categories

One way of considering and assessing the magnitude of the differences between the deprivation categories is by comparison with the changes over time for the whole population considered in the previous chapters of this report. Table 15.5 shows the changes in the statistics for the population distributions from 1993 to 2017 from Table 14.4 in Chapter 14, along with the changes from D1 to D5 for 2017 from Tables 15.1 and 15.2.

In terms of the increase in the mean, reduction in the healthy percentage, and increase in all obese the changes for men between D1 and D5 are not as great as from 1993 to 2017. The differences between D1 and D5 are still substantial, though. The one exception is that the increase in the severely obese percentage is slightly larger than from 1993 to 2017.

On the other hand, the BMI increases from D1 to D5 for women are quite a bit larger than the changes from 1993 to 2017. This emphasises the magnitude of the differences between the D1 and D5 category distributions for women in 2017 as greater even than the change over time for the population from 1993 to 2017.

Table 15.5 Change in the descriptive statistics from 1993 to 2017 and from D1 to D5

	1993 to 2017 Men	1993 to 2017 Women	D1 to D5 Men 2017	D1 to D5 Women 2017
Mean	1.58	1.71	0.88	2.42
Median	1.25	1.44	0.74	2.42
Modal interval mid-point	1.00	1.00	0.50	1.50
Standard deviation	1.17	1.17	0.97	1.20
Skewness	0.22	-0.21	0.05	-0.28
Underweight (BMI < 18.5)	0.1%	-0.1%	0.7%	0.1%
Healthy (18.5 ≤ BMI < 25.0)	-9.7%	-10.8%	-3.5%	-15.2%
Overweight (25.0 ≤ BMI < 30.0)	-3.6%	-0.5%	-4.8%	-0.5%
Obese I (30.0 ≤ BMI < 35.0)	7.8%	5.1%	3.1%	6.6%
Obese II (35.0 ≤ BMI < 40.0)	3.5%	3.5%	2.3%	4.6%
Severely obese (BMI ≥ 40.0)	1.9%	2.8%	2.2%	4.3%
All obese (BMI ≥ 30.0)	13.3%	11.4%	7.6%	15.6%

The changes can also be compared using the changes in the percentile values. Figure 15.7 shows the increases in the corresponding percentile values from D1 to D5 for men and women for both 2003 and 2017. The increases for the whole populations from 1993 to 2017 are also shown as the dotted and dashed lines.

This chart shows some of the differences already identified with the changes between D1 and D5 being greater in 2017 than in 2003 for both men and women, and the changes for women being greater than for men in each year. All the patterns are approximately linear. For women this indicates that the differences between the D1 and D5 distributions are approximately a linear scaling as evidenced by the distributions getting more stretched out in Figures 15.2 and 15.4. For men, the pattern is similar but with negative values for the lowest BMI values reflecting the left tail extending further for D5 compared to D1 as well as the right. The comparison with the percentile increases from 1993 to 2017 again shows that the increases from D1 to D5 are not as big for men as for the whole population from 1993 to 2017 but are larger for women (for most of the starting BMI values) than for the whole population from 1993 to 2017.

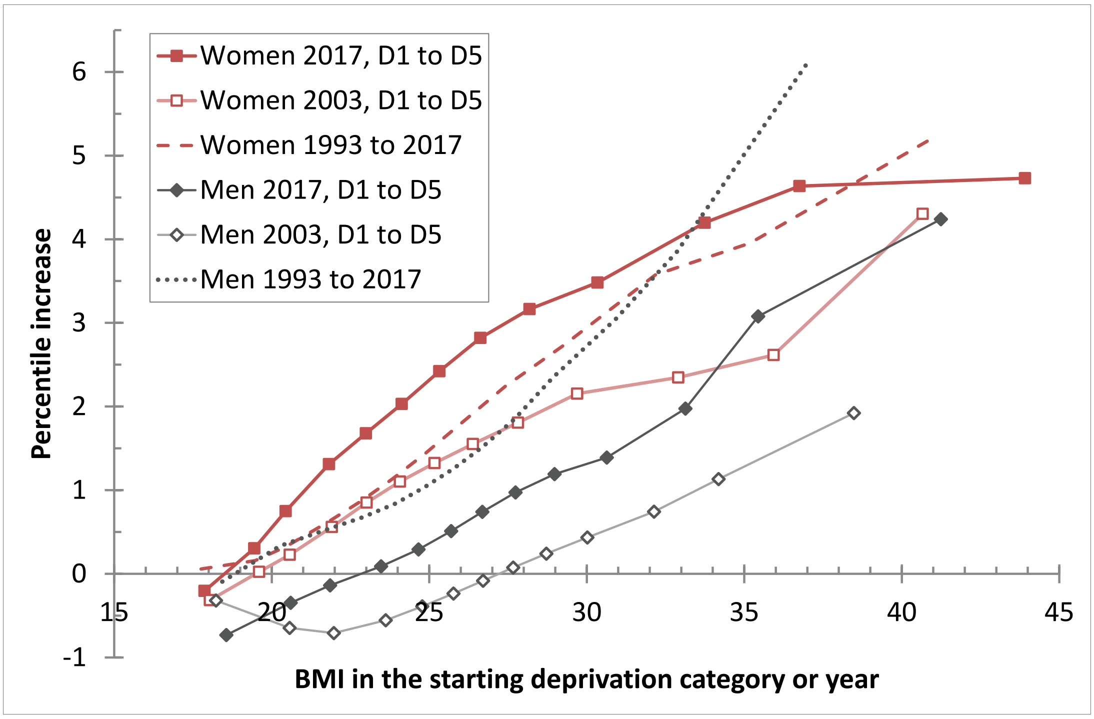


Figure 15.7 Percentile increases from D1 to D5 and from 2003 to 2017.

Another comparison that can be made is with the distributions of earlier years. The D1 category has the lowest BMI levels. If it were possible for the whole population to have the BMI distribution of the D1 category for 2017 then we can consider what this would correspond to in terms of the past population distributions produced in this report for 1993, 1997, 2001, 2005, 2009, 2013, and 2017.

For men, the D1 category distribution is closest to the distribution for 2005. For women, the D1 category distribution does not quite match any of the distributions but is most similar to the distributions for 1997 and 2001 with a mean value in between them. The mean and healthy category percentage for D1 is closer to 1997 than 2001.

Therefore, if the whole population had the BMI distribution for D1 then this would return approximately to the better population BMI levels of 2005 for men and of the late 1990s for women.

15.3 Height

15.3.1 Method used for the height BMI distributions

Another variable that can be considered for BMI is the association with a person's height, h. The work in this section divides the data into three roughly equal height categories and compares the BMI distributions.

Both the analysis and the discussion on this are fairly brief. This is because the interpretation of the results is uncertain. BMI includes height in the equation by dividing by height squared, and this is designed to adjust for height (Section 1.2.1 in Chapter 1). However, it could be that the equation does not account for the effect of height entirely accurately. Differences between the BMI distributions of the height categories could be due to genuine differences in the distribution of obesity levels (i.e., levels of body fat), or the effect of height in the equation, or a mixture of both.

The data was split approximately into thirds. For men this was using heights of 172 and 178 cm as the category boundaries. For women the boundary values used were 159 and 165 cm. The usual methodology in this report was applied to produce BMI distributions, as described for example in Section 15.2.

As for the deprivation BMI distributions (Section 15.2.2), the three different smoothing options of 11 point, 13 point and double smoothing were compared. The samples sizes are a little larger than for the deprivation distributions as the data is split into three categories rather than five. As a result, the distributions are generally a bit smoother. As for the deprivation analysis the double smoothing version was chosen here because it gives the smoothest distribution and this is considered the most plausible for a population BMI distribution. The differences between the distributions for the three smoothing options are small. For example, the BMI category percentages rounded to one decimal place differ by 0.3% between the three options for only two of the 72 values (six categories in the 12 distributions produced), by 0.2% for four of the values, and otherwise differ by 0.1% or are all the same.

15.3.2 Relationship between height and age

The average height of the population has tended to increase in recent times. The Our World in Data website has a web page on "Human Height" (Roser, Appel, and Ritchie, 2013) that has various charts and statistics showing mean height increasing globally over the past 200 years, but with little change in earlier times [44]. The result for the current population is that younger adult age groups tend to be slightly taller on average than older age groups. Figure 15.8 shows the mean height for each of the age groups in the 2015-2018 data, using the case weights (although the values for unweighted data are similar). As with previous charts the mid-point of the integer age values is used, such as 18-19 being plotted at 18.5. The value plotted at age 90 is the age group 90+. There is an increase in height from age group 18-19 to 20-24, particularly for women, presumably because the body is still growing at those ages. However, from ages in the 30s and older the average height generally decreases with age.

The consequence is that the age profiles vary a lot for the different height categories. For men the tallest category of $h \geq 178$ has higher frequencies in age categories up to age 54 compared to the category of $h < 172$, and lower frequencies for ages 55+. The frequencies for age 18-54 and 55+ for the three categories are $h < 172$: 49.7%, 50.3%; $172 \leq h < 178$: 65.6%, 34.4%; $h \geq 178$: 76.1%, 23.9%. The age profile for the middle category is similar to the overall 2011-2014 population age profile used throughout the work in this report. For women the patterns in the age profiles are similar to this. The tallest category of $h \geq 165$ has higher frequencies than $h < 159$ up to age 55-59. The frequencies for age 18-59 and age 60+ are $h < 159$: 55.1%, 44.9%; $159 \leq h < 165$: 72.0%, 28.0%; $h \geq 165$: 84.2%, 15.8%.

Since the younger age groups tend to have a lower BMI (Chapter 7) it is important to adjust the data to the standard population age profile to eliminate the effect of age on BMI. The case weights of the data were adjusted to the standard age profile for the 2011-2014 population (Section 3.3) in the

[44] https://ourworldindata.org/human-height

same way as for all the previous analysis presented in the report (discussed in Section 15.2.2 for example).

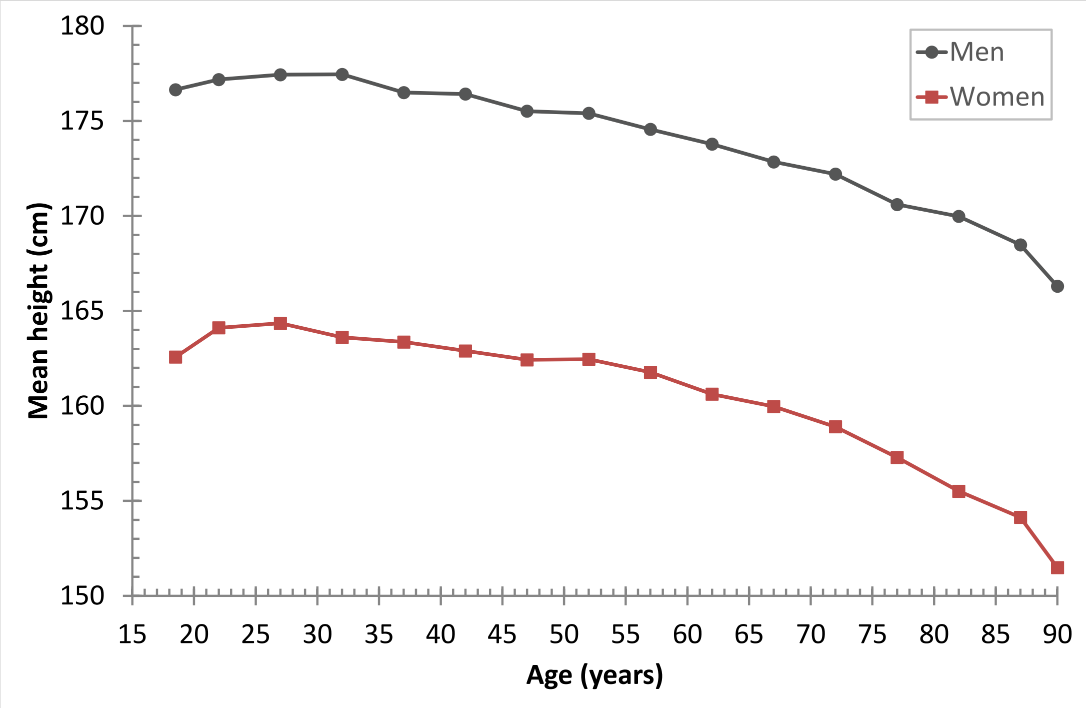


Figure 15.8 Mean height for age groups for the 2015-2018 data for men and women.

15.3.3 Results for the height category BMI distributions

The BMI distributions for the three height categories for 2017 (using 2015-2018 data) are shown in Figure 15.9 for men and Figure 15.10 for women. The statistics are in Tables 15.6 and 15.7. The BMI distributions have lower values for the higher height categories. The differences are greater for the data for women than for men. The difference column in Tables 15.6 and 15.7 gives the difference in statistics values between the tallest and shortest height categories. For the values for the mean, healthy category percentage and all obese category percentage the differences for men are -0.31, 3.6%, and -2.7%, and for women are -1.11, 9.2%, and -6.7%. Hence, the differences are quite large for the data for women.

In the distributions for men, the $h \geq 178$ BMI distribution has a slightly taller peak that is more to the left than the other distributions. The $h < 172$ distribution has a slightly higher left tail with a little more in the underweight category. There are some irregularities in the $h < 172$ distribution with a bump in the right tail. The right tail looks slightly more extended in this distribution, although there is not much difference for BMI values above 40. Overall, the differences are fairly small.

The distributions for women are very similar to each other at the start of the left tail. Each shorter height category has a BMI distribution stretched out more to the right, with a slightly lower peak that is more to the right and a higher right tail. Although again there is not much difference for BMI above 40. As already noted from the statistics, the differences in the distributions are much greater for women than for men.

Distributions were also calculated for 1993 using the first four years of the HSE data. These are shown in Figures 15.11 and 15.12. The statistics are in Tables 15.8 and 15.9. The BMI distributions for men are very similar to each other, with those for $h \geq 178$ and $172 \leq h < 178$ being almost identical. The distribution for $h < 172$ has a slightly lower peak and a right tail that it is little higher around a BMI of 35 but these are still very small differences. The distributions for women have greater differences and the differences between the height categories look similar to those for the 2017 distributions in Figure 15.10.

The statistics are in Tables 15.8 and 15.9, and naturally reflect the patterns in the charts. For example, the differences across the height categories for the tallest compared to the shortest for the mean, healthy category percentage and all obese category percentage are: men -0.21, 0.4%, -2.0%; women -1.29, 10.5%, -7.0%. These differences are large for the data for women and are similar to the 2017 distributions.

The percentiles were also calculated from 1993 to 2017 for each height category. The percentile increases are shown for each of the 10% percentiles along with 5% and 95% in Figures 15.13 and 15.14. The values for the complete population (from Figure 14.14 in Chapter 14) are also shown. The charts also including 1% and 99% are in Appendix 15.II (Figure 15.II.1 and Figure 15.II.2). The increases are quite similar for each height category, although the tallest height category has slightly higher increases than the other two for the data for women. The pattern of the increases over time, as for the population as a whole, is approximately that of a linear scaling.

15.3.4 Discussion of the height category results

As mentioned at the start of this section, the interpretation of these results depends on whether there is any bias in the way that the BMI equation takes account of height. If the BMI distribution adjusts for height in a suitable and fair way, then the implication is that taller categories have slightly lower obesity (i.e., body fat) levels, particularly for women. There could be genetic or physiological reasons for this. There could be diet reasons such as diet during childhood affecting height and obesity.

The Our World in Data web page on human height (Roser, Appel, and Ritchie, 2013) has a section discussing why height has stopped increasing in the last few years [45]. One of the references cited is a study by Schönbeck et al. (2013) looking at data of Dutch children that showed that heights are no longer increasing. In considering the reasons for this, the paper includes the comment: “Unhealthy eating habits may lead to inadequate nutrient intake, which may result in lower height. Furthermore, an unhealthy diet in combination with less energy expenditure due to a sedentary lifestyle leads to an increase in overweight and obesity, a phenomenon that has also been observed in Dutch children. Higher BMI is associated with earlier onset of maturation and menarche, which, in turn, are related to lower height”. In other words, a poor diet may tend to cause both higher BMI and lower height. It is possible, therefore, that this could be the main reason, or at least part of the reason, for the relationships seen in the data in this section with the lower height categories having higher BMI values.

Without knowing with any certainty the reasons for the differences in BMI between the height categories, it is difficult to draw any policy implications from these results. As discussed in Chapter 1, ultimately the interest with obesity is in the risk of health conditions. Potentially, height could be included in studies looking at the prevalence of health conditions to see if it is significant or not as an additional variable to BMI in the association with health risks.

[45] https://ourworldindata.org/human-height section headed “Why has growth in human height stagnated in rich countries?”

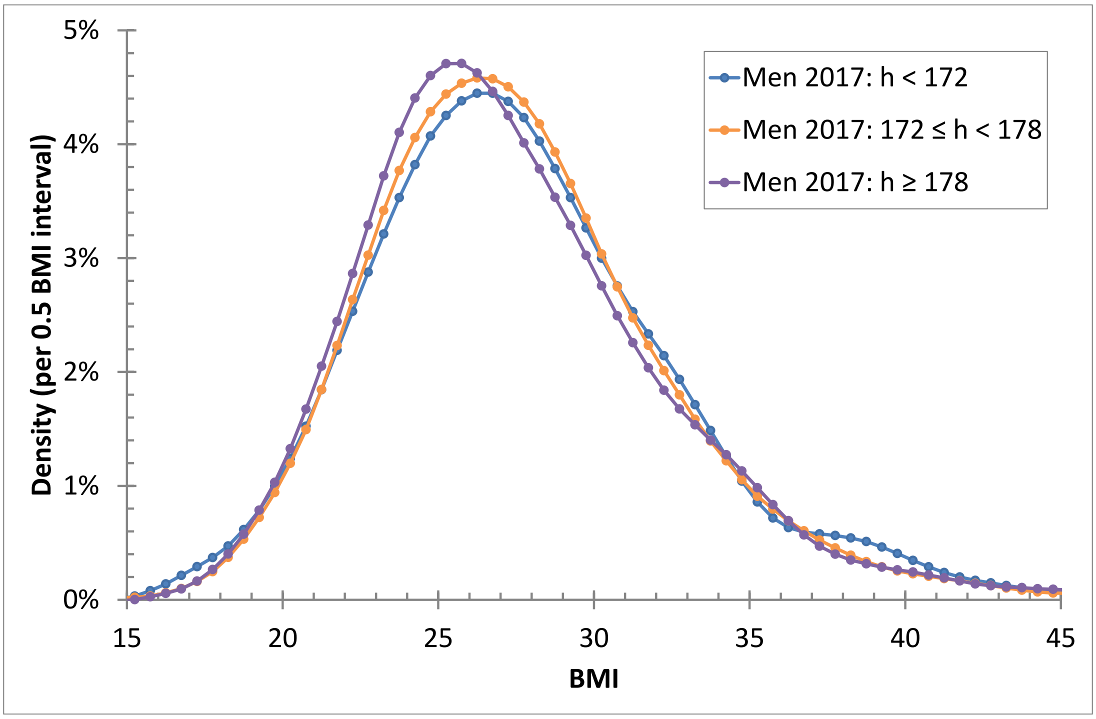


Figure 15.9 BMI distributions for the height categories for men for 2017.

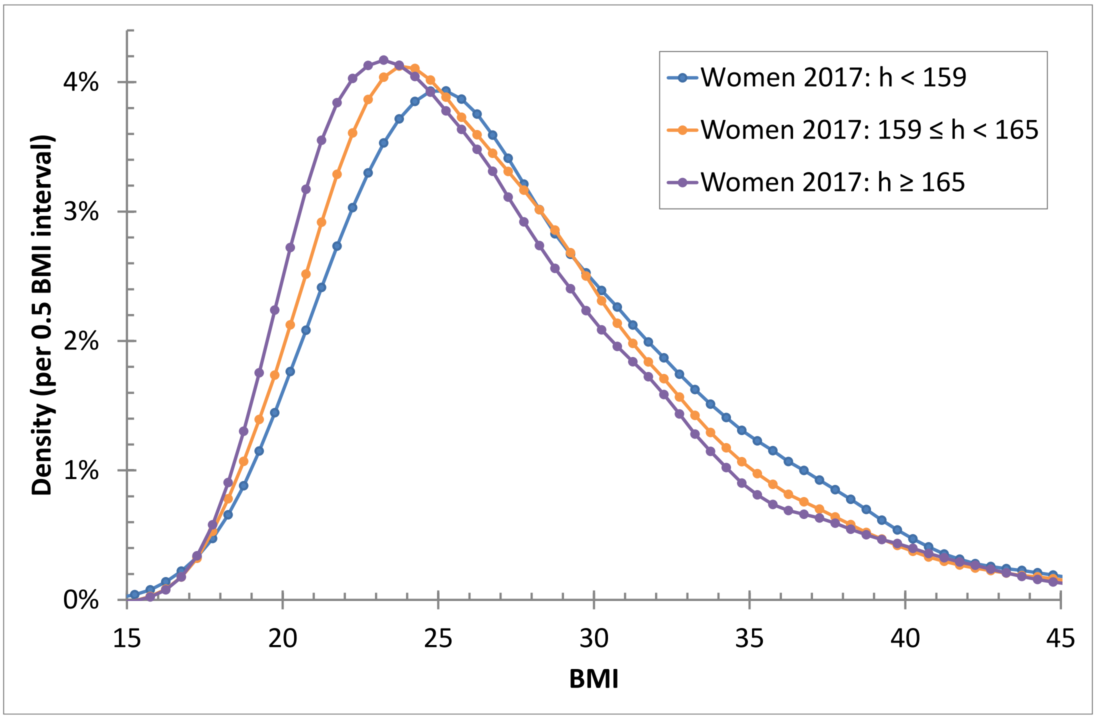


Figure 15.10 BMI distributions for the height categories for women for 2017.

Table 15.6 Descriptive statistics for the 2017 distributions for height categories for men.

	$h < 172$	$172 \le h < 178$	$h \ge 178$	Difference
Number of cases	3828	3824	3993	
Mean	27.76	27.62	27.46	-0.31
Median	27.19	27.08	26.74	-0.46
Modal interval mid-point	26.75	26.25	25.75	-1.00
Standard deviation	5.13	4.93	5.03	-0.10
Skewness	0.80	1.04	1.04	0.24
Underweight (BMI < 18.5)	1.5%	1.0%	1.0%	-0.6%
Healthy (18.5 ≤ BMI < 25.0)	29.2%	30.2%	32.9%	3.6%
Overweight (25.0 ≤ BMI < 30.0)	40.7%	42.1%	40.4%	-0.3%
Obese I (30.0 ≤ BMI < 35.0)	20.2%	19.6%	18.4%	-1.8%
Obese II (35.0 ≤ BMI < 40.0)	5.9%	5.2%	5.2%	-0.7%
Severely obese (BMI ≥ 40.0)	2.4%	1.9%	2.2%	-0.2%
All obese (BMI ≥ 30.0)	28.5%	26.7%	25.8%	-2.7%

Table 15.7 Descriptive statistics for the 2017 distributions for height categories for women.

	$h < 159$	$159 \le h < 165$	$h \ge 165$	Difference
Number of cases	4824	4865	4608	
Mean	27.99	27.38	26.87	-1.11
Median	26.87	26.24	25.66	-1.21
Modal interval mid-point	25.25	23.75	23.25	-2.00
Standard deviation	6.14	6.04	5.99	-0.15
Skewness	0.96	1.21	1.20	0.24
Underweight (BMI < 18.5)	1.9%	1.8%	2.1%	0.1%
Healthy (18.5 ≤ BMI < 25.0)	33.8%	38.8%	43.0%	9.2%
Overweight (25.0 ≤ BMI < 30.0)	32.8%	32.2%	30.2%	-2.6%
Obese I (30.0 ≤ BMI < 35.0)	18.2%	16.5%	15.0%	-3.3%
Obese II (35.0 ≤ BMI < 40.0)	8.9%	6.8%	6.1%	-2.8%
Severely obese (BMI ≥ 40.0)	4.4%	3.9%	3.7%	-0.6%
All obese (BMI ≥ 30.0)	31.4%	27.2%	24.8%	-6.7%

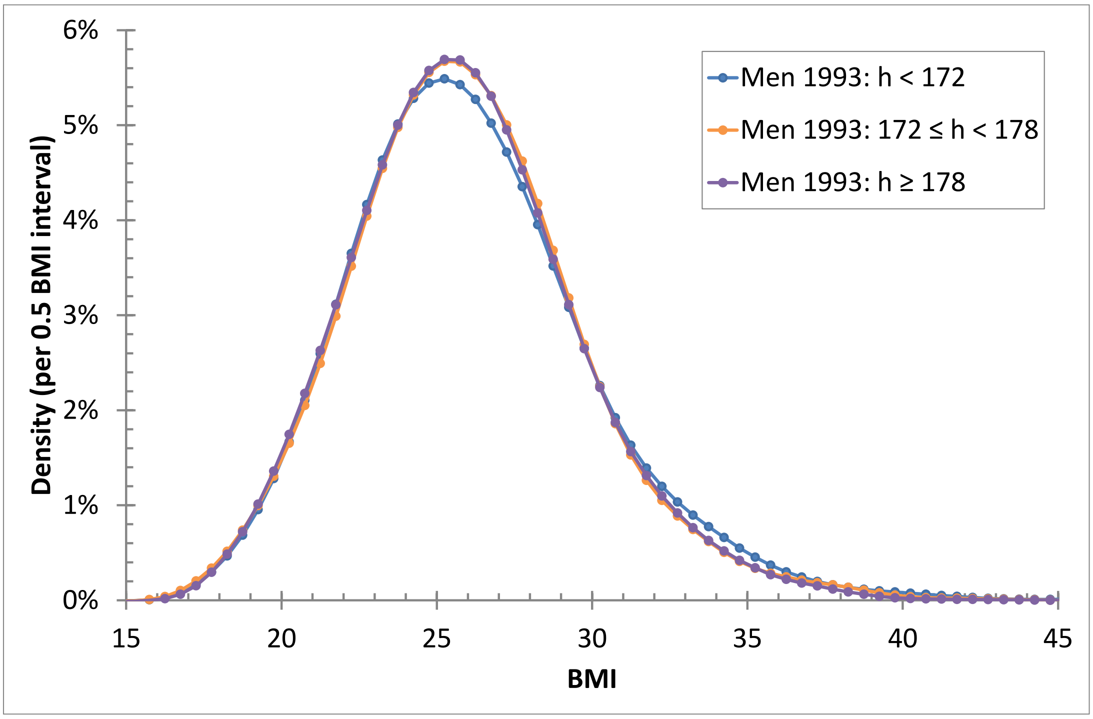


Figure 15.11 BMI distributions for the height categories for men for 1993.

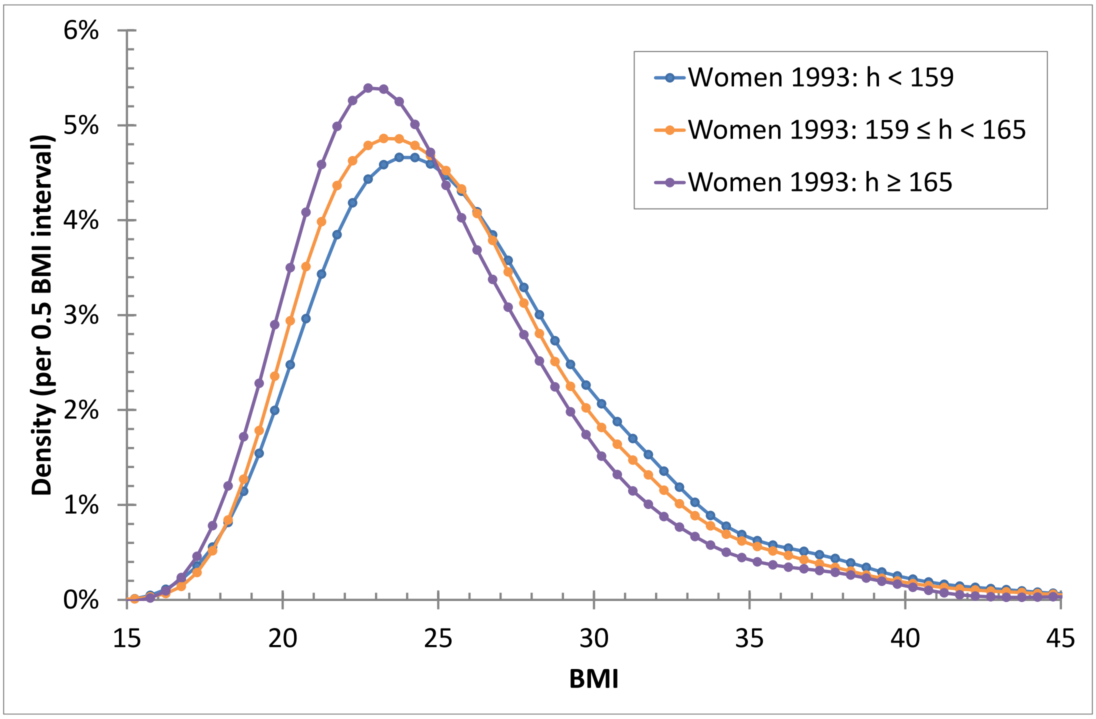


Figure 15.12 BMI distributions for the height categories for women for 1993.

Table 15.8 Descriptive statistics for the 1993 distributions for height categories for men.

	$h < 172$	$172 \le h < 178$	$h \ge 178$	Difference
Number of cases	5951	5403	5271	
Mean	26.14	26.01	25.93	-0.21
Median	25.77	25.76	25.71	-0.06
Modal interval mid-point	25.25	25.25	25.25	0.00
Standard deviation	3.99	3.81	3.67	-0.32
Skewness	0.82	0.77	0.46	-0.37
Underweight (BMI < 18.5)	1.0%	1.2%	1.0%	0.0%
Healthy (18.5 ≤ BMI < 25.0)	40.6%	40.2%	41.0%	0.4%
Overweight (25.0 ≤ BMI < 30.0)	43.5%	45.5%	45.1%	1.7%
Obese I (30.0 ≤ BMI < 35.0)	12.3%	11.1%	11.3%	-1.0%
Obese II (35.0 ≤ BMI < 40.0)	2.2%	1.8%	1.5%	-0.7%
Severely obese (BMI ≥ 40.0)	0.4%	0.2%	0.1%	-0.4%
All obese (BMI ≥ 30.0)	14.9%	13.1%	12.9%	-2.0%

Table 15.9 Descriptive statistics for the 1993 distributions for height categories for women.

	$h < 159$	$159 \le h < 165$	$h \ge 165$	Difference
Number of cases	7023	6477	5338	
Mean	26.25	25.75	24.97	-1.29
Median	25.38	24.93	24.19	-1.19
Modal interval mid-point	23.75	23.25	22.75	-1.00
Standard deviation	5.22	4.88	4.49	-0.73
Skewness	1.43	1.24	1.10	-0.32
Underweight (BMI < 18.5)	2.1%	1.9%	2.8%	0.7%
Healthy (18.5 ≤ BMI < 25.0)	44.5%	48.8%	55.1%	10.5%
Overweight (25.0 ≤ BMI < 30.0)	34.0%	32.9%	29.8%	-4.2%
Obese I (30.0 ≤ BMI < 35.0)	13.1%	11.4%	8.8%	-4.3%
Obese II (35.0 ≤ BMI < 40.0)	4.4%	3.7%	2.9%	-1.6%
Severely obese (BMI ≥ 40.0)	1.9%	1.4%	0.7%	-1.2%
All obese (BMI ≥ 30.0)	19.4%	16.4%	12.4%	-7.0%

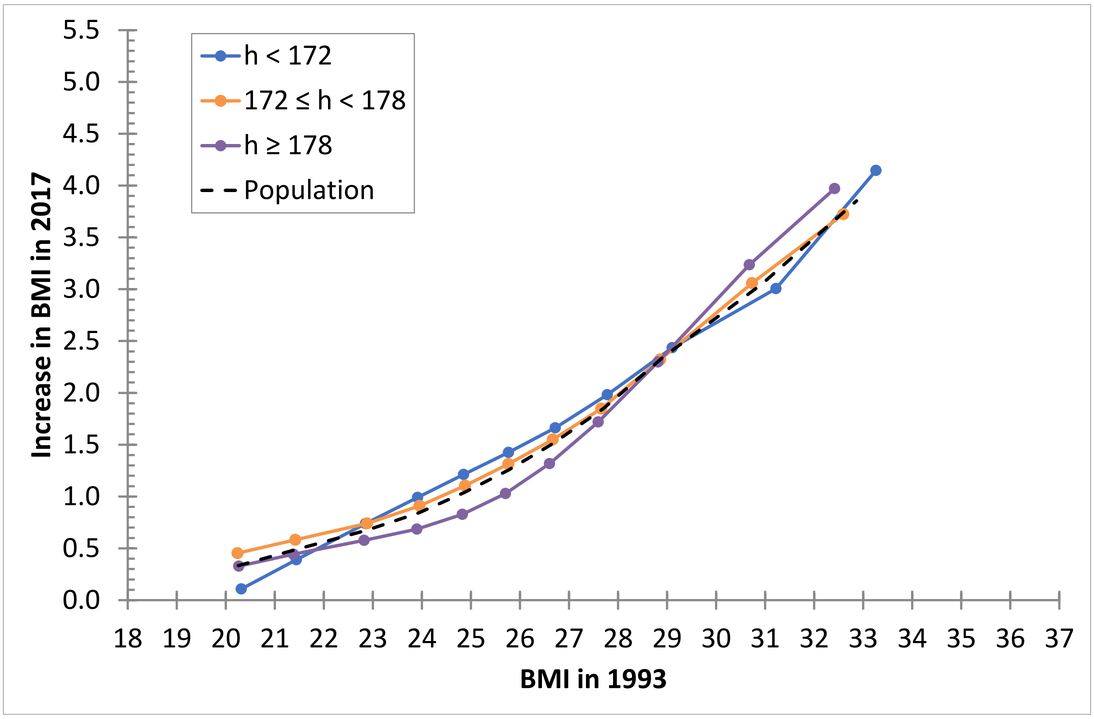


Figure 15.13 Percentile increases for the height BMI distributions for men from 1993 to 2017.

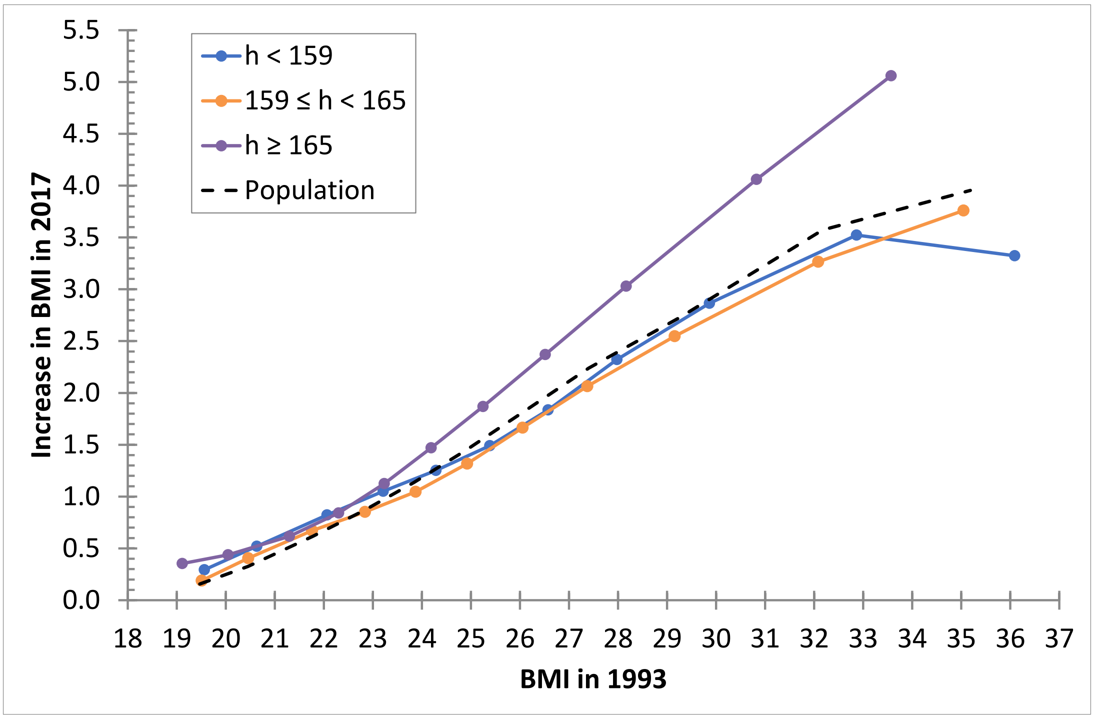


Figure 15.14 Percentile increases for the height BMI distributions for women from 1993 to 2017.

Chapter 16. Contributions and conclusions

Key points from this chapter:

- A summary is given of the main contribution and conclusions of the work in this report. The contributions include:
  - Developing a method for deriving population BMI distributions from the HSE sample data.
  - Comparing the rates of increase in BMI over time and identifying that the pattern of the changes is approximately a linear scaling transformation.
  - Deriving a BMI aging profile from age and cohort analysis.
  - Analysing how the BMI distributions vary with age.
  - Producing future BMI scenarios.
  - Converting BMI changes into calorie values.
  - Showing how the method can be applied to the relationship with other variables such as deprivation.
- Implications of the increases in BMI over time include an expected increase in the prevalence of obesity related health conditions. This is likely to result in reduced quality of life for some in the population, and also increased health costs. It could be a factor that significantly reduces average life expectancy.
- Various possible future work is suggested. This includes extending the results from the report as new HSE data becomes available.

16.1 Summary of contributions

This section gives some brief comments on the main contributions and conclusions of the work. The main points from the report are also listed in the Report Summary at the front of the report, and in addition each chapter gives a list of key points at the start of the chapter. The research questions are set out in Chapter 1 in Section 1.4, and the work in the report has addressed each of them.

Overweight and obesity (being high levels of fat in the body) is an important and significant current health issue. It is associated with an increased risk of a range of serious health conditions. One measure of body type that is often used as an indicator of obesity is body mass index (BMI). The work set out in this report aims to give a detailed analysis of how adult BMI has changed for the population of England since the early 1990s.

The data used is from the annual Health Survey for England (HSE), which is an excellent data source. The initial work used the HSE data for 1991-2014. Prior to the report being completed, HSE data for 2015-2018 was available which enabled most of the results to be extended and updated.

The main method developed in this research was a way to derive population BMI distributions from the HSE sample data. This method involves combining several years of HSE data, using a consistent age profile, grouping the data into narrow BMI intervals of width 0.5, and applying the Savitzky-Golay smoothing method (Savitzky and Golay, 1964; Steinier et al., 1972). It was found that the Savitzky-Golay method works very well in smoothing the irregularities in the sample data distribution whilst maintaining the general shape of the data. It also leaves the mean, standard deviation and skewness value unchanged from the values in the data distribution.

The initial application of this method used the HSE data in groups of four years to produce population BMI distributions for 1993, 1997, 2001, 2005, 2009, 2013, and 2017. Comparison of the distributions enabled the magnitude and the nature of the increases to be investigated. The increases in BMI are much greater in the 1990s than in recent years. From looking at the changes in the shape of the distribution and the increases in the percentiles, the pattern of the increases appeared to be approximately a linear scaling. Linear scaling transformations were therefore applied to model the changes from each distribution to the next and were able to match the distributions well. The overall changes from 1993 to 2013 and from 1993 to 2017 could also be modelled well either by combining the linear scaling models between the successive populations or as a single linear scaling transformation.

The scaling models along with the analysis of the increases in the percentiles and the charts of the distributions give good insights into exactly how BMI has changed over time. The pattern is that the BMI increases between the equivalent percentiles in the populations tend to be greater for those with a higher BMI than those with a lower BMI. More specifically, the BMI increases between the percentiles in the two populations are related to the starting BMI in an approximately linear relationship with a positive gradient.

When interpreting changes in the population BMI distributions at two points in time, the difference is that the two populations have lived through different time periods. Therefore, they have experienced different conditions and have different lifestyles. These result from changes in many areas of life such as the food industry, work, leisure, the economy, and society. One implication of the way that the BMI distributions have increased over time is that the lifestyle changes that have happened have affected the BMI of those with a higher BMI much more than those with a lower BMI.

Having established the changes over time at an overall population level, further analysis was then done to look at the relationship between mean BMI and age, and how this has changed over time. A key part of this was to consider the data within narrow birth cohorts. This showed a consistent pattern within each of the cohorts of BMI increasing with age, but with the rate of increase getting smaller with age. In particular, for the adult ages considered in this report of 18 and over, BMI increases the most at the youngest ages. Older cohorts are at lower levels, presumably because of experiencing different conditions earlier in life, but show a similar pattern of BMI increasing with age. Over the past few years, the cohorts have tended to increase in level and they all seem to be converging towards an aging profile reflecting modern conditions.

BMI distributions were also produced for age groups and for single years of ages. The changes over time for each of the age groups are similar to that of the population as a whole with the largest increases in the earlier years of the HSE and the pattern of increases being approximately a linear scaling transformation.

The changes with age were also investigated by comparing the distributions for the different ages. The increases in BMI for men over about age 24 are mainly a shift in the distribution along the x-axis (i.e., a translation), with little change in shape. In other words, the tendency (on average) to increase BMI is quite similar for each BMI level. For women, the increases with age are a combination of a shift and a linear scaling and so this is a somewhat different pattern with some change in the shape of the distribution. This means that the tendency for BMI to increase with age for women is greater for those with higher BMI.

The current BMI distributions for age 18 are relatively healthy compared to older adults, albeit that they were even better in the 1990s. BMI increases rapidly from age 18 with each year of age. This indicates that targeting health messages and interventions at those of age of about 18 could be a particularly effective and important strategy.

BMI increases with age for both men and women but with different patterns. One notable point is that the distributions for men and women for age 18 are extremely similar. However, the distributions for men and women become quite different as age increases. In the ages in the 30s, 40s and 50s, the distributions for men are more symmetrical than those for women, with a peak further to the right and a lower right tail. These different shapes mean that the healthy percentages at these ages are much lower for men and the overall percentage being overweight or obese is much higher. However, high BMI values are more prevalent for women.

There may be some implications for health interventions in the different ways that the BMI distributions change with age for men and women. BMI has a tendency to increase in a similar way for all BMI levels for men. This means that even those with a fairly low BMI tend to have considerable increases in BMI with age. The result is a low percentage for men in the healthy BMI category as age increases. For women, those with a low BMI tend to have relatively low increases in BMI with age, but those with a high BMI tend to have higher increases. This results in more women staying within the healthy BMI category than for men as age increases. However, it also results in more women than men having very high levels of BMI such as the severely obese category. Hence, for general strategy for health interventions, a broader strategy may be better for men and a more targeted approach may be better for women.

The next aspect of the work was to apply the patterns identified to model future BMI scenarios. A lognormal distribution (with a slight modification for men) was found to represent the distributions well for each year of age. Mean and standard deviation values are required for each year of age. These are used as the parameter values for the lognormal distributions, which are then combined together to give the population BMI distribution. Particular scenarios can be modelled by specifying the set of mean and standard deviation values for all the ages.

One scenario that was examined, called the "current trend scenario", considers what would happen if the aging profile in the future follows the hypothesised pattern for modern current conditions. This is based on data and patterns for the recent years for the HSE that are considered here. This gives a small increase in BMI compared to the 2013 distributions, with mean BMI values for men and for women about 0.4 higher and healthy category percentages about 3% less. Subsequent comparison with the 2017 distributions showed that the prevalence of some of the highest BMI values in the 2017 distributions have actually already exceeded the prevalence in the current trend model.

The other scenario examined was a "low BMI" scenario. The aim with this one was to apply a suitable aging pattern curve to generate population BMI distributions similar to 1993. The scenario therefore gives one way of getting back to 1993 BMI levels, and the mean and standard deviation values for each age could therefore be used as target values to aim for. The benefits would be considerable. For example, the healthy category percentages in this scenario are 9.1% higher than for 2017 for men, and 9.7% higher for women. There are considerable reductions in each of the obese category percentages compared to 2017.

The next piece of analysis considered what the BMI changes correspond to in terms of calories. This used energy balance equations from the literature to produce values to convert from BMI to calories. This was applied to the overall increase in BMI from 1993 to 2013 and from 1993 to 2017, and to the increase in BMI from age 18 to 24. In each of these cases the calculated increase in daily calories was relatively small, despite the large changes in the BMI distributions. For example, the increase in calories from 1993 to 2017 for a BMI of 30 at a median activity level was calculated as 141 kcal / day for men and 102 kcal / day for women. From age 18 to 24 most of the values for men for the median activity level are about 170 kcal / day, and the values vary for women up to a maximum of about 140 kcal / day. These are all estimated long term average amounts at a population level, and so they are not applicable to individuals because of the variations of circumstances and physiology that exist at an individual level.

The last element of the work looked at the relationship of BMI with the variables of deprivation and height. This illustrated how the method developed in the report can be applied to look at the effect of other variables. Deprivation was examined using the five categories in the HSE data. BMI tends to be higher for the more deprived areas with particularly large differences in the data for women. Three categories were used for height, with BMI tending to be slightly lower for the tallest height category.

Overall, the work presented here hopefully gives a detailed and useful insight into how BMI has increased for the adult population of England since the early 1990s. This should help with understanding and identifying the reasons for the increases, and hence with considering the best ways of trying to improve the situation. The methods developed and described in the report can also be applied to future data, as it becomes available, to monitor closely the changes in BMI. This should help with the evaluations of any interventions at a population level. It should also be useful for identifying the effects of any changes at a population level in the various obesity factors.

16.2 Implications for health and life expectancy

The data analysis has shown that BMI and levels of obesity have increased substantially since the early 1990s, particularly during the 1990s but also in the years since then. Given the assumed link with a variety of health conditions this would be expected to lead to increased prevalence of those health conditions. This will reduce the quality of life for those affected and, in some cases, will reduce life span. Hence, there is likely to be an effect on average life expectancy data. There is also likely to be impacts on health costs in treating the conditions.

I am not sure of the exact timescales of such effects but there is likely to be a time lag between changes in obesity levels and changes in the prevalence of health conditions. The risks are also likely to increase with the length of time that a person has higher levels of BMI. Changes in the 1990s may therefore have their main effects many years later. For example, it may have had a significant effect on life expectancy in recent years. Analysis of these relationships and estimations of the effects of the increases in BMI on health, disease prevalence, and costs is outside the scope of this report but would be interesting further work.

16.3 Further work

Ideas for further work include the following areas:

- The results in this report can be updated in the future once new data becomes available. This assumes that the HSE continues and that the same variables that were used here are still provided.
- Following the methodology used in the report, the next population distributions for men and women would be for 2021 and would be based on the HSE data for 2019-2022. The HSE survey could not be completed for 2020 because of the Covid-19 pandemic and so no data is available for 2020. However, the 2021 distributions could still be produced using the HSE data for 2019, 2021, and 2022, although this would likely to be a smaller sample size of data than for the previous distributions. Based on previous schedules the 2022 HSE data will probably be available at the end of 2023. Alternatively, the next distribution produced could again be based on four years of data and use the data for 2019 and 2021-2023.
- Population distributions could be produced using different groups of data for particular reasons. This does not have to be in groups of four years of data. For example, using groups of years of HSE data immediately before and after the Covid-19 pandemic to examine changes in BMI that might be due to the effects of the pandemic.
- The methodology used here could be applied to other variables, such as alternative obesity measures (Section 1.2.2) as long as suitable data is available. In particular, the same analysis as presented in this report could be done for any variables in the HSE. This would, however, be a substantial piece of work.
- It would be interesting to analyse older datasets to look at BMI values further in the past. I am not sure exactly what data is available, but perhaps it may be possible to do some analysis from data in military service records, for example.
- It would also be a useful comparison to apply the same analysis approach to data from other countries to compare the results and the patterns.
- Confidence intervals could potentially be calculated for some of the results, probably with a bootstrapping approach (Section 4.8.4).
- The models and patterns identified here for how BMI varies with age might be useful for studies looking at the association between BMI and health conditions. This potentially could enable comparisons or adjustments where BMI is measured at different ages for different people, by estimating lifetime BMI trajectories.
- Any models for the relationship of BMI to disease prevalence or health costs could be combined with the results here to estimate the effects of the increases in BMI on disease levels and costs. These could also be compared with actual data and, for example, could help with understanding and explaining any changes in disease prevalence.
- The characteristics and lifestyle variables of the cases in chosen BMI categories, such as high BMI, could be examined using other variables in the HSE.
- The analysis and models here could be used to predict the effects of certain policy interventions. This might include predictions of the long term BMI aging trajectory from interventions targeted at the younger age groups. It also might include predictions of the effects of changes in calorie consumption from the models and analysis in Chapter 13.

16.4 Conclusions

The objective of this work is stated in Section 1.4 as: "The overall objective of the work presented in this report is to identify the patterns of the changes over time in the BMI distributions for men and for women in England since the start of the HSE surveys in 1991. One of the main aims of this report is to provide a detailed record of how BMI has changed since the early 1990s, which will hopefully be a useful resource as a long term reference."

As indicated in the objective, one of the main contributions of the work in this report is to provide considerable detail on the patterns and the changes in BMI. This has included producing distributions of BMI using intervals of narrow width 0.5, and analysing the changes over time with percentiles and scaling models. This gives additional information to that previously available such as from the Excel Data Tables (Section 1.2.5).

The standard BMI categories that are often used in obesity research and in data analysis are very wide, covering a considerable range of weights (Section 1.2.4). Statistics for these BMI categories are useful and they have been used in this report. However, the information from the BMI category statistics is inevitably limited because they are broad. Looking at BMI in more detail than this using the much narrower intervals in the distributions produced here has given very useful insights into how BMI has changed since the early 1990s. In general, this level of detail will often be a much better approach in understanding BMI patterns and relationships. This also includes studies looking at the relationship between BMI and health conditions.

Other elements of the work here have included looking at factors and relationships with BMI. One major area is how BMI changes with age, with an important aspect being to take account of different birth years by look at age cohorts. This gives insights into how and why BMI has changed over time. The relationship of BMI with age has also led on to a model framework, which enabled scenarios to be produced for some different situations.

Other relationships investigated included translating the BMI results into calorie amounts using models from the literature, and comparing the differences in BMI for different categories of deprivation and height.

The approach taken here provides a general method for producing and analysing population BMI distributions from the HSE or from similar data. This can be applied to investigate other aspects of obesity data such as new data in the future, the relationship of BMI with other variables, data from other countries, and other measures of obesity. In particular, hopefully the results here can be extended in the future as new HSE data becomes available.

References

The references are listed here by chapter and then in alphabetical order. The references for the Health Survey for England (HSE) are in the last subsection at the end of these references.

Report summary

Henry, C. J. K. (2005). Basal metabolic rate studies in humans: measurement and development of new equations, Public Health Nutrition, 8(7A), 1133–1152.
DOI: https://doi.org/10.1079/PHN2005801

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 36(8), 1627-1639.
DOI: https://doi.org/10.1021/ac60214a047

Steinier, J., Termonia, Y., and Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 44(11), 1906-1909.
DOI: https://doi.org/10.1021/ac60319a045

WHO (2021). Obesity and overweight, Fact sheet, Published online.
https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

Chapter 1

Ahern, A. L., Wheeler G. M., Aveyard P., et al. (2017). Extended and standard duration weight loss referrals for adults in primary care (WRAP): a pragmatic randomised controlled trial, Lancet, 389(10085), 2214-2225.
DOI: https://doi.org/10.1016/S0140-6736(17)30647-5

Astbury, N. M., Aveyard, P., Nickless, A., Hood, K., Corfield, K., Lowe, R., and Jebb S. A. (2018). Doctor Referral of Overweight People to Low Energy total diet replacement Treatment (DROPLET): pragmatic randomised controlled trial, BMJ, 362, k3760.
DOI: https://doi.org/10.1136/bmj.k3760

Aune, D., Sen, A., Prasad, M., Norat, T., Janszky, I., Tonstad, S., Romundstad, P., and Vatten, L. J. (2016). BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants, BMJ, 353(8056), i2156.
DOI: https://doi.org/10.1136/bmj.i2156

Aveyard P., Lewis A., Tearne S., et al. (2016). Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial, Lancet, 388(10059), 2492-2500.
DOI: https://doi.org/10.1016/S0140-6736(16)31893-1

Blackburn, H., and Jacobs Jr., D. (2014). Commentary: Origins and evolution of body mass index (BMI): continuing saga, International Journal of Epidemiology, 43(3), 665–669.
DOI: https://doi.org/10.1093/ije/dyu061

Di Angelantonio, E, et al. (2016). Body-mass index and all-cause mortality: individual participant-data meta-analysis of 239 prospective studies in four continents, Lancet, 388 (10046), 776-786.
DOI: https://doi.org/10.1016/S0140-6736(16)30175-1

Eknoyan, G. (2008). Adolphe Quetelet (1796–1874)—the average man and indices of obesity, Nephrology Dialysis Transplantation, 23, 47-51.
DOI: https://doi.org/10.1093/ndt/gfm517

Flegal, K. M., Kit, B. K., Orpana, H., and Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis, JAMA: The Journal of the American Medical Association, 309(1): 71–82.
DOI: https://doi.org/10.1001/jama.2012.113905

Guh, D. P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C. L., and Anis, A. H. (2009). The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis, BMC Public Health, 9(88).
DOI: https://doi.org/10.1186/1471-2458-9-88

Hazlehurst, J.M., Logue, J., Parretti, H.M. et al. (2020). Developing Integrated Clinical Pathways for the Management of Clinically Severe Adult Obesity: a Critique of NHS England Policy, Current Obesity Reports, 9, 530-543.
DOI: https://doi.org/10.1007/s13679-020-00416-8

Heggie L., Mackenzie, R. M., Ells, L. J., Simpson, S. A., and Logue, J. (2020). Tackling reporting issues and variation in behavioural weight management interventions: Design and piloting of the standardized reporting of adult behavioural weight management interventions to aid evaluation (STAR-LITE) template, Clinical Obesity, 10(5), e12390.
DOI: https://doi.org/10.1111/cob.12390

Heymsfield, S.B., and Waki, M. (1991). Body composition in humans: advances in the development of multicompartment chemical models, Nutrition Reviews, 49(4), 97-108.
DOI: https://doi.org/10.1111/j.1753-4887.1991.tb02997.x

Inoue S. et al. (2000). The Asia-Pacific perspective: redefining obesity and its treatment, Sponsored by the Regional Office for the Western Pacific of the World Health Organisation, the International Association for the Study of Obesity and the International Obesity Task Force, Published by Health Communications Australia Pty Limited.
https://www.vepachedu.org/TSJ/BMI-Guidelines.pdf

Johnson, P.C.D., Logue, J., McConnachie, A., Abu-Rmeileh, N.M.E., Hart, C., Upton, M.N., Lean, M., Sattar, N., and Watt, G. (2012). Intergenerational change and familial aggregation of body mass index, European Journal of Epidemiology, 27, 53–61.
DOI: https://doi.org/10.1007/s10654-011-9639-5

Kent, S., Fusco, F., Gray, A., Jebb, S. A., Cairns, B. J., and Mihaylova, B. (2017). Body mass index and healthcare costs: a systematic literature review of individual participant data studies, Obesity Reviews 18(8), 869-879.
DOI: https://doi.org/10.1111/obr.12560

Keys, A., Fidanza, F., Karvonen, M. J., Kimura, N., Taylor, H. L. (1972). Indices of relative weight and obesity, Journal of Chronic Diseases, 25, 329–343. Reprinted in: International Journal of Epidemiology, 43(3), 655–665, 2014.
DOI: https://doi.org/10.1093/ije/dyu058

Koliaki, C., Liatis, S., and Kokkinos, A. (2019). Obesity and cardiovascular disease: revisiting an old relationship, Metabolism Clinical and Experimental, 92, 98-107.
https://doi.org/10.1016/j.metabol.2018.10.011

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., and Straif, K. (2016). Body fatness and cancer — viewpoint of the IARC working group,
New England Journal of Medicine; 375(8), 794-798.
DOI: https://doi.org/10.1056/NEJMsr1606602

Lim, S. S., Vos, T., Flaxman, A. D. et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010, Lancet, 380, 2224-2260.
DOI: https://doi.org/10.1016/S0140-6736(12)61766-8

Liu, X., and Brooks, R. J. (2016). Agent-based simulation of herding in financial markets, Proceedings of the Operational Research Society Simulation Workshop 2016 (SW16), Operational Research Society, 45-53.
https://eprints.lancs.ac.uk/id/eprint/79266

Lukaski, H. C. (2014). Commentary: Body mass index persists as a sensible beginning to comprehensive risk assessment, International Journal of Epidemiology, 43(3), 669-671.
DOI: https://doi.org/10.1093/ije/dyu059

Ma, C., Avenell, A., Bolland, M., Hudson, J., Stewart, F., Robertson, C., Sharma, P., Fraser, C., MacLennan, G. (2017). Effects of weight loss interventions for adults who are obese on mortality, cardiovascular disease, and cancer: systematic review and meta-analysis, BMJ, 359, j4849.
DOI: https://doi.org/10.1136/bmj.j4849

Murray, C. J. L., et al. (2013). UK health performance: findings of the Global Burden of Disease Study 2010, Lancet, 381(9871), 997-1020.
DOI: https://doi.org/10.1016/S0140-6736(13)60355-4

Murray, C. J. L., et al. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet, 396(10258), 1223-1249.
DOI: https://doi.org/10.1016/S0140-6736(20)30752-2

NCD-RisC: NCD Risk Factor Collaboration (2016), Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants, Lancet, 387(10026), 1377-1396.
DOI: https://doi.org/10.1016/S0140-6736(16)30054-X

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Sattar, N., and Preiss, D. (2017). Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? Circulation, 135, 2369–2372.
DOI: https://doi.org/10.1161/CIRCULATIONAHA.117.028307

Tuson, M. (2019). Understanding the Impact of Social Networks on the Spread of Obesity, Ph.D. Thesis, Cardiff University, U.K.
http://orca.cf.ac.uk/128597/

Wardle, J. and Boniface, D. (2008). Changes in the distributions of body mass index and waist circumference in English adults, 1993/1994 to 2002/2003, International Journal of Obesity, 32, 527–532.
https://doi.org/10.1038/sj.ijo.0803740

Wells, J. C. K. (2014). Commentary: The paradox of body mass index in obesity assessment: not a good index of adiposity, but not a bad index of cardio-metabolic risk, International Journal of Epidemiology, 43(3), 672–674.
DOI: https://doi.org/10.1093/ije/dyu060

WHO (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies, Lancet, 363(9403), 157-163.
DOI: https://doi.org/10.1016/S0140-6736(03)15268-3

WHO (2011). Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation, Geneva, 8–11 December 2008, World Health Organisation.
https://www.who.int/publications/i/item/9789241501491

WHO (2021). Obesity and overweight, Fact sheet, Published online.
https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

Yu, E., Stokes, A. C., Ley, S. H., Manson, J. E., Willett, W., Satija, A., and Hu, F. B. (2017). Weight history, all-cause and cause-specific mortality in three prospective cohort studies, Annal of Internal Medicine, 166(9), 613–620.
DOI: https://doi.org/10.7326/M16-1390

Chapter 2

Wardle, J. and Boniface, D. (2008). Changes in the distributions of body mass index and waist circumference in English adults, 1993/1994 to 2002/2003, International Journal of Obesity, 32, 527–532.
https://doi.org/10.1038/sj.ijo.0803740

Chapter 3

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 36(8), 1627-1639.
DOI: https://doi.org/10.1021/ac60214a047

Steinier, J., Termonia, Y., and Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 44(11), 1906-1909.
DOI: https://doi.org/10.1021/ac60319a045

Chapter 4

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Wardle, J. and Boniface, D. (2008). Changes in the distributions of body mass index and waist circumference in English adults, 1993/1994 to 2002/2003, International Journal of Obesity, 32, 527–532.
https://doi.org/10.1038/sj.ijo.0803740

Chapter 6

Blackshaw, J., Ewins, M., and Chang, M. (2019). Health Matters: Addressing the food environment as part of a local whole systems approach to obesity, UK Health Security Agency, blog.
https://ukhsa.blog.gov.uk/2019/08/08/health-matters-addressing-the-food-environment-as-part-of-a-local-whole-systems-approach-to-obesity/

Broadberry, S., Campbell, B., Klein, A., Overton, M., and van Leeuwen, B. (2015), British Economic Growth, 1270-1870, Cambridge University Press.

Butland, B., Jebb, S., Kopelman, P., McPherson, K., Thomas, S., Mardell, J., and Parry, V. (2007). Foresight Tackling Obesities: Future Choices – Project report, 2nd edition, Government Office for Science and Department of Health and Social Care.
https://www.gov.uk/government/collections/tackling-obesities-future-choices

Clements, K. W., and Si, J. (2018), Engel's Law, diet diversity, and the quality of food consumption, American Journal of Agricultural Economics, 100(1), 1-22.
DOI: https://doi.org/10.1093/ajae/aax053

Engel, E. (1857). Die Productions und Consumptionsverhaltnisse des Konigreichs Sachsen. [The Production and Consumption Conditions of the Kingdom of Saxony], originally in Zeitschrift des Statistischen Bureaus des Koniglich Sdchsischen Ministerium des Inneren, Nos. 8 and 9, November 22nd, 1857, reprinted in Engel's Die Lebenskosten belgischer Atbeiter-Familien [Cost of Living in Belgian Working-Class Families], Dresden, 1895. [In German].

Frew, E., Ng, S. W ., Coast, J., Hollingsworth, B. P., and Smith, R. (2018). How can economics help tackle obesity? Obesity, 26(7), 1112-1113.
DOI: https://doi.org/10.1002/oby.22211

Houthakker, H. S. (1957). An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law, Econometrica, 25(4), 532-551.
DOI: https://doi.org/10.2307/1905382

Lee, A., Mhurchu, C. N., Sacks, G. et al. (2013). Monitoring the price and affordability of foods and diets globally, Obesity Reviews, 14(S1), 82-95.
DOI: https://doi.org/10.1111/obr.12078

Maes, H. H., Neale, M. C., Eaves, L. J. (1997). Genetic and environmental factors in relative body weight and human adiposity. Behavior Genetics, 27(4), 325-351.
DOI: https://doi.org/10.1023/a:1025635913927

Maruyama, M. (1963). The second cybernetics: deviation amplifying mutual causal processes, American Scientist, 51(2), 164–179.
https://www.jstor.org/stable/27838689

Naukkarinen, J., Rissanen, A., Kaprio, J., Pietiläinen, K. H. (2012). Causes and consequences of obesity: the contribution of recent twin studies, International Journal of Obesity, 36, 1017-1024.
DOI: https://doi.org/10.1038/ijo.2011.192

Pidd, M. (2009). Tools for Thinking; Modelling in Management Science. (3rd ed.) John Wiley and Sons Ltd.

Roser, M. (2013). Web page suggested citation: Max Roser (2013) - "Economic Growth". Published online at OurWorldInData.org
Retrieved from: https://ourworldindata.org/economic-growth [Online Resource]

Roser, M., and Ritchie, H. (2021). Web page suggested citation: Max Roser and Hannah Ritchie (2021) - "Food Prices". Published online at OurWorldInData.org
Retrieved from: https://ourworldindata.org/food-prices [Online Resource]

Roser, M., Ritchie, H., and Rosado, P. (2013). Web page suggested citation: Max Roser, Hannah Ritchie and Pablo Rosado (2013) - "Food Supply". Published online at OurWorldInData.org
Retrieved from: https://ourworldindata.org/food-supply [Online Resource]

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Thomas, R. and Dimsdale, N. (2017). A Millennium of UK Data, Bank of England OBRA dataset.
https://www.bankofengland.co.uk/statistics/research-datasets Spreadsheet link: A millennium of macroeconomic data

Thompson, D. (2012). Food Is Cheap, The Atlantic, online article.
https://www.theatlantic.com/business/archive/2012/04/food-is-cheap/255516/

Vandenbroeck, P., Goossens, J., and Clemens, M. (2007a). Foresight Tackling Obesities: Future Choices — Obesity System Atlas, Government Office for Science and Department of Health and Social Care.
https://www.gov.uk/government/collections/tackling-obesities-future-choices

Vandenbroeck, P., Goossens, J., and Clemens, M. (2007b). Foresight Tackling Obesities: Future Choices – Building the Obesity System Map, Government Office for Science and Department of Health and Social Care.
https://www.gov.uk/government/collections/tackling-obesities-future-choices

Wardle, J. and Boniface, D. (2008). Changes in the distributions of body mass index and waist circumference in English adults, 1993/1994 to 2002/2003, International Journal of Obesity, 32, 527–532.
https://doi.org/10.1038/sj.ijo.0803740

USDoL: U.S. Department of Labor (Elaine L. Chao, Secretary) and U.S. Bureau of Labor Statistics (Kathleen P. Utgoff, Commissioner) (2006), 100 Years of U.S. Consumer Spending: Data for the Nation, New York City, and Boston, Report 991.
https://www.bls.gov/opub/100-years-of-u-s-consumer-spending.pdf

Zimmerman, C. C. (1932). Ernst Engel's Law of expenditures for food, The Quarterly Journal of Economics, 47(1), 78–101.
DOI: https://doi.org/10.2307/1885186

Chapter 7

Cranfield J., Inwood K., and Ross J. A. (2015). Before the obesity epidemic: the body mass index of Canadians in the First and Second World Wars, Canadian Bulletin of Medical History = Bulletin Canadien D'histoire de la Medecine, 32(2), 319-335.
DOI: https://doi.org/10.3138/cbmh.32.2.319

Chapter 13

Hall, K. D., Sacks, G., Chandramohan, D., Chow, C. C., Wang, Y. C., Gortmaker, S. L., and Swinburn, B. A. (2011). Quantification of the effect of energy imbalance on bodyweight, Lancet, 378(9793), 826–837.
DOI: https://doi.org/10.1016/S0140-6736(11)60812-X

Henry, C. J. K. (2005). Basal metabolic rate studies in humans: measurement and development of new equations, Public Health Nutrition, 8(7A), 1133–1152.
DOI: https://doi.org/10.1079/PHN2005801

Marlett, K. L., and Ravussin, E. (2018). Energy expenditure and obesity, in: Handbook of Obesity Treatment, edited by Wadden, T. A., and Bray, G.A., 2nd Edition, The Guilford Press, New York and London, pages 38-63.

Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., Koh, Y.O. (1990). A new predictive equation for resting energy expenditure in healthy individuals. The American Journal of Clinical Nutrition, 51(2), 241-247.
DOI: https://doi.org/10.1093/ajcn/51.2.241

Roser, M., Ritchie, H., and Rosado, P. (2013). Web page suggested citation: Max Roser, Hannah Ritchie and Pablo Rosado (2013) - "Food Supply". Published online at OurWorldInData.org
Retrieved from: https://ourworldindata.org/food-supply [Online Resource]

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Chapter 14

Henry, C. J. K. (2005). Basal metabolic rate studies in humans: measurement and development of new equations, Public Health Nutrition, 8(7A), 1133–1152.
DOI: https://doi.org/10.1079/PHN2005801

Roser, M., Ritchie, H., and Rosado, P. (2013). Web page suggested citation: Max Roser, Hannah Ritchie and Pablo Rosado (2013) - "Food Supply". Published online at OurWorldInData.org
Retrieved from: https://ourworldindata.org/food-supply [Online Resource]

SACN: Scientific Advisory Committee on Nutrition (2012), Dietary Reference Values for Energy, The Stationery Office (TSO), London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 36(8), 1627-1639.
DOI: https://doi.org/10.1021/ac60214a047

Steinier, J., Termonia, Y., and Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 44(11), 1906-1909.
DOI: https://doi.org/10.1021/ac60319a045

Wardle, J. and Boniface, D. (2008). Changes in the distributions of body mass index and waist circumference in English adults, 1993/1994 to 2002/2003, International Journal of Obesity, 32, 527–532.
https://doi.org/10.1038/sj.ijo.0803740

Chapter 15

Schönbeck, Y., Talma, H., van Dommelen, P., Bakker, B., Buitendijk, S. E., HiraSing, R. A., van Buuren, S. (2013). The world's tallest nation has stopped growing taller: the height of Dutch children from 1955 to 2009, Pediatric Research 73(3), 371-377.
DOI: https://doi.org/10.1038/pr.2012.189

Roser, M., Appel, C., and Ritchie, H. (2013): Max Roser, Cameron Appel and Hannah Ritchie (2013) - "Human Height". Published online at OurWorldInData.org. Article "updated in May 2019".
Retrieved from: 'https://ourworldindata.org/human-height' [Online Resource]

Chapter 16

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 36(8), 1627-1639.
DOI: https://doi.org/10.1021/ac60214a047

Steinier, J., Termonia, Y., and Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 44(11), 1906-1909.
DOI: https://doi.org/10.1021/ac60319a045

Health Survey for England (HSE) dataset references

The citations for the latest editions of the HSE datasets for each of the years from 1991 to 2019, as given on the UK Data Service website (home page: https://ukdataservice.ac.uk), are given below. The HSE data for 1991-2018 was used in the work in this report. The HSE data for 2019 was not used, but the accompanying statistics data tables for HSE 2019 are referenced in the report (e.g., in Section 1.2.5).

Office of Population Censuses and Surveys, Social Survey Division. (1997). Health Survey for England, 1991-1992: Combined Data File. [data collection]. 2nd Edition. UK Data Service. SN: 3238.
DOI: http://doi.org/10.5255/UKDA-SN-3238-1

Office of Population Censuses and Surveys, Social Survey Division. (1997). Health Survey for England, 1993. [data collection]. 2nd Edition. UK Data Service. SN: 3316.
DOI: http://doi.org/10.5255/UKDA-SN-3316-1

Joint Health Surveys Unit of Social and Community Planning Research and University College London. (2017). Health Survey for England, 1994. [data collection]. 5th Edition. UK Data Service. SN: 3640.
DOI: http://doi.org/10.5255/UKDA-SN-3640-2

Joint Health Surveys Unit of Social and Community Planning Research and University College London. (2017). Health Survey for England, 1995. [data collection]. 5th Edition. UK Data Service. SN: 3796.
DOI: http://doi.org/10.5255/UKDA-SN-3796-2

Joint Health Surveys Unit of Social and Community Planning Research and University College London. (2017). Health Survey for England, 1996. [data collection]. 5th Edition. UK Data Service. SN: 3886. DOI: http://doi.org/10.5255/UKDA-SN-3886-2

Joint Health Surveys Unit of Social and Community Planning Research and University College London. (2017). Health Survey for England, 1997. [data collection]. 4th Edition. UK Data Service. SN: 3979. DOI: http://doi.org/10.5255/UKDA-SN-3979-2

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 1998. [data collection]. 5th Edition. UK Data Service. SN: 4150. DOI: http://doi.org/10.5255/UKDA-SN-4150-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 1999. [data collection]. 4th Edition. UK Data Service. SN: 4365. DOI: http://doi.org/10.5255/UKDA-SN-4365-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2011). Health Survey for England, 2000. [data collection]. 4th Edition. UK Data Service. SN: 4487, DOI: http://doi.org/10.5255/UKDA-SN-4487-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2001. [data collection]. 3rd Edition. UK Data Service. SN: 4628.
DOI: http://doi.org/10.5255/UKDA-SN-4628-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2002. [data collection]. 2nd Edition. UK Data Service. SN: 4912.
DOI: http://doi.org/10.5255/UKDA-SN-4912-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2003. [data collection]. 2nd Edition. UK Data Service. SN: 5098.
DOI: http://doi.org/10.5255/UKDA-SN-5098-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2004. [data collection]. 2nd Edition. UK Data Service. SN: 5439.
DOI: http://doi.org/10.5255/UKDA-SN-5439-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2011). Health Survey for England, 2005. [data collection]. 3rd Edition. UK Data Service. SN: 5675.
DOI: http://doi.org/10.5255/UKDA-SN-5675-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2011). Health Survey for England, 2006. [data collection]. 4th Edition. UK Data Service. SN: 5809. DOI: http://doi.org/10.5255/UKDA-SN-5809-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2007. [data collection]. 2nd Edition. UK Data Service. SN: 6112 DOI: http://doi.org/10.5255/UKDA-SN-6112-1

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2013). Health Survey for England, 2008. [data collection]. 4th Edition. UK Data Service. SN: 6397. DOI: http://doi.org/10.5255/UKDA-SN-6397-2

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. (2015). Health Survey for England, 2009. [data collection]. 3rd Edition. UK Data Service. SN: 6732.
DOI: http://doi.org/10.5255/UKDA-SN-6732-2

NatCen Social Research, Royal Free and University College Medical School. Department of Epidemiology and Public Health. (2015). Health Survey for England, 2010. [data collection]. 3rd Edition. UK Data Service. SN: 6986.
DOI: http://doi.org/10.5255/UKDA-SN-6986-3

NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2013). Health Survey for England, 2011. [data collection]. UK Data Service. SN: 7260.
DOI: http://doi.org/10.5255/UKDA-SN-7260-1

NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2014). Health Survey for England, 2012. [data collection]. UK Data Service. SN: 7480.
DOI: http://doi.org/10.5255/UKDA-SN-7480-1

NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2015). Health Survey for England, 2013. [data collection]. UK Data Service. SN: 7649.
DOI: http://doi.org/10.5255/UKDA-SN-7649-1

NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2018). Health Survey for England, 2014. [data collection]. 3rd Edition. UK Data Service. SN: 7919.
DOI: http://doi.org/10.5255/UKDA-SN-7919-3

NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2019). Health Survey for England, 2015. [data collection]. 2nd Edition. UK Data Service. SN: 8280.
DOI: http://doi.org/10.5255/UKDA-SN-8280-2

NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2019). Health Survey for England, 2016. [data collection]. 3rd Edition. UK Data Service. SN: 8334.
DOI: http://doi.org/10.5255/UKDA-SN-8334-3

University College London, Department of Epidemiology and Public Health, National Centre for Social Research (NatCen). (2019). Health Survey for England, 2017. [data collection]. 2nd Edition. UK Data Service. SN: 8488.
DOI: http://doi.org/10.5255/UKDA-SN-8488-2

NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2022). Health Survey for England, 2018. [data collection]. 2nd Edition. UK Data Service. SN: 8649.
DOI: http://doi.org/10.5255/UKDA-SN-8649-2

NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2021). Health Survey for England, 2019. [data collection]. UK Data Service. SN: 8860.
DOI: http://doi.org/10.5255/UKDA-SN-8860-1