Risks of COVID-19 by occupation in NHS workers in England

van der Plaat DA[1], Madan I[2,3]*, Coggon D[4], van Tongeren M[5], Edge R[6], Muiry R[2], Parsons V[2,3], Cullinan P[1]

1. National Heart and Lung Institute, Imperial College, London, SW3 6LY UK
2. Occupational Health Service, Guy's and St Thomas NHS Foundation Trust, St Thomas Hospital, Westminster Bridge Rd, London SE1 7NJ UK
3. Faculty of Life Sciences and Medicine, King's College London, London SE1 7EH UK
4. MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton UK
5. University of Manchester, Centre for Occupational and Environmental Health, School of Health Sciences, Manchester, M13 9PL UK
6. Lancaster University, Lancaster Medical School Bailrigg Lancaster, LA1 4YW, UK

*corresponding author

Ira Madan

Occupational Health Service

St Thomas Hospital

Guy's and St Thomas' NHS Foundation Trust

Westminster Bridge Rd London, SE1 7NJ UK

Email: Ira.madan@kcl.ac.uk

Key messages

1. What is already known about this subject?

Healthcare workers and other keyworkers (workers whose job was considered essential to societal functioning) had a higher likelihood of testing positive for COVID-19 than other workers during the first lockdown in England. Amongst healthcare workers, those working in inpatient settings had the highest rate of infection.

2. What are the new findings?

Between March and July 2020, the overall risk of COVID-19 sickness absence in National Health Service staff in England was lower at older ages, higher in non-white staff, and (in comparison with administrative and clerical staff) more than doubled in registered nurses and among workers such as healthcare assistants providing support to health professionals. Risk in health care scientists was little different from that in administrative and clerical occupations

3. How might this impact on policy or clinical practice in the foreseeable future?

Our results suggest that the risk reduction strategies that were in place for healthcare scientists were adequate. However, the protection for nursing and supporting health professionals was insufficient. In the event of a further 'wave' of infections resulting in high hospital admissions, attention should be paid to ensuring that risk reduction strategies for nurses and supporting health professionals are improved.

Abstract

Objective

To quantify occupational risks of Covid-19 among healthcare staff during the first wave (9.3.2020–31.7.2020) of the pandemic in England

Methods

We used pseudonymised data on 902,813 individuals employed by 191 National Health Service trusts to explore demographic and occupational risk factors for sickness absence ascribed to Covid-19 (n = 92,880). We estimated odds ratios (ORs) by multivariable logistic regression.

Results

With adjustment for employing trust, demographic characteristics, and previous frequency of sickness absence, risk relative to administrative/clerical occupations was highest in “additional clinical services” (care assistants and other occupations directly supporting those in clinical roles) (OR 2.31 [2.25-2.37]), registered nursing and midwifery professionals (OR 2.28 [2.23-2.34]) and allied health professionals (OR 1.94 [1.88-2.01]), and intermediate in doctors and dentists (OR 1.55 [1.50-1.61]). Differences in risk were higher after the employing trust had started to care for documented Covid-19 patients, and were reduced, but not eliminated, following additional adjustment for exposure to infected patients or materials, assessed by a job-exposure matrix. For prolonged Covid-19 sickness absence (episodes lasting >14 days), the variation in risk by staff group was somewhat greater.

Conclusions

After allowance for possible bias and confounding by non-occupational exposures, we estimated that relative risks for Covid-19 among most patient-facing occupations were between 1.5 and 2.5. The highest risks were in those working in additional clinical services, nursing and midwifery and in allied health professions. Better protective measures for these staff groups should be a priority. Covid-19 may meet criteria for compensation as an occupational disease in some healthcare occupations.

Keywords: Covid-19, occupation, risk, healthcare workers, compensation

Introduction

Covid-19, like many communicable diseases, poses an occupational hazard to healthcare workers. When the first wave of the pandemic hit the UK early in March 2020, precautions were implemented to reduce transmission to healthcare staff, including identification and segregation of infected patients, and the use of personal protective equipment (PPE). In the early weeks, however, these measures were far from ideal. Adequate PPE often was in short supply and much of the UK's pandemic stockpile contained equipment suitable for an influenza outbreak, but not for more infectious diseases (1). Additionally, a lack of capacity meant that testing of patients who might be carrying SARS-CoV-2 was insufficient (2). Cases of occupationally-acquired disease were therefore to be expected. However, the level of risk has been uncertain, as has the extent to which it varied between different healthcare occupations. Better understanding would help in prioritisation of preventive strategies during further waves of the pandemic, and in the management of similar infectious diseases. It is also needed to inform decisions on possible compensation for Covid-19 as an occupational disease in healthcare workers.

Evidence to date has indicated that in England and Wales, male healthcare workers (taken as a group), nurses, nursing assistants and auxiliaries of both sexes have had higher age-adjusted mortality from Covid-19 than the general population (3). Several studies have found that patient-facing healthcare workers were infected with COVID-19 at substantially higher rates than non-healthcare workers during the first wave (March–July 2020), although they differ in their findings as to which groups were at greatest risk (4-6). Mortality, however, depends not only on risk of contracting Covid-19, but also on personal vulnerability when infection occurs, which may vary importantly between occupations. Furthermore, differences in the incidence of infection by occupation may be driven not only by exposures in the workplace (through proximity to infected colleagues as well as contact with patients and infected materials), but also away from work. For example, rates of infection have been higher among people living in large, crowded households (7)

To get further insight regarding occupational risks of Covid-19 in healthcare workers, we analysed data on sickness absence among employees of National Health Service (NHS) trusts in England, before and after they started to care for patients known to have the disease.

Methods

With approval by the NHS Health Research Authority (reference 20/SC/0282), we were allowed access to two pseudonymised databases prepared by the NHS Electronic Staff Record (ESR) Central Team. They contained information on demographic and occupational characteristics of all staff continuously employed by NHS trusts (organisational units serving a geographical area or specialised function) in England from 01.01.2019 to 31.07.2020, and on all their absences from work during that period, other than for annual leave. The latter included the reason for absence, and the start and end date of each episode.

Supplementary File A describes the methods by which we used the two databases to create a file for statistical analysis. We first checked for missing and inconsistent data, and corrected clear anomalies in a small minority of records by imputation according to a standard set of rules. We also reclassified some variables into aggregated categories that would facilitate more meaningful analysis. We then generated a file with one record for each individual, which included the variables listed in Table 1, and also the start and end dates of all absences during 01.01.2019 to 31.07.2020, with the reason for absence.

Table 1. Distribution of risk factors in study sample and cumulative prevalence of new Covid-19 sickness absence during 9 March to 31 July 2020

Risk factor	Frequency of risk factor in study sample		Any Covid-19 sickness absence starting 9 March to 31 July 2020		Any prolonged* Covid-19 sickness absence starting 9 March to 16 July 2020	
	N	(%)[a]	N	(%)[b]	N	(%)[b]
Sex						
Female	696,357	77.1	72,420	10.4	16,413	2.4
Male	206,456	22.9	20,460	9.9	4,575	2.2
Age (years)						
<30	109,277	12.1	12,941	11.8	1,587	1.5
30-34	107,563	11.9	12,043	11.2	1,954	1.8
35-39	103,499	11.5	10,830	10.5	2,111	2.0
40-44	113,523	12.6	12,421	10.9	2,752	2.4
45-49	125,802	13.9	13,863	11.0	3,572	2.8
50-54	133,721	14.8	13,403	10.0	3,686	2.8
55-60	133,908	14.8	11,732	8.8	3,375	2.5
>60	75,520	8.4	5,647	7.5	1,951	2.6
Ethnicity						
White	687,174	76.1	63,630	9.3	12,495	1.8
South Asian	61,861	6.9	8,033	13.0	2,376	3.8
Other or unspecified Asian	37,956	4.2	7,402	19.5	2,273	6.0
Black	54,267	6.0	6,705	12.4	2,054	3.8
Mixed	16,026	1.8	1,838	11.5	385	2.4
Other	12,937	1.4	1,982	15.3	622	4.8

Risk factor	Frequency of risk factor in study sample		Any Covid-19 sickness absence starting 9 March to 31 July 2020		Any prolonged* Covid-19 sickness absence starting 9 March to 16 July 2020	
	N	(%)[a]	N	(%)[b]	N	(%)[b]
Unknown	32,592	3.6	3,290	10.1	783	2.4
Episodes of sickness absence in 2019						
0	302,258	33.5	20,840	6.9	4,361	1.4
1	234,871	26.0	22,805	9.7	5,242	2.2
2-3	267,763	29.7	33,707	12.6	7,972	3.0
>3	97,921	10.8	15,528	15.9	3,413	3.5
Staff group at 9 March 2020						
Administrative and clerical	193,983	21.5	11,236	5.8	2,340	1.2
Additional clinical services	176,558	19.6	23,967	13.6	6,148	3.5
Additional professional scientific and technical	40,874	4.5	2,960	7.2	509	1.2
Allied health professionals	67,067	7.4	7,584	11.3	1,192	1.8
Estates and ancillary	58,313	6.5	4,684	8.0	1,212	2.1
Healthcare scientists	20,657	2.3	1,492	7.2	241	1.2
Medical and dental	76,184	8.4	6,203	8.1	1,061	1.4
Nursing and midwifery registered	265,486	29.4	34,390	13.0	8,232	3.1
Students	1,797	0.2	201	11.2	24	1.3
Multiple or unknown	1,894	0.2	163	8.6	29	1.5
Exposure category at 9 March 2020**						
Care of patients much more likely to have Covid-19 than general population	64,977	7.2	9,004	13.9	1,514	2.3
Care for patients who may be more likely to have Covid-19 than general population	292,692	32.4	41,808	14.3	10,651	3.6
Care of patients with similar or lower prevalence of Covid-19 than general population	250,863	27.8	22,262	8.9	4,682	1.9
No patient care but often in areas where patients have higher prevalence of Covid-19 than general population	554	0.1	39	7.0	11	2.0
No patient care but often in areas where patients have similar or lower prevalence of Covid-19 than general population	16,295	1.8	1,686	10.3	246	1.5
No patient care, occasionally in patient areas	155,963	17.3	10,944	7.0	2,538	1.6
Unlikely to be in patient areas, but work with material potentially contaminated by coronavirus	23,659	2.6	1,976	8.4	344	1.5
Other or unknown	97,810	10.8	5,161	5.3	1,002	1.0

[a]prevalence % in study sample (total N = 902,813). [b]prevalence % among those with risk factor. *prolonged COVID sickness absence defined as episodes lasting >14 days. **exposure categories as specified in job-exposure matrix. For more information see Supplementary file A.

Staff group was assigned to 10 categories, according to a classification used in the ESR records (Supplementary File A – Table A1). The ESR system also held more detailed occupational data, but to protect privacy, that could not be released. Instead, four members of the team (an occupational hygienist and three occupational physicians with experience in the NHS) compiled a job-exposure matrix (JEM), which the ESR Management Team then used to reclassify detailed occupational categories (n=659) to the eight exposure categories listed in Table 1.

Within the ESR database, reasons for absence (of any type) were described by four variables (Supplementary File A). The 192 different combinations were collapsed into 60 categories, of which 32 were related to sickness absence. Using the information on absence episodes, we defined a variable which for each individual represented the number of new episodes of sickness absence (for any cause) that had started during 2019 (classified as 0, 1, 2-3 and >3). This was intended as a marker for long-term propensity to take sickness absence, which can vary importantly between individuals independently of morbidity (8). In addition, we distinguished episodes of Covid-19 sickness absence, which we defined as being for any of five categories of sickness (cough/flu, chest/respiratory, infectious diseases, other or unknown) with Covid-19 recorded as a related reason. Such episodes were classed as prolonged if their duration exceeded 14 days (see Supplementary File B).

Data on the date by which each trust was known to have admitted at least three Covid-19 cases were obtained from an NHS COVID-19 daily situation report published on 12.11.2020 (9). We took 09.03.2020 as the date from which Covid-19 sickness absence could reasonably be assumed to reflect coronavirus infection. That was at least 10 days before most hospitals started to admit documented Covid-19 cases (see Supplementary File B for further justification).

Two collaborating trusts provided data on antibody tests that had been carried out on staff members before 07.08.2020. Individuals were identified by an encrypted code number that had been assigned by the ESR Management Team, allowing anonymised linkage with the other records to which we had access.

Statistical analysis

Statistical analysis was carried out with R (version 4.0.4) software. We first generated descriptive statistics summarising the distributions of the main variables. We then fitted two multivariable logistic regression models to estimate odds ratios (ORs) with 95% confidence intervals (95%CIs) for the start of any episode of Covid-19 sickness absence from 09.03.2020 to 31.07.2020. Model 1 included sex, age group, ethnicity, episodes of sickness absence in 2019 and staff group, while in model 2 the exposure category variable was additionally included to help understand the extent to which associations with staff group reflected patient-related exposures.

Next, the analysis was repeated, distinguishing between onset of the Covid-19 sickness absence before and after the employing trust had first cared for at least three documented Covid-19 cases. Our aim was to distinguish periods when acquisition of Covid-19 through transmission from patients was less and more likely; we incorporated a lag of four days to allow for an interval between exposure to infection and development of symptoms.

Further logistic regression models were used to explore risk factors for prolonged Covid-19 sickness absence starting during 09.03.2020 16.07.2020 (because records were complete only up to 31.07.2020, we could not be confident of accurately distinguishing prolonged episodes that started after 16.07.2020).

Finally, to check on the reliability of Covid-19 sickness absence as a marker for the disease, we used data from two collaborating trusts to compare the prevalence of positive antibody tests in employees who underwent testing before 07.08.2020, according to their history of Covid-19 sickness absence.

In sensitivity analyses, we excluded individuals in whom one or more of age, sex or ethnicity was imputed because of inconsistencies, those with multiple jobs or whose job changed over the study period, and those with a missing or imputed end date of an absence.

Results

After exclusion of 21,775 employees who were absent from work continuously from 09.03.2020 to 31.07.2020 (mainly because of maternity or study leave), and 56,543 at nine trusts which never coded whether sickness absence was related to Covid-19, analysis was based on 902,813 individuals (77% female) from 191 trusts. Most (89%) were aged between 25 and 60 years, and 76% were of white ethnicity. A total of 92,880 (10%) had one or more episodes of Covid-19 sickness absence during the study period, including 20,988 (2.3%) in whom at least one episode was prolonged. Table 1 gives further information about the distribution of risk factors in the study sample, and the cumulative prevalence of Covid-19 sickness absence over the study period, according to those risk factors.

Table 2 shows associations of Covid-19 sickness absence at any time during the study period with the main risk factors of interest. After adjustment for other covariates, risk was similar in men and women, and in age groups below 55 years, but lower at older ages (OR for age >60 relative to <30 years in fully adjusted model: 0.76). Risk was generally higher for non-white relative to white ethnicity, and particularly for those of Asian origin (ORs 1.43 and 1.73 in fully adjusted model). Frequency of sickness absence during 2019 was a further risk factor, with an OR of 2.41 for >3 relative to 0 episodes in the fully adjusted model.

Table 2. Associations of risk factors at baseline with start of any Covid-19 sickness absence during 9 March to 31 July 2020

Risk estimates were derived from two logistic regression models that included all of the variables for which results are presented, together with trust (191 categories).

Risk factor	Model 1		Model 2	
	OR	(95% CI)	OR	(95% CI)
Sex				
Female	*ref.*	*ref.*	*ref.*	*ref.*
Male	1.01	0.99 - 1.03	1.02	1 - 1.03
Age (years)				
<30	*ref.*	*ref.*	*ref.*	*ref.*
30-34	0.96	0.94 - 0.99	0.97	0.94 - 0.99
35-39	0.97	0.95 - 1.00	0.98	0.96 - 1.01
40-44	0.99	0.96 - 1.02	1.00	0.97 - 1.02
45-49	1.00	0.98 - 1.03	1.00	0.98 - 1.03
50-54	0.98	0.95 - 1.00	0.98	0.95 - 1.00
55-60	0.89	0.86 - 0.91	0.89	0.86 - 0.91
>60	0.76	0.74 - 0.79	0.76	0.73 - 0.79
Ethnicity				
White	*ref.*	*ref.*	*ref.*	*ref.*
South Asian	1.43	1.4 - 1.47	1.41	1.37 - 1.45

Risk factor	Model 1		Model 2	
	OR	(95% CI)	OR	(95% CI)
Other or unspecified Asian	1.73	1.67 - 1.78	1.65	1.60 - 1.70
Black	1.15	1.12 - 1.19	1.14	1.10 - 1.17
Mixed	1.14	1.08 - 1.20	1.13	1.08 - 1.19
Other	1.48	1.41 - 1.56	1.44	1.37 - 1.51
Unknown	1.07	1.03 - 1.11	1.07	1.03 - 1.11
Episodes of sickness absence in 2019				
0	*ref.*	*ref.*	*ref.*	*ref.*
1	1.39	1.37 - 1.42	1.38	1.36 - 1.41
2-3	1.83	1.79 - 1.86	1.80	1.77 - 1.84
>3	2.41	2.36 - 2.47	2.38	2.32 - 2.43
Staff group at 9 March 2020				
Administrative and clerical	*ref.*	*ref.*	*ref.*	*ref.*
Additional clinical services	2.31	2.25 - 2.37	1.63	1.55 - 1.72
Additional professional scientific and technical	1.37	1.31 - 1.43	1.05	0.98 - 1.12
Allied health professionals	1.94	1.88 - 2.01	1.33	1.25 - 1.41
Estates and ancillary	1.45	1.39 - 1.50	1.30	1.25 - 1.35
Healthcare scientists	1.17	1.10 - 1.24	1.03	0.95 - 1.11
Medical and dental	1.55	1.50 - 1.61	1.09	1.03 - 1.15
Nursing and midwifery registered	2.28	2.23 - 2.34	1.57	1.49 - 1.65
Students	1.87	1.60 - 2.20	1.35	1.14 - 1.59
Multiple or unknown	1.62	1.37 - 1.92	1.17	0.98 - 1.39
Exposure category at 9 March 2020*				
Care of patients much more likely to have Covid-19 than general population	-	-	1.48	1.40 - 1.57
Care for patients who may be more likely to have Covid-19 than general population	-	-	1.43	1.36 - 1.51
Care of patients with similar or lower prevalence of Covid-19 than general population	-	-	1.06	1.01 - 1.12
No patient care but often in areas where patients have higher prevalence of Covid-19 than general population	-	-	0.72	0.52 - 1.01
No patient care but often in areas where patients have similar or lower prevalence of Covid-19 than general population	-	-	1.28	1.18 - 1.38
No patient care, occasionally in patient areas	-	-	*ref.*	*ref.*
Unlikely to be in patient areas, but work with material potentially contaminated by coronavirus	-	-	0.92	0.85 - 0.99
Other or unknown	-	-	0.73	0.71 - 0.76

*exposure categories are from the JEM. For more information see Supplementary file A.

With no adjustment for exposure category, ORs varied more than twofold across the ten staff groups, the lowest risk being in administrative and clerical jobs (the reference for other risk estimates), and the highest in additional clinical services (OR 2.31), registered nursing and midwifery professionals (OR 2.28), allied health professionals (OR 1.94) and students (OR 1.87). Risk in doctors and dentists was intermediate (OR 1.55), while that in health care scientists was little different from administrative and clerical occupations (OR 1.17).

Exposure category showed an expected gradient of risk, with the highest ORs (relative to no patient care and only occasionally in patient areas) for hands-on or face-to-face care of patients likely to have a higher prevalence of Covid-19 than the general population (ORs 1.48 and 1.43). After adjustment for exposure category, the risk estimates for other staff groups relative to administrative and clerical jobs were all reduced. However, Covid-19 sickness absence was still notably more frequent among those working in additional clinical services (OR 1.63) and in registered nursing and midwifery professionals (OR 1.57). Re-analysis excluding individuals with imputed or missing data gave similar results (Supplementary File C – Table C1).

Most (75%) of the 191 trusts had cared for at least three documented Covid-19 patients by 12 April 2020, but 25 (12.5%) had still not done so by 31.07.2020. The latter were mainly mental health and specialist (e.g. orthopaedic) trusts. Before trusts had cared for three documented Covid-19 patients, ORs for Covid-19 sickness absence relative to administrative and clerical workers were highest in additional clinical services (1.85), registered nurses and midwives (1.81), doctors and dentists (1.66) and allied health professionals (1.62) (Table 3). After trusts had started to care for Covid-19 patients, the ranking of risks by staff group was broadly similar, but the divergence of ORs was greater (2.71 for additional clinical services and 2.70 for registered nurses and midwives). For doctors and dentists, the OR was somewhat reduced (1.45).

Table 3 Associations of staff group with a first episode of Covid-19 sickness absence during 9 March to 31 July 2020, according to whether the employing trust had yet cared for at least three documented Covid-19 patients

Risk estimates were derived from two logistic regression models, each of which included all of the variables from Model 1 in Table 2.

Staff group at 9 March 2020	Before trust had cared for 3 Covid-19 cases[a]					After trust had cared for ≥3 Covid-19 cases[a]			
	Number at risk	Number of cases	OR	(95%CI)		Number at risk	Number of cases	OR	(95%CI)
Administrative and clerical	193,983	6,086	*ref.*	*ref.*		187,897	5,150	*ref.*	*ref.*
Additional clinical services	176,558	10,742	1.85	1.79 - 1.91		165,816	13,225	2.71	2.62 - 2.80
Additional professional scientific and technical	40,874	1,586	1.20	1.13 - 1.27		39,288	1,374	1.53	1.44 - 1.63
Allied health professionals	67,067	2,995	1.62	1.55 - 1.70		64,072	4,589	2.25	2.15 - 2.35
Estates and ancillary	58,313	2,094	1.21	1.15 - 1.27		56,219	2,590	1.68	1.60 - 1.77
Healthcare scientists	20,657	786	1.24	1.15 - 1.34		19,871	706	1.12	1.03 - 1.22
Medical and dental	76,184	3,459	1.66	1.58 - 1.74		72,725	2,744	1.45	1.38 - 1.53
Nursing and midwifery registered	265,486	15,785	1.81	1.76 - 1.87		249,701	18,605	2.70	2.61 - 2.79
Students	1,797	37	0.83	0.60 - 1.16		1,760	164	2.76	2.30 - 3.30
Multiple or unknown	1,894	89	1.56	1.25 - 1.94		1,805	74	1.65	1.30 - 2.10

[a]with a lag of four days to allow for the interval between exposure to infection and development of symptoms (see text)

Table 4 presents an analysis similar to that in Table 2, but with prolonged Covid-19 sickness absence as the outcome. Individuals with only shorter durations of Covid19 sickness absence were excluded, and risk estimates are relative to no Covid-19 sickness absence. Notable differences from the findings for all Covid-19 sickness absence were a progressive increase in risk across age bands (OR for age >60 vs.<30 years 2.15 in fully adjusted model), higher ORs for non-white vs. white ethnicity, higher risk estimates for additional clinical services and registered nurses and midwives (ORs of 2.88 and 2.59 respectively, reducing to 1.88 and 1.60 after adjustment for exposure category), and lower risk estimates for medical and dental staff (ORs 1.10 and 0.77 before and after adjustment for exposure category).

Table 4. Associations of risk factors at baseline with start of any episode of prolonged Covid-19 sickness absence during 9 March to 16 July 2020

Risk estimates were derived from two logistic regression models that included all of the variables for which results are presented, together with trust (191 categories). An episode of Covid-19 sickness absence was classed as prolonged if it lasted >14 days. Individuals who had only short-term Covid-19 sickness absence were excluded from these analyses (see text).

Risk factor	Model 1		Model 2	
	OR[a]	(95% CI)	OR[a]	(95% CI)
Sex				
Female	*ref.*	*ref.*	*ref.*	*ref.*
Male	1.03	0.99 - 1.06	1.04	1.00 - 1.08
Age (years)				
<30	*ref.*	*ref.*	*ref.*	*ref.*
30-34	1.25	1.17 - 1.34	1.25	1.17 - 1.34
35-39	1.56	1.45 - 1.66	1.56	1.46 - 1.67
40-44	1.73	1.63 - 1.85	1.72	1.62 - 1.84
45-49	2.01	1.89 - 2.14	1.99	1.87 - 2.11
50-54	2.18	2.05 - 2.32	2.17	2.04 - 2.30
55-60	2.12	1.99 - 2.25	2.10	1.97 - 2.23
>60	2.19	2.04 - 2.35	2.15	2.01 - 2.31
Ethnicity				
White	*ref.*	*ref.*	*ref.*	*ref.*
South Asian	2.54	2.42 - 2.67	2.47	2.35 - 2.59
Other or unspecified Asian	2.90	2.75 - 3.05	2.68	2.54 - 2.82
Black	1.72	1.63 - 1.81	1.68	1.59 - 1.77
Mixed	1.38	1.24 - 1.54	1.36	1.23 - 1.52
Other	2.41	2.21 - 2.62	2.27	2.08 - 2.48
Unknown	1.28	1.18 - 1.38	1.28	1.18 - 1.38
Episodes of sickness absence in 2019				
0	*ref.*	*ref.*	*ref.*	*ref.*
1	1.48	1.42 - 1.55	1.47	1.41 - 1.53
2-3	2.01	1.93 - 2.09	1.98	1.91 - 2.06
>3	2.59	2.47 - 2.72	2.53	2.41 - 2.66
Staff group at 9 March 2020				
Administrative and clerical	*ref.*	*ref.*	*ref.*	*ref.*
Additional clinical services	2.88	2.74 - 3.02	1.88	1.70 - 2.09
Additional professional scientific and technical	1.19	1.08 - 1.31	1.01	0.88 - 1.16
Allied health professionals	1.73	1.61 - 1.86	1.14	1.01 - 1.28
Estates and ancillary	1.59	1.48 - 1.71	1.41	1.30 - 1.52
Healthcare scientists	0.94	0.82 - 1.08	0.90	0.76 - 1.06
Medical and dental	1.10	1.02 - 1.19	0.77	0.68 - 0.87
Nursing and midwifery registered	2.59	2.47 - 2.71	1.60	1.44 - 1.78
Students	2.00	1.33 - 3.03	1.47	0.96 - 2.25
Multiple or unknown	1.45	1.00 - 2.11	0.98	0.67 - 1.44
Exposure category at 9 March 2020*				
Care of patients much more likely to have Covid-19 than general population	-	-	1.41	1.25 - 1.59

Risk factor	Model 1		Model 2	
	OR[a]	(95% CI)	OR[a]	(95% CI)
Care for patients who may be more likely to have Covid-19 than general population	-	-	1.65	1.49 - 1.83
Care of patients with similar or lower prevalence of Covid-19 than general population	-	-	1.02	0.92 - 1.14
No patient care but often in areas where patients have higher prevalence of Covid-19 than general population	-	-	1.01	0.55 - 1.85
No patient care but often in areas where patients have similar or lower prevalence of Covid-19 than general population	-	-	0.94	0.79 - 1.12
No patient care, occasionally in patient areas	-	-	*ref.*	*ref.*
Unlikely to be in patient areas, but work with material potentially contaminated by coronavirus	-	-	0.77	0.66 - 0.89
Other or unknown	-	-	0.68	0.63 - 0.74

[a]odds ratio relative to no new Covid-19 sickness absence during study period. *exposure categories are based on the constructed JEM. For more information see Supplementary file A.

At the two collaborating trusts, results from antibody tests performed by 7 August 2020 were available for 11,050 staff members. The overall prevalence of positive results among those who had taken Covid-19 sickness absence (37.0%) was 3.3 times that in those who had not (11.1%). There were no differences in this ratio by staff group that could not easily be attributable to random sampling variation (Table 5).

Table 5. Results of antibody tests at two trusts according to risk factors
Antibody test results, prior to 7 August 2020, were provided by Cambridge University Hospitals NHS Foundation Trust and Guys and St Thomas's Trust.

Risk factor	No Covid-19 sickness absence				Covid-19 sickness absence				Ratio of proportions with at least one positive test[c]
	At least one test performed		At least one test positive		At least one test performed		At least one test positive		
	N	(%)[a]	N	(%)[b]	N	(%)[a]	N	(%)[b]	
All employees	9,502	54.8	1,053	11.1	1,548	64.6	573	37.0	3.3
Staff group at 9 March 2020									
Administrative and clerical	2,120	49.6	205	9.7	198	55.8	72	36.4	3.8
Additional clinical services	981	51.6	126	12.8	175	57.8	66	37.7	2.9
Additional professional scientific and technical	479	66.3	45	9.4	51	68.0	13	25.5	2.7
Allied health professionals	725	63.3	73	10.1	137	76.5	44	32.1	3.2
Estates and ancillary	537	46.2	135	25.1	112	57.4	58	51.8	2.1
Healthcare scientists	357	55.9	18	5.0	32	66.7	11	34.4	6.8
Medical and dental	1,076	54.4	78	7.2	129	55.6	44	34.1	4.7
Nursing and midwifery registered	3,202	58.5	370	11.6	712	70.7	265	37.2	3.2
Students	7	87.5	3	42.9	0	0	0	0	0
Multiple or unknown	18	72.0	0	0	2	100	0	0	-

[a]prevalence % of having at least one test.
[b]prevalence % among those tested
[c]proportion among those with Covid-19 sickness absence / proportion among those with no Covid-19 sickness absence

Discussion

After allowance for employing trust, demographic characteristics, and previous frequency of sickness absence, we found more than twofold variation in the risk of Covid-19 sickness absence across major NHS staff groups in England. Differences were reduced, but not eliminated, following adjustment for potential exposure to infected patients or materials, assessed by a JEM. For prolonged Covid-19 sickness absence, the variation in risk was greater.

The analysis benefitted from a large sample size, giving high statistical power, and from its use of data collected prospectively in a standardised format. Information about employing trust, sex, age, staff group and frequency of earlier sickness absence should all have been highly reliable, and we would not expect serious misclassification between the specified categories of ethnicity. A limitation was that staff group distinguished only broad categories of work. Ideally, analysis would have discriminated between occupations in finer detail, but access to that level of information was precluded by data protection rules. We therefore constructed a JEM to group the 659 occupations in the ESR database to eight exposure categories.

As an indicator of occupational exposure to infection from patients, the JEM should have been superior to staff group. For example, within medical and dental personnel, it distinguished specialists in intensive care, expected to have high exposure to patients with Covid-19, from orthopaedic surgeons, whose patients would be expected to have lower prevalence of the disease. However, even in the detailed occupational classification to which the JEM was applied, some job categories were heterogeneous (e.g. nurses in medical wards could not be distinguished from those working in surgery). Moreover, it did not allow for changes in duties during the epidemic, or for use of PPE and its effectiveness (including possible changes over time as a consequence of modified protection policies). In early April 2020, workers with a long-term condition such as asthma, were advised by Government that they should ‘shield’ and either work from home or not work at all. The health-related characteristics that prompted advice to shield are associated with higher risk of severe outcomes (vulnerability) should an individual contract Covid-19, but not with a higher risk of contracting infection. To bias associations of staff group with sickness absence for Covid-19 importantly, shielding would need to have been substantially more prevalent in some occupational groups than others. This seems unlikely, but if anything, redeployment out of patient-facing roles would be expected to reduce risk estimates for patient-facing occupations.

The varying specificity of occupational categories in the JEM complicates interpretation of numerical estimates of risk for exposure levels. Also, the heterogeneous mix of occupations in

individual exposure categories, makes it harder to assess the potential for confounding by non-occupational exposures. For these reasons, we focused principally on risk by staff group (a well-established classification of jobs), and used exposure category to help understand the extent to which associations with staff group reflected patient-related exposures. .

The other major limitation was the incomplete validity of sickness absence as a marker for Covid-19. Early in the epidemic, diagnostic tests were not widely available, and clinical diagnoses may not have been accurate. Nevertheless, at the trusts which provided data, antibody tests were more than three times as likely to be positive among individuals who had taken Covid-19 sickness absence.

In assessing relative risks by staff group, we adjusted for demographic variables, for trust and for frequency of sickness absence in 2019. The latter was intended as a marker of individual propensity to take sickness absence when ill, and showed an expected association with Covid-19 sickness absence. Adjustment for trust was important because rates of infection were known to have varied geographically (10). Moreover, there may have been systematic differences between trusts in the ascertainment and coding of reasons for absence.

In all analyses, we took administrative and clerical workers as the reference for risks in other staff groups. Making up 21.5% of the study sample, they encompassed a range of occupations, including senior managers as well as middle-grade administrative occupations, clerical workers and receptionists. Most will have been office-based, with little or no direct patient contact, and during the epidemic, some may have worked partially or totally from home. Their work may have entailed social contact with colleagues, but not at a level higher than in many occupations outside healthcare. Furthermore, their socio-economic circumstances will have been neither exceptionally good nor poor. Thus, within the demographic strata that we distinguished, their exposures to SARS-CoV-2 should have been representative of the wider working population in their local area.

An indication of differences in risk between staff groups for reasons other than patient-care comes from analysis restricted to the period before each trust began to care for documented Covid-19 cases (Table 3). During that phase, much of the observed variation in risk might be expected to reflect exposure to infection away from work, or through proximity to infected colleagues. However, the highest ORs (between 1.6 and 1.9) were all in patient-facing occupations, suggesting that there may also have been some unrecognised contact with infected patients.

Once trusts were known to be caring for Covid-19 patients, the ORs for most of these occupations were higher, excess relative risks (estimated as OR-1) increasing by 0.6-0.9 (Table 3). An exception were doctors and dentists, in whom ORs were lower when trusts were known to be caring for Covid-19 patients. This may have been because in the early phase of the epidemic, some doctors contracted infection from undiagnosed patients, but that risk of was reduced once testing became more widely available.

Another clue to the impact of patient-related exposures on differences in risk between staff groups is the effect of adjusting risk estimates for exposure category (Table 2). ORs reduced for all staff groups, as expected given a partial correlation between staff group and exposure category. However, the reductions were greatest for patient-facing occupations. For example, the OR for additional clinical services (a group that included care assistants) fell from 2.31 to 1.63, and that for registered nurses and midwives from 2.28 to 1.57. Such changes point to an important contribution from patient-related exposures, but because of the limitations of the JEM, may not have captured them fully.

When allowance is made for the inaccuracy of sickness absence as a marker for disease, and the possibility of a small occupational risk in the reference group of administrative and clerical workers, the results in Tables 2 and 3 suggest that occupational exposures increased the risk of contracting Covid-19 in additional clinical services, registered nurses and midwives, and allied health professionals by a factor of between 1.5 and 2.5. The average relative risk in doctors and dentists appears to have been somewhat lower, but still elevated. Few studies have explored infection rates of Covid-19 in healthcare staff by occupational group during the first wave of infection in England. Zheng, in a study of 1045 staff at a London hospital tested in March/April 2020, found a higher than expected rate of Covid-19 positivity and correspondingly high Covid-19 sickness absence in medical and dental, nursing, midwifery and additional clinical services staff (2). In a study of 11,500 staff at Oxford University Hospitals, tested between March and early June, porters and cleaners had the highest rates of Covid-19 positivity (3).

In our study, it is notable that risk among laboratory scientists was little higher than in administrative and clerical occupations. This suggests that even early in the epidemic, precautions against transmission of SARS-CoV-2 through the handling of clinical samples were fairly adequate.

While our main outcome measure was cumulative prevalence of any Covid-19 sickness absence, we also explored risk factors for longer episodes, expecting that prolonged absence

might have higher specificity as a marker for Covid-19. Moreover, it would tend to reflect more disabling disease of the type most likely to be considered for compensation. A complication is that it will have depended not only the risk of contracting infection, but also on personal vulnerability once infection occurred. Thus, while risk of any Covid-19 sickness absence was lowest in the oldest age group, that of prolonged absence increased with age (a major determinant of vulnerability (11)). Similarly, the higher risk of prolonged Covid-19 sickness absence among non-white ethnic groups may have been a consequence of higher vulnerability (11). This will be explored further in a separate report.

For most staff groups, ORs were higher for prolonged than for any Covid-19 sickness absence (Table 4), reinforcing the case for a relative risk in the order of two from occupational exposures. The occupational hazard in medical and dental personnel may have been obscured by relatively low vulnerability to severe disease.

Our analysis suggests that during the first wave of the Covid-19 pandemic in England, occupationally-attributable relative risks for Covid-19 among most patient-facing occupations in healthcare workers were in the order of 1.5 to 2.5. For medical and dental personnel, relative risks were a little lower, but still elevated. Better protective measures for these groups should be a priority in the future. Whether relative risks are sufficient to warrant compensation for Covid-19 as an occupational disease in healthcare workers will depend on the regulatory framework, and the required confidence of occupational attribution.

Ethical approval statement

Approval granted by the NHS Health Research Authority (reference 20/SC/0282). The study was registered at ISRCTN: 36352994

Contributorship statement

All authors contributed to the planning, conduct, analyses and reporting of this manuscript as outlined below.

Diana van der Platt (statistician): was responsible for the statistical aspects of analysis and interpretation of the quantitative aspects of the study.

Ira Madan (Consultant Occupational Physician and Reader): was co-chief investigator with responsibility for advising on study design, analysis and interpretation of results.

David Coggon (Emeritus Professor of Occupational and Environmental Medicine): was responsible for advising on methodological design, analysis and interpretation of results.

Martie van Tongeren (Professor of Occupational and Environmental Medicine): was responsible for advising on study design, analysis and interpretation of results.

Rhiannon Edge (Lecturer in Population Health): was responsible for advising on study design, analysis and interpretation of results.
Rupert Muiry (Research assistant): was responsible for scoping out and reviewing the emerging literature.
Vaughan Parsons (Research manager): was responsible for overseeing the set-up and delivery of the study, and facilitated data collection.
Paul Cullinan (Professor in Occupational and Environmental Respiratory Disease): was chief investigator with responsibility for advising on study design, analysis and interpretation of results. Had overall responsibility for the management and delivery of the study.

Data Sharing Statement

With permission, source data is available upon request from the NHS Electronic Staff Record (ESR) Warehouse (NHS England)

Acknowledgments

We are very grateful to the following, without whom the study would not have been possible: Sam Wright, Workforce Information Advisor, NHS Electronic Staff Record, and Mike Vickerman, Workforce Information and Analysis, DHSC. Dr Gavin Debrera (Public Health England) and Dr Kit Harling gave invaluable help in planning the study.

We are grateful too, to Lee Isidore, Manal Sadik and Victoria Thorpe for their helpful input into the interpretation of our findings.

We would also like to thank Cambridge University Hospitals NHS Foundation Trust (Dr Mark Ferris), Guys and St Thomas's NHS Foundation Trust (Dr Ali Hashtroudi) and Bolton NHS Foundation Trust (Dr Martin Seed) for providing results from their staff antibody testing programmes.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

Funding

This study was funded by a grant from the COLT Foundation. Award reference: N/A

References

1. Foster P, Neville S. How poor planning left the UK without enough PPE: Financial Times; 202001/03/2021 [Available from: https://www.ft.com/content/9680c20f-7b71-4f65-9bec-0e9554a8e0a7.
2. Iacobucci G. Covid-19: Lack of testing led to patients being discharged to care homes with virus, say auditors. BMJ. 2020;369:m2375.
3. ONS. Coronavirus (COVID-19) related deaths by occupation, England and Wales: deaths registered between 9 March and 28 December 2020. Office for National Statistics; 2021.
4. Zheng C, Hafezi-Bakhtiari N, Cooper V, Davidson H, Habibi M, Riley P, et al. Characteristics and transmission dynamics of COVID-19 in healthcare workers at a London teaching hospital. The Journal of hospital infection. 2020;106(2):325–9.
5. Eyre DW, Lumley SF, O'Donnell D, Campbell M, Sims E, Lawson E, et al. Differential occupational risks to healthcare workers from SARS-CoV-2 observed during a prospective observational study. eLife. 2020;9(e60675).
6. Jones NK, Rivett L, Sparkes D, Forrest S, Sridhar S, Young J, et al. Effective control of SARS-CoV-2 transmission between healthcare workers during a period of diminished community prevalence of COVID-19. eLife. 2020;9(e59391).
7. Timson A, Clair A. Better housing is crucial for our health and the COVID-19 recovery. The Health Foundation; 2020.
8. Beemsterboer W, Stewart R, Groothoff J, Nijhuis F. A literature review on sick leave determinants (1984-2004). International Journal of Occupational Medicine and Environmental Health. 2009;22(2):169-79.
9. NHS England. COVID-19 NHS Situation Report. UK: NHS England; 2021.
10. Kulu H, Dorey P. Infection rates from Covid-19 in Great Britain by geographical units: A model-based estimation from mortality data. Health & Place. 2021;67(102460).
11. Service I. Poverty, Inequality and COVID-19. 2020.

Supplementary file A

SOURCE MATERIAL AND PREPARATION OF DATA FOR ANALYSIS

SOURCE MATERIAL

After ethical approval of our study protocol by the National Health Service (NHS) Health Research Authority (reference 20/SC/0282), the NHS Electronic Staff Record (ESR) Central Team prepared two pseudonymised data files to which we were given access.

SIP File

The first file ("SIP") comprised records for all individuals who were continuously employed by NHS trusts in England and Wales from 1 January 2019 (or before) to 31 July 2020. Each record had the following fields:

Coded identifier
A unique, encrypted 8 to 9 digit serial number, assigned by the ESR Central Team, which would enable linkage to other data (see below)

Sex
Two categories

Age band
Nine categories defined by the individual's age at 15th of September 2020.

Ethnicity
Sixty categories

Trust
Separate categories for each of 200 trusts – individuals were classified according to the trust by which they were employed at 31 July 2020

NHS region
The NHS region in which the employing trust was located (nine categories)

Staff group
A single variable with 10 categories specified as in Table A1 (further information on this classification can be found at https://digital.nhs.uk/binaries/content/assets/website-assets/data-and-information/data-sets/nwd-and-nhs-occupation-codes/nwd_v3.0_data_set_specification_v1.0_final.xlsx).
Individuals were classed according to their staff group at 31 July 2020.

Table A1. Specification of staff groups

Staff group	Definition	Example job roles
Administrative and clerical	Non-clinical staff including non-clinical managers, administrative officers, executive board members who do not have significant patient contact as part of their role	Accountant, chief executive, clerical worker, receptionist
Additional clinical services	Staff directly supporting those in clinical roles. Support to nursing, allied health professionals, health care scientists and other scientific staff are included. Have significant patient contact as part of their role.	Call operator, emergency care assistant, healthcare assistant, nursery nurse
Additional professional scientific and technical	Scientific staff including registered pharmacists, psychologists, social workers, and other roles such as technicians and psychological therapists	Pharmacist, chaplain, social worker, osteopath
Allied health professionals	Registered clinical staff providing diagnostic, technical and therapeutic patient care, including dietitians, radiographers and physiotherapists. Includes qualified ambulance staff such as paramedics	Dietitian, physiotherapist, paramedic, drama therapist, specialist practitioner
Estates and ancillary	Non-clinical support and maintenance staff, including gardeners, plumbers, cooks and housekeepers who do not have significant patient contact as part of their role	Electrician, housekeeper, telephonist
Healthcare scientists	Registered qualified and other staff working in a defined healthcare scientist role, including clinical scientists and biomedical scientists and technicians working in healthcare science. Also includes public health scientific staff	Healthcare scientist, consultant healthcare scientist, healthcare science practitioner
Medical and dental	Registered doctors and dentists	Consultant, clinical assistant, dental officer, foundation year 1, specialty doctor
Nursing and midwifery registered	Registered nurses and midwives	Staff nurse, midwife, community nurse, modern matron, nurse consultant
Students	Directly employed staff undertaking formal education, including student nurses and midwives	Student midwife, student dietitian, student orthoptist
No staff group specified		

Exposure category

This was a single variable with eight categories (Table A2). Before compilation of the data files, the ESR Management Team provided us with a list of 659 detailed occupational codes, which are part of the National Workforce Data Set (NWD) owned by NHS Digital and are used across workforce data in the NHS. These occupational codes were used to classify occupations in the ESR database. Because of the need to protect privacy, it was not possible to release individual information at this level to the research team. However, based on their personal knowledge, four members of the study team (an occupational hygienist and three senior occupational physicians with long experience of work in the NHS) constructed a job-exposure matrix (JEM) classifying each occupational category to one of the nine exposure categories. The ESR Management Team then applied the JEM to each individual's detailed occupational code at 31 July 2020, to generate the exposure category in the SIP file. The specificity with which exposures could be assigned, varied by occupation. For example, the occupational coding scheme distinguished doctors working in intensive care from those working in general surgery, but it did not distinguish between nurses in different adult and general services.

Table A2. Specification of exposure categories

1. Hands-on or face-to-face care of patients much more likely to have Covid-19 than general population

2. Hands-on or face-to-face care of patients who may be more likely to have Covid-19 than general population

3. Hands-on or face-to-face care of patients whose prevalence of Covid-19 is likely to be similar to, or lower than in the general population

4. No hands-on or face-to-face care of patients, but often working in patient areas where patients are more likely to have Covid-19 than general population

5. No hands-on or face-to-face care of patients, but often working in patient areas where the prevalence of Covid-19 among patients is likely to be similar to, or lower than, in the general population

6. No hands-on or face-to-face care of patients, but occasionally in patient areas

7. Unlikely to be in patient areas, but work with material (blood/urine/clothing/equipment/installations) potentially contaminated by virus

8. Other occupation (i.e. not any of 1-7) or occupation unknown

ABS File

The second (ABS) file included records of all absences during 1 January 2019 to 31 July 2020, other than for annual leave, among staff included in the SIP file. There was one record for each absence, with the following fields:

Coded identifier
Specified as for the SIP file

Sex
Specified as for the SIP file

Age band
Specified as for the SIP file

Ethnicity
Specified as for the SIP file

Trust
Coded to 200 categories as in the SIP file, except that individuals were classified according to the trust by which they were employed at the date when their absence started

NHS region
The NHS region in which the employing trust (at the date when the absence started) was located (nine categories)

Staff group
Classified to 10 categories as in the SIP file, but based on the job held at the date when the absence began

Exposure category
A single variable with eight categories, specified as in the SIP file, but based on the job held at the date when the absence began

Reason for absence
Described by four variables:

- Attendance type (19 categories)
- Absence category (11 categories
- Attendance reason (77 categories)
- Related reason (5 categories)

The categories of the first three of these variables can be found in Table A4 below. The categories of related reason were:

- Coronavirus (COVID-19)
- Coronavirus (COVID-19) – Household Member Symptoms
- Coronavirus (COVID-19) – Post travel Quarantine
- Coronavirus (COVID-19) – Test and Trace Contact
- Menopause

Date absence started
Date month and year

Date absence ended
Date month and year

Numbers of records

In total, 981,131 individuals had at least one record in the SIP file, and of those, 835,180 had at least one recorded absence in the ABS file (no individuals had a record in the ABS file but not in the SIP file). The ABS file contained records on a total of 4,159,991 absence episodes.

INITIAL STEPS

Removal of complete duplicates

We first removed all records that were exact duplicates from the SIP file (n = 2,880) and ABS file (53,288).

Reclassification of ethnicity

The scheme that was used to classify ethnicity in the two data files distinguished 77 different categories. Some were quite rare, limiting the potential for their analysis as separate entities. Also, there were frequent instances of overlap in the specification of categories (e.g. “Indian” with “Asian or Asian British – Indian” and “White English” with White British”).

We therefore aggregated the starting categories into two broader classifications of ethnicity (“Ethnicity group 1” and “Ethnicity group 2”), one nested within the other, with the aim of ensuring adequate numbers for meaningful analysis in each aggregated category, and reducing overlap and ambiguity. Table A3 shows the frequency of each category of ethnicity in the original data files, and the aggregated categories to which it was assigned in the two broader classifications.

Reclassification of reason for absence

The first three variables that characterised reason for absence (attendance type, absence category, and attendance reason) occurred in 192 different combinations. To facilitate further analysis, these combinations were collapsed into 60 categories of a new variable which we labelled as “Collapsed absence category” (Table A4).

Table A3. Reclassification of ethnicity

Original category	Frequency across all SIP and ABS records	Allocation in revised classifications	
		Ethnicity group 1	Ethnicity group 2
0 White	243	White	White
1 Black-Caribbean	173	Black or Black British - Caribbean	Black or Black British
2 Black-African	911	Black or Black British - African	Black or Black British
3 Black-Other	27	Black or Black British - Any other or unspecified Black background	Black or Black British
4 Indian	181	Asian or Asian British - Indian	Asian or Asian British - South Asian
5 Pakistani	12	Asian or Asian British - Pakistani	Asian or Asian British - South Asian
7 Chinese	9	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
9 Not given	47	Not stated	Not stated
A White - British	3477821	White	White
B White - Irish	56427	White	White
C White - Any other White background	188395	White	White
C2 White Northern Irish	1167	White	White
C3 White Unspecified	7271	White	White
CA White English	36682	White	White
CB White Scottish	3378	White	White
CC White Welsh	2343	White	White
CD White Cornish	1686	White	White
CE White Cypriot (non specific)	206	White	White
CF White Greek	2157	White	White
CG White Greek Cypriot	490	White	White
CH White Turkish	612	White	White
CJ White Turkish Cypriot	271	White	White
CK White Italian	5880	White	White
CL White Irish Traveller	38	White	White

Original category	Frequency across all SIP and ABS records	Allocation in revised classifications	
		Ethnicity group 1	Ethnicity group 2
CM White Traveller	73	White	White
CN White Gypsy/Romany	251	White	White
CP White Polish	12075	White	White
CQ White ex-USSR	873	White	White
CR White Kosovan	163	White	White
CS White Albanian	452	White	White
CT White Bosnian	40	White	White
CU White Croatian	378	White	White
CV White Serbian	145	White	White
CW White Other Ex-Yugoslav	280	White	White
CX White Mixed	2785	Mixed	Mixed
CY White Other European	26591	White	White
D Mixed - White & Black Caribbean	27712	Mixed	Mixed
E Mixed - White & Black African	13526	Mixed	Mixed
F Mixed - White & Asian	19742	Mixed	Mixed
G Mixed - Any other mixed background	23086	Mixed	Mixed
GA Mixed - Black & Asian	472	Mixed	Mixed
GB Mixed - Black & Chinese	44	Mixed	Mixed
GC Mixed - Black & White	1179	Mixed	Mixed
GD Mixed - Chinese & White	461	Mixed	Mixed
GE Mixed - Asian & Chinese	440	Mixed	Mixed
GF Mixed - Other/Unspecified	1706	Mixed	Mixed
H Asian or Asian British - Indian	238011	Asian or Asian British - Indian	Asian or Asian British - South Asian
J Asian or Asian British - Pakistani	76282	Asian or Asian British - Pakistani	Asian or Asian British - South Asian
K Asian or Asian British - Bangladeshi	27145	Asian or Asian British - Bangladeshi	Asian or Asian British - South Asian
L Asian or Asian British - Any other Asian background	132728	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified

Original category	Frequency across all SIP and ABS records	Allocation in revised classifications	
		Ethnicity group 1	Ethnicity group 2
LA Asian Mixed	1601	Mixed	Mixed
LB Asian Punjabi	1328	Asian or Asian British - Indian	Asian or Asian British - South Asian
LC Asian Kashmiri	250	Asian or Asian British - Indian	Asian or Asian British - South Asian
LD Asian East African	482	Other	Other
LE Asian Sri Lankan	2468	Asian or Asian British - Sri Lankan	Asian or Asian British - South Asian
LF Asian Tamil	985	Asian or Asian British - Indian	Asian or Asian British - South Asian
LG Asian Sinhalese	180	Asian or Asian British - Sri Lankan	Asian or Asian British - South Asian
LH Asian British	6769	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
LJ Asian Caribbean	743	Other	Other
LK Asian Unspecified	6325	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
M Black or Black British - Caribbean	87878	Black or Black British - Caribbean	Black or Black British
N Black or Black British - African	206147	Black or Black British - African	Black or Black British
Nan	21209	Not stated	Not stated
P Black or Black British - Any other Black background	18128	Black or Black British - Any other or unspecified Black background	Black or Black British
PA Black Somali	2367	Black or Black British - African	Black or Black British
PB Black Mixed	503	Mixed	Mixed
PC Black Nigerian	9720	Black or Black British - African	Black or Black British
PD Black British	11356	Black or Black British - Any other or unspecified Black background	Black or Black British
PE Black Unspecified	1435	Black or Black British - Any other or unspecified Black background	Black or Black British
R Chinese	17949	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
S Any Other Ethnic Group	64865	Other	Other
SA Vietnamese	213	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified

Original category	Frequency across all SIP and ABS records	Allocation in revised classifications	
		Ethnicity group 1	Ethnicity group 2
SB Japanese	273	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
SC Filipino	55696	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
SD Malaysian	705	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
SE Other Specified	6487	Other	Other
Z Not Stated	169316	Not stated	Not stated
Missing	4056	Not stated	Not stated

Table A4. Specification of collapsed absence categories

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Paid Part Day	Paid Leave	Adoption Appointment	Adoption appointment
Special Decreasing Bal	Special Leave	Adoption Appointment	Adoption appointment
Special Increasing Bal	Special Leave	Adoption Appointment	Adoption appointment
Unpaid Authorised Special	Special Leave	Adoption Appointment	Adoption appointment
Unpaid Authorised Special Hrs	Special Leave	Adoption Appointment	Adoption appointment
Adoption	Adoption	Adoption Leave	Adoption leave
Adoption	Adoption		Adoption leave
Unpaid Authorised Special	Special Leave	Adoption Leave	Adoption leave
Unpaid Authorised Special Hrs	Special Leave	Adoption Leave	Adoption leave
Paternity Adoption	Paternity Adoption	Paternity Leave	Adoption leave paternity
Paternity Adoption	Paternity Adoption		Adoption leave paternity
Shared Parental Adoption	Shared Parental Adoption		Adoption leave shared parental adoption
Unpaid Authorised Special Hrs	Special Leave	Annual Leave	Annual leave
Special Decreasing Bal	Special Leave	Antenatal	Antenatal
Special Increasing Bal	Special Leave	Antenatal	Antenatal
Paid Part Day	Paid Leave	Attendance at Public Bodies	Attendance public body
Special Decreasing Bal	Special Leave	Attendance at Public Bodies	Attendance public body
Special Increasing Bal	Special Leave	Attendance at Public Bodies	Attendance public body
Unpaid Authorised Special	Special Leave	Attendance at Public Bodies	Attendance public body
Unpaid Authorised Special Hrs	Special Leave	Attendance at Public Bodies	Attendance public body
Special Decreasing Bal	Special Leave	Bereavement	Bereavement
Special Increasing Bal	Special Leave	Bereavement	Bereavement
Unpaid Authorised Special	Special Leave	Bereavement	Bereavement
Unpaid Authorised Special Hrs	Special Leave	Bereavement	Bereavement
Special Decreasing Bal	Special Leave	Career Break	Career break
Special Increasing Bal	Special Leave	Career Break	Career break
Unpaid Authorised Special	Special Leave	Career Break	Career break
Unpaid Authorised Special Hrs	Special Leave	Career Break	Career break
Paid Part Day	Paid Leave	Carer's Leave	Carer's leave
Special Decreasing Bal	Special Leave	Carer's Leave	Carer's leave
Special Decreasing Bal	Special Leave	Emergency Leave/Time Off for Dependants	Carer's leave

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Special Increasing Bal	Special Leave	Carer's Leave	Carer's leave
Special Increasing Bal	Special Leave	Emergency Leave/Time Off for Dependants	Carer's leave
Unpaid Authorised Special	Special Leave	Carer's Leave	Carer's leave
Unpaid Authorised Special	Special Leave	Emergency Leave/Time Off for Dependants	Carer's leave
Unpaid Authorised Special Hrs	Special Leave	Carer's Leave	Carer's leave
Unpaid Authorised Special Hrs	Special Leave	Emergency Leave/Time Off for Dependants	Carer's leave
Paid Part Day	Paid Leave	Compassionate Leave	Compassionate leave
Special Decreasing Bal	Special Leave	Compassionate Leave	Compassionate leave
Special Increasing Bal	Special Leave	Compassionate Leave	Compassionate leave
Unpaid Authorised Special	Special Leave	Compassionate Leave	Compassionate leave
Unpaid Authorised Special Hrs	Special Leave	Compassionate Leave	Compassionate leave
Special Decreasing Bal	Special Leave	Court Appearance	Court appearance
Special Increasing Bal	Special Leave	Court Appearance	Court appearance
Unpaid Authorised Special	Special Leave	Court Appearance	Court appearance
Unpaid Authorised Special Hrs	Special Leave	Court Appearance	Court appearance
Paid Part Day	Paid Leave	Disability Leave	Disability leave
Special Decreasing Bal	Special Leave	Disability Leave	Disability leave
Special Increasing Bal	Special Leave	Disability Leave	Disability leave
Unpaid Authorised Special	Special Leave	Disability Leave	Disability leave
Unpaid Authorised Special Hrs	Special Leave	Disability Leave	Disability leave
Paid Part Day	Paid Leave	Industrial Action	Industrial action
Special Increasing Bal	Special Leave	Industrial Action	Industrial action
Unpaid Authorised Special	Special Leave	Industrial Action	Industrial action
Unpaid Authorised Special Hrs	Special Leave	Industrial Action	Industrial action
Unpaid Unauth Special Hrs	Special Leave	Industrial Action	Industrial action
Unpaid Unauthorised Special	Special Leave	Industrial Action	Industrial action
Special Decreasing Bal	Special Leave	Infection Precaution	Infection precaution
Special Increasing Bal	Special Leave	Infection Precaution	Infection precaution
Paid Part Day	Paid Leave	Interview Leave	Interview leave
Special Decreasing Bal	Special Leave	Interview Leave	Interview leave
Special Increasing Bal	Special Leave	Interview Leave	Interview leave
Unpaid Authorised Special	Special Leave	Interview Leave	Interview leave
Unpaid Authorised Special Hrs	Special Leave	Interview Leave	Interview leave

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Special Decreasing Bal	Special Leave	Jury Service	Jury service
Special Increasing Bal	Special Leave	Jury Service	Jury service
Unpaid Authorised Special	Special Leave	Jury Service	Jury service
Unpaid Authorised Special Hrs	Special Leave	Jury Service	Jury service
Maternity	Maternity	Maternity Leave	Maternity leave
Maternity	Maternity		Maternity leave
Medical Suspension with Pay	Paid Leave	Allergy	Medical suspension
Medical Suspension with Pay	Paid Leave	Infection	Medical suspension
Medical Suspension with Pay	Paid Leave	Needlestick	Medical suspension
Medical Suspension with Pay	Paid Leave	Other	Medical suspension
Medical Suspension with Pay	Paid Leave		Medical suspension
Special Decreasing Bal	Special Leave	Medical Suspension	Medical suspension
Special Increasing Bal	Special Leave	Medical Suspension	Medical suspension
Paid Part Day	Paid Leave	Medical/Dental Appointment	Medical/dental appointment
Special Decreasing Bal	Special Leave	Medical/Dental Appointment	Medical/dental appointment
Special Increasing Bal	Special Leave	Medical/Dental Appointment	Medical/dental appointment
Unpaid Authorised Special	Special Leave	Medical/Dental Appointment	Medical/dental appointment
Unpaid Authorised Special Hrs	Special Leave	Medical/Dental Appointment	Medical/dental appointment
Paid Part Day	Paid Leave	Other	Other part day
Paid Part Day	Paid Leave	Paid Part Day	Other part day
Paid Part Day	Paid Leave	Time Off in Lieu - Other	Other part day
Paid Part Day	Paid Leave	Time Off in Lieu - Overtime/Time Owed	Other part day
Paid Part Day	Paid Leave	Time Off in Lieu - Worked Public Holiday	Other part day
Paid Part Day	Paid Leave	Trade Union Duties	Other part day
Paid Part Day	Paid Leave		Other part day
Special Decreasing Bal	Special Leave	Other	Other special leave
Special Decreasing Bal	Special Leave	Time Off in Lieu - Other	Other special leave
Special Decreasing Bal	Special Leave	Time Off in Lieu - Overtime/Time Owed	Other special leave
Special Decreasing Bal	Special Leave	Time Off in Lieu - Worked Public Holiday	Other special leave
Special Decreasing Bal	Special Leave	Trade Union Duties	Other special leave
Special Decreasing Bal	Special Leave		Other special leave
Special Increasing Bal	Special Leave	Other	Other special leave
Special Increasing Bal	Special Leave		Other special leave

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Unpaid Authorised Special	Special Leave	Other	Other special leave
Unpaid Authorised Special	Special Leave		Other special leave
Unpaid Authorised Special Hrs	Special Leave	Other	Other special leave
Unpaid Authorised Special Hrs	Special Leave		Other special leave
Unpaid Unauth Special Hrs	Special Leave	Other	Other special leave
Unpaid Unauth Special Hrs	Special Leave	Unauthorised Leave	Other special leave
Unpaid Unauth Special Hrs	Special Leave		Other special leave
Unpaid Unauthorised Special	Special Leave	Other	Other special leave
Unpaid Unauthorised Special	Special Leave	Unauthorised Leave	Other special leave
Unpaid Unauthorised Special	Special Leave		Other special leave
Special Increasing Bal	Special Leave	Time Off in Lieu - Other	Other special leave time off in lieu
Special Increasing Bal	Special Leave	Time Off in Lieu - Overtime/Time Owed	Other special leave time off in lieu
Special Increasing Bal	Special Leave	Time Off in Lieu - Worked Public Holiday	Other special leave time off in lieu
Special Increasing Bal	Special Leave	Trade Union Duties	Other special leave trade union
Unpaid Authorised Special	Special Leave	Trade Union Duties	Other special leave trade union
Special Decreasing Bal	Special Leave	Parental Leave	Parental leave
Special Increasing Bal	Special Leave	Parental Leave	Parental leave
Unpaid Authorised Special	Special Leave	Parental Leave	Parental leave
Unpaid Authorised Special Hrs	Special Leave	Parental Leave	Parental leave
Shared Parental Birth	Shared Parental Birth		Parental leave shared parental birth
Paid Part Day	Paid Leave	Paternity Antenatal Appointment	Paternity appointment
Special Decreasing Bal	Special Leave	Paternity Antenatal Appointment	Paternity appointment
Special Increasing Bal	Special Leave	Paternity Antenatal Appointment	Paternity appointment
Unpaid Authorised Special	Special Leave	Paternity Antenatal Appointment	Paternity appointment
Unpaid Authorised Special Hrs	Special Leave	Paternity Antenatal Appointment	Paternity appointment
Additional Pat Leave Birth	Additional Paternity Birth		Paternity leave
Paternity Birth	Paternity Birth	Paternity Leave	Paternity leave
Paternity Birth	Paternity Birth		Paternity leave
Unpaid Authorised Special	Special Leave	Paternity Leave	Paternity leave
Unpaid Authorised Special Hrs	Special Leave	Paternity Leave	Paternity leave
Paid Part Day	Paid Leave	Phased Return to Work	Phased return
Special Decreasing Bal	Special Leave	Phased Return to Work	Phased return
Special Increasing Bal	Special Leave	Phased Return to Work	Phased return

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Unpaid Authorised Special	Special Leave	Phased Return to Work	Phased return
Unpaid Authorised Special Hrs	Special Leave	Phased Return to Work	Phased return
Special Decreasing Bal	Special Leave	Magisterial/Local Government/Parliamentary Candidate	Political candidate
Special Increasing Bal	Special Leave	Magisterial/Local Government/Parliamentary Candidate	Political candidate
Unpaid Authorised Special	Special Leave	Magisterial/Local Government/Parliamentary Candidate	Political candidate
Unpaid Authorised Special Hrs	Special Leave	Magisterial/Local Government/Parliamentary Candidate	Political candidate
Sickness	Sickness	S14 Asthma	Sickness asthma
Sickness	Sickness	S11 Back Problems	Sickness back problems
Sickness	Sickness	S18 Blood disorders	Sickness blood disorders
Sickness	Sickness	S20 Burns, poisoning, frostbite, hypothermia	Sickness burns, poisoning, frostbite, hypothermia
Sickness	Sickness	S17 Benign and malignant tumours, cancers	Sickness cancer
Sickness	Sickness	S19 Heart, cardiac & circulatory problems	Sickness cardiac and circulatory
Sickness	Sickness	S15 Chest & respiratory problems	Sickness chest and respiratory
Sickness	Sickness	S13 Cold, Cough, Flu - Influenza	Sickness cough, flu
Sickness	Sickness	S22 Dental and oral problems	Sickness dental and oral problems
Sickness	Sickness	S21 Ear, nose, throat (ENT)	Sickness ear, nose, throat
Sickness	Sickness	S24 Endocrine / glandular problems	Sickness endocrine, glandular problems
Sickness	Sickness	S23 Eye problems	Sickness eye problems
Sickness	Sickness	S25 Gastrointestinal problems	Sickness gastrointestinal problems
Sickness	Sickness	S26 Genitourinary & gynaecological disorders	Sickness genitourinary, gynaecological problems
Sickness	Sickness	S16 Headache / migraine	Sickness headache, migraine
Sickness	Sickness	S27 Infectious diseases	Sickness infectious diseases
Sickness	Sickness	S28 Injury, fracture	Sickness injury, fracture
Sickness	Sickness	S10 Anxiety/stress/depression/other psychiatric illnesses	Sickness mental health
Sickness	Sickness	Stress	Sickness mental health
Sickness	Sickness	S29 Nervous system disorders	Sickness nervous system disorders
Sickness	Sickness	S98 Other known causes - not elsewhere classified	Sickness other
Sickness	Sickness	Musculo-skeletal Back	Sickness other MSDs
Sickness	Sickness	Musculo-skeletal Other Joint, Lower Limb	Sickness other MSDs
Sickness	Sickness	S12 Other musculoskeletal problems	Sickness other MSDs
Paid Part Day	Paid Leave	Sickness	Sickness part day
Sickness	Sickness	Pregnancy Related	Sickness pregnancy related disorders
Sickness	Sickness	S30 Pregnancy related disorders	Sickness pregnancy related disorders

Attendance Type	Absence Category	Attendance Reason	Collapsed absence category
Sickness	Sickness	S31 Skin disorders	Sickness skin disorders
Sickness	Sickness	S32 Substance abuse	Sickness substance abuse
Sickness	Sickness	Surgery	Sickness surgery
Sickness	Sickness	Not Known	Sickness unknow cause
Sickness	Sickness	S99 Unknown causes / Not specified	Sickness unknow cause
Study Decreasing Bal	Study Leave	Professional Leave	Study leave
Study Decreasing Bal	Study Leave	Study Leave	Study leave
Study Decreasing Bal	Study Leave		Study leave
Study Increasing Bal	Study Leave	Study Leave	Study leave
Study Increasing Bal	Study Leave		Study leave
Training Development	Paid Leave	Development	Study leave
Training Development	Paid Leave	External Training	Study leave
Training Development	Paid Leave	Internal Training	Study leave
Training Development	Paid Leave	Other	Study leave
Training Development	Paid Leave		Study leave
Special Decreasing Bal	Special Leave	Suspended - Paid	Suspended
Special Increasing Bal	Special Leave	Suspended - Paid	Suspended
Unpaid Authorised Special	Special Leave	Suspended - Unpaid	Suspended
Unpaid Authorised Special Hrs	Special Leave	Suspended - Unpaid	Suspended
Special Decreasing Bal	Special Leave	Deployment with Reserve Forces	Volunteer
Special Decreasing Bal	Special Leave	Training with Reserve and Cadet Forces	Volunteer
Special Decreasing Bal	Special Leave	Volunteer Leave	Volunteer
Special Increasing Bal	Special Leave	Deployment with Reserve Forces	Volunteer
Special Increasing Bal	Special Leave	Training with Reserve and Cadet Forces	Volunteer
Special Increasing Bal	Special Leave	Volunteer Leave	Volunteer
Unpaid Authorised Special	Special Leave	Training with Reserve and Cadet Forces	Volunteer
Unpaid Authorised Special	Special Leave	Volunteer Leave	Volunteer
Unpaid Authorised Special Hrs	Special Leave	Training with Reserve and Cadet Forces	Volunteer
Unpaid Authorised Special Hrs	Special Leave	Volunteer Leave	Volunteer

INCONSISTENCIES AND ANOMALIES IN DATA ON INDIVIDUAL ABSENCE EPISODES

We next considered each individual episode of absence, looking for inconsistencies and anomalies in the recorded data. These took several forms.

Inconsistent dates of absence

Two episodes, both for maternity leave, had an end date before their start date. In both cases, we assumed that the record was corrupted as it is not possible for a user to enter an end date that is prior to the start date, and therefore we changed the incorrect end date to one year later. This meant that the duration of the leave in each case was a little less than 12 months, which was a common duration for maternity leave in the dataset as a whole. In one of the two cases, there was also an absence for “medical suspension” linked to Covid-19 which started four weeks after the maternity leave. However, such overlap occurred quite frequently with maternity leave and was not judged to invalidate the imputed end date.

Implausible durations of absence

18 end dates, all of which were after 31 July 2020, gave implausibly long durations that were incompatible with the reason for absence. These appeared to have arisen from transposition of digits in data entry (e.g. 2019 being entered as 2109 or 2018 being entered as 2028), and were revised accordingly.

Missing end dates for absences

The approach to missing end dates depended on whether there was another episode of absence with a later start date.

Where there was no subsequent episode with a later start date, we assumed that an absence continued beyond 31 July 2020, provided that would not make it implausibly long (n = 60,805). In this context, plausibility was determined from the distribution of durations for all absences in the same collapsed absence category with known start and end dates. Where continuation beyond 31 July 2020 would imply an implausibly long duration, we imputed an end date by adding the median duration for the collapsed absence category to the start date (n = 466).

Where the individual had another episode with a later start date, the missing end date was imputed as the earlier of: a) the start date of the episode plus the median duration for the collapsed absence category; and b) the start date of the next episode for that individual minus one day (n = 6,938).

ELIMINATION OF ABSENCES THAT ENDED BEFORE THE START OF THE STUDY PERIOD

Having made these adjustments, we removed from the ABS file, all records for absences that ended before 01 January 2019 (n = 1,567), leaving a total of 4,103,602 remaining absence records.

OTHER INCONSISTENCIES AND ANOMALIES

Where an individual had one or more records in the ABS file as well as a record in the SIP file, it was important to check that all of his/her records were mutually consistent. Some characteristics (sex, age group and ethnicity) could not change over time and should therefore have been identical in all records for the same individual. Others (trust, staff group, exposure category) were specific to the date of the record (31 July 2020 for the SIP file and date absence started for the ABS file) and might have changed over time if people changed jobs. Furthermore, it was possible that some individuals held more than one job simultaneously, and that trust, staff group and/or exposure category differed between those jobs. It was also important to check that for each individual, the start and end dates of all absence episodes were mutually compatible.

When exploring inconsistencies, we considered the possibility that records might have been assigned to the wrong individual. A pointer to this would be inconsistencies for multiple demographic variables in the same person. However, only one individual had inconsistencies in both of sex and ethnicity, one in both age group and ethnicity, and none in both sex and age group. We therefore judged such error to be unlikely.

Sex

Sex was recorded inconsistently in 33 individuals. This may have occurred because first names were not clearly sex-specific. Each individual was assigned the modal value for sex across all of his/her records, giving preference to the value from the SIP file in the event of a tie.

Age group

Age group was recorded inconsistently in 45 individuals. Each was assigned the modal value for age group across all of his/her records, giving preference to the value from the SIP file in the event of a tie.

Ethnicity

Across all of their records in the SIP and ABS files, 6,872 individuals were assigned to two or three different categories of ethnicity. In many cases this reflected missing information or a lack of specificity in some records as compared with others. Table A5 shows the conventions that we adopted in reclassifying each combination of ethnicity categories to a single category in Ethnicity group 1 and Ethnicity group 2. Table A5 also shows the frequency with which each combination occurred in the dataset.

Table A5. Assignment of Ethnicity groups 1 and 2 where more than one category of ethnicity was assigned to the same individual in the SIP and ABS data files

Combination of different ethnicity categories for same individual			Number of individuals	Allocation in revised classifications	
				Ethnicity group 1	Ethnicity group 2
White	Asian or Asian British - Any other or unspecified Asian background		28	Mixed	Mixed
White	Asian or Asian British - Bangladeshi		6	Mixed	Mixed
White	Asian or Asian British - Indian		37	Mixed	Mixed
White	Asian or Asian British - Pakistani		15	Mixed	Mixed
White	Black or Black British - African		11	Mixed	Mixed
White	Black or Black British - Any other or unspecified Black background		6	Mixed	Mixed
White	Black or Black British - Caribbean		26	Mixed	Mixed
White	Mixed	Not stated	1	Mixed	Mixed
White	Mixed		291	Mixed	Mixed
White	Not stated		3675	White	White
White	Other	Not stated	1	Mixed	Mixed
White	Other		124	Mixed	Mixed
Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Bangladeshi		16	Asian or Asian British - Bangladeshi	Asian or Asian British - South Asian
Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Indian		103	Asian or Asian British - Indian	Asian or Asian British - South Asian
Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Pakistani		30	Asian or Asian British - Pakistani	Asian or Asian British - South Asian

Combination of different ethnicity categories for same individual			Number of individuals	Allocation in revised classifications	
				Ethnicity group 1	Ethnicity group 2
Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Sri Lankan		11	Asian or Asian British - Sri Lankan	Asian or Asian British - South Asian
Asian or Asian British - Any other or unspecified Asian background	Black or Black British - African		3	Mixed	Mixed
Asian or Asian British - Any other or unspecified Asian background	Black or Black British - Any other or unspecified Black background		6	Mixed	Mixed
Asian or Asian British - Any other or unspecified Asian background	Not stated		302	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
Asian or Asian British - Any other or unspecified Asian background	Other		374	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - Other or unspecified
Asian or Asian British - Bangladeshi	Asian or Asian British - Pakistani		3	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - South Asian
Asian or Asian British - Bangladeshi	Not stated		38	Asian or Asian British - Bangladeshi	Asian or Asian British - South Asian
Asian or Asian British - Bangladeshi	Other		1	Asian or Asian British - Bangladeshi	Asian or Asian British - South Asian
Asian or Asian British - Indian	Asian or Asian British - Bangladeshi		6	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - South Asian
Asian or Asian British - Indian	Asian or Asian British - Pakistani		19	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - South Asian
Asian or Asian British - Indian	Asian or Asian British - Sri Lankan		1	Asian or Asian British - Any other or unspecified Asian background	Asian or Asian British - South Asian
Asian or Asian British - Indian	Black or Black British - Any other or unspecified Black background		2	Mixed	Mixed

Combination of different ethnicity categories for same individual			Number of individuals	Allocation in revised classifications	
				Ethnicity group 1	Ethnicity group 2
Asian or Asian British - Indian	Not stated		279	Asian or Asian British - Indian	Asian or Asian British - South Asian
Asian or Asian British - Indian	Other	Not stated	1	Asian or Asian British - Indian	Asian or Asian British - South Asian
Asian or Asian British - Indian	Other		18	Asian or Asian British - Indian	Asian or Asian British - South Asian
Asian or Asian British - Pakistani	Not stated		76	Asian or Asian British - Pakistani	Asian or Asian British - South Asian
Asian or Asian British - Pakistani	Other		6	Asian or Asian British - Pakistani	Asian or Asian British - South Asian
Asian or Asian British - Sri Lankan	Not stated		1	Asian or Asian British - Sri Lankan	Asian or Asian British - South Asian
Asian or Asian British - Sri Lankan	Other		3	Asian or Asian British - Sri Lankan	Asian or Asian British - South Asian
Black or Black British - African	Asian or Asian British - Indian		4	Mixed	Mixed
Black or Black British - African	Asian or Asian British - Pakistani		3	Mixed	Mixed
Black or Black British - African	Black or Black British - Any other or unspecified Black background	Not stated	1	Black or Black British - African	Black or Black British
Black or Black British - African	Black or Black British - Any other or unspecified Black background		212	Black or Black British - African	Black or Black British
Black or Black British - African	Black or Black British - Caribbean		46	Black or Black British - Any other or unspecified Black background	Black or Black British
Black or Black British - African	Not stated		211	Black or Black British - African	Black or Black British
Black or Black British - African	Other	Not stated	2	Black or Black British - African	Black or Black British
Black or Black British - African	Other		32	Black or Black British - African	Black or Black British
Black or Black British - Any other or unspecified Black background	Not stated		35	Black or Black British - Any other or unspecified Black background	Black or Black British
Black or Black British - Any other or unspecified Black background	Other		8	Black or Black British - Any other or unspecified Black background	Black or Black British

Combination of different ethnicity categories for same individual			Number of individuals	Allocation in revised classifications	
				Ethnicity group 1	Ethnicity group 2
Black or Black British - Caribbean	Asian or Asian British - Pakistani		1	Mixed	Mixed
Black or Black British - Caribbean	Black or Black British - Any other or unspecified Black background	Not stated	2	Black or Black British - Caribbean	Black or Black British
Black or Black British - Caribbean	Black or Black British - Any other or unspecified Black background		88	Black or Black British - Caribbean	Black or Black British
Black or Black British - Caribbean	Not stated		101	Black or Black British - Caribbean	Black or Black British
Black or Black British - Caribbean	Other		11	Black or Black British - Caribbean	Black or Black British
Mixed	Asian or Asian British - Any other or unspecified Asian background	Not stated	1	Mixed	Mixed
Mixed	Asian or Asian British - Any other or unspecified Asian background		49	Mixed	Mixed
Mixed	Asian or Asian British - Bangladeshi		2	Mixed	Mixed
Mixed	Asian or Asian British - Indian		22	Mixed	Mixed
Mixed	Asian or Asian British - Pakistani		8	Mixed	Mixed
Mixed	Black or Black British - African		64	Mixed	Mixed
Mixed	Black or Black British - Any other or unspecified Black background		13	Mixed	Mixed
Mixed	Black or Black British - Caribbean		38	Mixed	Mixed
Mixed	Not stated		160	Mixed	Mixed
Mixed	Other		78	Mixed	Mixed
Other	Not stated		160	Other	Other

Trust

There were no inconsistencies across the ABS and SIP files in the trust to which each individual was assigned.

Staff group

As indicated above, it was to be expected that some people had changed staff group over the course of the study period. However, there were 19,467 individuals with two or more different recorded values for staff group, of which 4,234 had different recorded values for staff group either in the SIP file or on the same date in the ABS file. Review of the staff group categories that occurred in combination did not identify any that clearly looked anomalous, and we therefore assumed that the individuals concerned had held two different jobs simultaneously. To account for this, we derived a new category of “Multiple” where an individual had two different staff groups on the date to which the record referred.

Where an individual was recorded as having different staff groups on different dates, each was treated as valid for the date in question, there being no basis for judging that one was implausible.

Exposure category

As with staff group, it was to be expected that some people had changed exposure group over the course of the study period. However, there were 40,094 individuals with two or more different values for exposure category recorded, of which 7,066 were recorded simultaneously, either in the SIP file or in the ABS file. Review of the exposure codes that occurred in combination indicated that in almost all instances, one category clearly indicated greater potential for occupational exposure than the other. We therefore adopted a convention by which, as far as possible, for the date in question, the individual was assigned the higher of the two exposure categories, applying the conventions summarised in Table A6. Table A6 also indicates (in italics and underscored) combinations of exposure category for which it was less clear which was higher, and the adopted value was therefore somewhat arbitrary.

Where an individual was recorded as having different exposure categories on different dates, each was treated as valid for the date in question, there being no basis for judging that one was implausible.

Table A6: Assigned exposure category where two jobs with different exposure categories were held simultaneously

Assigned categories are shown in italics with an underscore where it was unclear which of a paired combination represented a higher potential for exposure to coronavirus.
Numbers in brackets indicate the frequency with which each combination of exposure categories occurred in two jobs held simultaneously (7,066 individuals had multiple exposure codes simultaneously during any time of the study of which 90 had 3 exposure codes simultaneously).

First exposure category	Second exposure category						
	2	**3**	**4**	**5**	**6**	**7**	**8**
1	1 (207)	1 (377)	1 (9)	1 (2)	1 (80)	1 (6)	1 (632)
2		2 (2046)	2 (18)	2 (7)	2 (799)	2 (81)	2 (304)
3			*3* (1)	3 (8)	3 (663)	*3* (38)	3 (384)
4				(0)	4 (2)	(0)	4 (2)
5					5 (66)	*5* (5)	5 (10)
6						*6* (77)	6 (1391)
7							7 (31)

Key to exposure categories:

1. Hands-on or face-to-face care of patients much more likely to have Covid-19 than general population

2. Hands-on or face-to-face care of patients who may be more likely to have Covid-19 than general population

3. Hands-on or face-to-face care of patients whose prevalence of Covid-19 is likely to be similar to, or lower than in the general population

4. No hands-on or face-to-face care of patients, but often working in patient areas where patients are more likely to have Covid-19 than general population

5. No hands-on or face-to-face care of patients, but often working in patient areas where the prevalence of Covid-19 among patients is likely to be similar to, or lower than, in the general population

6. No hands-on or face-to-face care of patients, but occasionally in patient areas

7. Unlikely to be in patient areas, but work with material (blood/urine/clothing/equipment/installations) potentially contaminated by virus

8. Other occupation (i.e. not any of 1-7) or unknown

Overlapping spells of absence in the same individual

There were 63,556 instances in which two or more spells of absence in the same individual overlapped. For 28,217 recordings, the same absence was recorded for different jobs held simultaneously. Among the remainder, in most cases (n = 29,835), one was nested within the other, but in 5,504 a second absence started during the first, and ended after it or had a missing end date. Review of the reasons ascribed to these overlapping absences indicated that they were generally plausible. Sickness absences cannot overlap with other sickness absences, only non-sickness absences can overlap with sickness absences or non-sickness absence, and we only observed the later (e.g. an episode of absence for infection precaution or medical suspension superimposed on a spell of absence for another reason). However, there was only one instance of overlap between spells of sickness absence for two different specific disease categories.

We therefore made no changes to the records in the ABS file in response to overlapping episodes of absence. However, we decided that in subsequent analyses:

a) When calculating period of absence from work irrespective of reason (e.g. to exclude people who were absent continuously throughout a specified period), we would treat overlapping periods of absence as a single episode running from the earliest start date to latest end date for the overlapping episodes.

b) Where the focus was on spells of absence for a specific reason (e.g. sickness absence for Covid-19), we would ignore overlaps with periods of absence for other reasons, and when multiple periods of absence for the reason of interest overlapped, we would treat them as a single episode of absence for the reason of interest, running from the earliest start date to the latest end date of the overlapping spells.

Supplementary File B

SPECIFICATION OF STUDY PERIOD, COVID-19 SICKNESS ABSENCE AND OCCUPATIONAL VARIABLES

Study period

National data on hospital admissions for Covid-19 by trust[B1] indicated that they first began about 19 March, and because of the lag between onset of illness and admission to hospital, it was to be expected that infection first started to emerge on any scale a week or two earlier. Furthermore, analysis of start dates for Covid-19 sickness absence (for definition see below) indicated a marked increase in daily numbers from 9 March. Also, analysis of results from antibody tests during late May to July in the subset of cohort members at Guys and St Thomas's Trust, showed that the ratio of positive to negative tests among people whose only episode of Covid-19 sickness absence began during the seven days from 9 March was substantially higher than that in people who had taken no such absence at any time.

The start of the study period was therefore taken as 9 March 2020. The end date of 31 July 2020 was that up to which complete data were available in the databases from the NHS Electronic Staff Record.

Covid-19 sickness absence

We sought measures of sickness absence that would serve as markers, albeit imperfect, for incident Covid-19 and for incident cases of more severe Covid-19. As well as the three variables that were used to specify "Collapsed absence category", the ABS file included a variable that denoted whether absences were related to Covid-19. This had five possible values: No relation recorded; Coronavirus (COVID-19); Coronavirus (COVID-19) – household member symptoms; Coronavirus (COVID-19) – post travel quarantine; and Coronavirus (COVID-19) – test and trace contact. The last three categories seemed unlikely to reflect illness in the staff member. However, it was possible that they were not used consistently, and that some absences to which they could have been applied, were instead assigned to the less specific category labelled simply as Coronavirus (COVID-19).

Among the 102,422 sickness absence episodes that began during the study period, and were labelled as Coronavirus (COVID-19), the large majority fell in the collapsed absence categories “Sickness – chest and respiratory” (n = 33,555, mean duration 13 days, median duration 8 days), “Sickness – cough, flu” (n = 27,937, mean duration 12 days, median duration 7 days) and “Sickness – infectious disease” (n= 39,110, mean duration 13 days, median duration 8 days). The similarity of the mean and median durations across the three categories suggested that they had been applied to similar types of illness. In addition, there were two less specific collapsed absence categories, which when associated with a label of Coronavirus (COVID-19), had comparable mean and median durations. These were “Sickness – other” (n= 430, mean duration 15 days, median duration 9 days), and “Sickness – unknown” (n= 668, mean duration 14 days, median duration 8 days). Together, these five categories were therefore aggregated as “Covid-19 sickness absence”.

The label, Coronavirus (COVID-19) also occurred in association with other categories of sickness absence, but these were less frequent, and had different durations (e.g. “Sickness – mental health”, n = 158, mean duration 52 days, median duration 42.5 days, “Sickness – part day”, n = 69, mean duration 1 day, median duration 1 day). It was less certain that such absences indicated an episode of illness from Covid-19, and we therefore excluded them from our definition of “Covid-19 sickness absence”.

Table B1 summarises the durations of the 101,700 episodes of Covid-19 sickness absence which started during the study period. Based on this information, we classed episodes of Covid-19 sickness absence as “prolonged” if their duration exceeded 14 days.

Table B1. Durations of Covid-19 sickness absence

	N	%	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
All	101700	100	1	6	8	12.41	14	144
≤7 days	50664	49.8	1	3	6	5.19	7	7
8-14 days	29414	28.9	8	9	11	11.02	14	14
15-21 days	10645	10.5	15	15	17	17.41	19	21
>21 days	10977	10.8	22	26	33	44.78	52	144

Staff group and exposure category

Both of these variables could vary over time (see Supplementary File A). In the analysis for this paper, we aimed to classify individuals according to the job that they held on 9 March 2020. Where staff group and/or exposure category changed over time, we therefore gave preference in the following order:

- That from the absence record with the most recent start date at 9 March 2020
- If there was no absence starting before 9 March 2020, that from the absence record with the earliest start date after 9 March 2020
- If there were no absences at all, that from the SIP file

References

B1. COVID-19 NHS Situation Reports, COVID-19 daily situation report. Data as reported on 05-Nov-20. Published: 12 November 2020. Downloaded from: https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-hospital-activity/

Supplementary File C

SENSITIVITY ANALYSES EXCLUDING INDIVIDUALS WITH IMPUTED OR MISSING DATA, OR MULTIPLE JOBS

Table C1. Associations of risk factors at baseline with start of an episode of Covid-19 sickness absence (any and prolonged) during 9 March to 31 July 2020 - excluding individuals with imputed or missing data, or multiple jobs

Excludes 6,554 individuals for whom one or more of age, sex or ethnicity was imputed because of inconsistencies in the raw data, 3 with unknown staff group, 44,134 with multiple staff group or exposure categories during the study period (either because they held multiple jobs simultaneously or changed their job), and 1,923 with a missing or imputed end date of an absence. Risk estimates were derived from two logistic regression models that included all of the variables for which results are presented, together with trust (191 categories), and are relative to no Covid-19 sickness absence. An episode of Covid-19 sickness absence was classed as prolonged if it had lasted >14 days by 31 July 2020.

	Any Covid-19 sickness absence					Any prolonged Covid-19 sickness absence			
Risk factor	**Model 1**		**Model 2**			**Model 1**		**Model 2**	
	RR	**(95% CI)**	**RR**	**(95% CI)**		**RR**	**(95% CI)**	**RR**	**(95% CI)**
Sex									
Female	*ref.*	*ref.*	*ref.*	*ref.*		*ref.*	*ref.*	*ref.*	*ref.*
Male	1.01	0.99 - 1.03	1.02	1.00 - 1.04		1.02	0.98 - 1.06	1.03	0.99 - 1.08
Age (years)									
<30	*ref.*	*ref.*	*ref.*	*ref.*		*ref.*	*ref.*	*ref.*	*ref.*
30-34	0.97	0.94 - 1.00	0.97	0.94 - 1.00		1.26	1.17 - 1.36	1.26	1.17 - 1.35
35-39	0.98	0.95 - 1.01	0.99	0.96 - 1.02		1.55	1.44 - 1.66	1.55	1.44 - 1.67
40-44	0.99	0.97 - 1.02	1.00	0.97 - 1.03		1.72	1.61 - 1.84	1.71	1.60 - 1.84
45-49	1.00	0.98 - 1.03	1.01	0.98 - 1.04		2.01	1.88 - 2.15	1.99	1.86 - 2.12
50-54	0.97	0.95 - 1.00	0.98	0.95 - 1.01		2.15	2.01 - 2.29	2.13	2.00 - 2.28
55-60	0.88	0.86 - 0.91	0.88	0.86 - 0.91		2.07	1.94 - 2.21	2.05	1.92 - 2.20
>60	0.75	0.72 - 0.78	0.74	0.72 - 0.77		2.10	1.95 - 2.26	2.06	1.91 - 2.22

	Any Covid-19 sickness absence					Any prolonged Covid-19 sickness absence			
Risk factor	**Model 1**		**Model 2**			**Model 1**		**Model 2**	
	RR	**(95% CI)**	**RR**	**(95% CI)**		**RR**	**(95% CI)**	**RR**	**(95% CI)**
Ethnicity									
White	*ref.*	*ref.*	*ref.*	*ref.*		*ref.*	*ref.*	*ref.*	*ref.*
South Asian	1.42	1.38 - 1.46	1.40	1.36 - 1.44		2.56	2.43 - 2.70	2.49	2.36 - 2.62
Other or unspecified Asian	1.72	1.66 - 1.77	1.64	1.59 - 1.69		2.93	2.78 - 3.10	2.71	2.56 - 2.86
Black	1.15	1.11 - 1.18	1.13	1.10 - 1.17		1.73	1.64 - 1.84	1.69	1.60 - 1.79
Mixed	1.14	1.07 - 1.20	1.13	1.07 - 1.20		1.39	1.24 - 1.56	1.37	1.22 - 1.54
Other	1.48	1.40 - 1.56	1.43	1.36 - 1.51		2.42	2.21 - 2.65	2.29	2.09 - 2.51
Unknown	1.07	1.03 - 1.11	1.07	1.03 - 1.11		1.31	1.21 - 1.42	1.31	1.21 - 1.42
Episodes of sickness absence in 2019									
0	*ref.*	*ref.*	*ref.*	*ref.*		*ref.*	*ref.*	*ref.*	*ref.*
1	1.39	1.36 - 1.42	1.38	1.35 - 1.41		1.49	1.43 - 1.56	1.48	1.42 - 1.55
2-3	1.82	1.78 - 1.86	1.80	1.76 - 1.83		2.00	1.92 - 2.08	1.97	1.89 - 2.05
>3	2.41	2.35 - 2.47	2.37	2.32 - 2.43		2.63	2.50 - 2.77	2.57	2.44 - 2.70
Staff group at 9 March 2020									
Administrative and clerical	*ref.*	*ref.*	*ref.*	*ref.*		*ref.*	*ref.*	*ref.*	*ref.*
Additional clinical services	2.30	2.24 - 2.36	1.62	1.54 - 1.71		2.93	2.78 - 3.09	1.82	1.61 - 2.04
Additional professional scientific and technical	1.36	1.31 - 1.43	1.04	0.98 - 1.12		1.19	1.07 - 1.33	0.97	0.83 - 1.13
Allied health professionals	1.95	1.89 - 2.02	1.33	1.26 - 1.41		1.73	1.60 - 1.87	1.06	0.93 - 1.21
Estates and ancillary	1.44	1.39 - 1.50	1.29	1.24 - 1.34		1.60	1.48 - 1.73	1.39	1.28 - 1.51
Healthcare scientists	1.17	1.10 - 1.24	1.03	0.95 - 1.11		0.96	0.83 - 1.11	0.88	0.73 - 1.05
Medical and dental	1.54	1.49 - 1.60	1.08	1.01 - 1.14		1.05	0.97 - 1.15	0.69	0.60 - 0.79
Nursing and midwifery registered	2.28	2.23 - 2.33	1.57	1.49 - 1.65		2.63	2.49 - 2.76	1.54	1.37 - 1.73
Students	1.88	1.60 - 2.20	1.34	1.13 - 1.59		1.73	1.03 - 2.91	1.19	0.70 - 2.02
Exposure category at 9 March 2020*									
Care of patients much more likely to have Covid-19 than general population	-	-	1.49	1.40 - 1.58		-	-	1.54	1.34 - 1.77
Care for patients who may be more likely to have Covid-19 than general population	-	-	1.42	1.35 - 1.50		-	-	1.71	1.52 - 1.93
Care of patients with similar or lower prevalence of Covid-19 than general population	-	-	1.06	1.00 - 1.12		-	-	1.05	0.93 - 1.18

	Any Covid-19 sickness absence					Any prolonged Covid-19 sickness absence			
Risk factor	**Model 1**		**Model 2**			**Model 1**		**Model 2**	
	RR	**(95% CI)**	**RR**	**(95% CI)**		**RR**	**(95% CI)**	**RR**	**(95% CI)**
No patient care but often in areas where patients have higher prevalence of Covid-19 than general population	-	-	0.75	0.54 - 1.05		-	-	1.12	0.57 - 2.19
No patient care but often in areas where patients have similar or lower prevalence of Covid-19 than general population	-	-	1.29	1.19 - 1.39		-	-	0.97	0.80 - 1.17
No patient care, occasionally in patient areas	-	-	*ref.*	*ref.*		-	-	*ref.*	*ref.*
Unlikely to be in patient areas, but work with material potentially contaminated by coronavirus	-	-	0.92	0.85 - 0.99				0.79	0.67 - 0.94
Other or unknown	-	-	0.73	0.70 - 0.76				0.66	0.60 - 0.72

*Exposure categories are based on the constructed Job Exposure Matrix (JEM). For more information see Supplementary file A.