Impact of cystic fibrosis on birthweight: a population based study of children in Denmark and Wales

Daniela K. Schlüter[1*], Rowena Griffiths[3], Abdulfatah Adam[4], Ashley Akbari[3], Martin Heaven[3], Shantini Paranjothy[5], Anne-Marie Nybo Andersen[4], Siobhán B. Carr[6], Tania Pressler[7], Peter J. Diggle[1], David Taylor-Robinson[2]

Affiliations:

1. Centre for Health Informatics, Computing and Statistics (CHICAS), Lancaster Medical School, Lancaster University, Lancaster, United Kingdom, LA1 4YW
2. Department of Public Health and Policy, Farr Institute, University of Liverpool, Liverpool, United Kingdom, L69 3GB
3. Farr Institute, Swansea University Medical School, Swansea University, United Kingdom, SA2 8PP
4. Section of Epidemiology, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5 Box 2099, DK-1014 Copenhagen K; Denmark
5. Division of Population Medicine, School of Medicine, Cardiff University, 3rd Floor Neuadd Meirionnydd, Health Park, Cardiff CF14 4YS
6. Department of Respiratory Paediatrics, Royal Brompton Hospital, London, UK
7. Copenhagen Cystic Fibrosis Centre, Department of Infectious Diseases, Rigshospitalet, Copenhagen University, Copenhagen, Denmark

[*]Corresponding author; e-mail: d.schlueter@lancaster.ac.uk

Keywords: Cystic fibrosis, birthweight, gestational age, data linkage, causal inference

Word count: 4351

ABSTRACT

Background

Poor growth during infancy and childhood is a characteristic feature of cystic fibrosis (CF). However, the impact of CF on intrauterine growth is unclear. We studied the effect of CF on birthweight in Denmark and Wales, and assessed whether any associations are due to differences in gestational age at birth.

Methods

We conducted national registry linkage studies in two countries, using data for 2.2 million singletons born in Denmark (between 1980 and 2010) and Wales (between 1998 and 2015). We used hospital in- and outpatient data to identify 852 children with CF. Using causal mediation methods we estimated the direct and indirect (via gestational age) effect of CF on birthweight after adjustment for sex, parity, and socio-economic background. We tested the robustness of our results by adjusting for additional factors such as maternal smoking during pregnancy in sub-populations where these data were available.

Results

Babies with CF were more likely to be born preterm and with low birth weight than non-CF babies (12.7% v 5% and 9.4% v 5.8% preterm; 11.9% v 4.2% and 11% v 5.4% low birth weight in Denmark and Wales, respectively). Using causal mediation methods the total effect of CF on birthweight was estimated to be -178.8g (95%CI -225.43g to -134.47g) in the Danish population and -210.08g (95%CI -281.97g to -141.5g) in the Welsh population. About 40% of this effect of CF on birthweight was mediated through gestational age.

Conclusions

CF significantly impacts on intrauterine growth and leads to lower birthweight in babies with CF, which is only partially explained by shorter gestation.

KEY MESSAGES

What is the key question?

Does having cystic fibrosis impact on birthweight and gestational age at birth?

What is the bottom line?

Cystic fibrosis has a large effect on birthweight, similar to the impact of maternal smoking during pregnancy, and is only partially explained by shorter gestation, indicating that having CF impacts directly on intrauterine growth.

Why read on?

Utilizing record-linked data for 2.2 million babies in two countries and modern methods of causal mediation analysis we show that babies with CF are born about 200g lighter than non-CF babies, and only around 40% of this effect is explained by reduced gestational age.

INTRODUCTION

Cystic fibrosis (CF) impacts negatively on the growth of children. Sub-optimal growth is associated with worse lung function and increased risk of premature death in people with CF [1-5]. Around the time of diagnosis, children with CF are on average lighter than non CF children [6] and a number of small studies have suggested that babies with CF may be born lighter than the non-CF population [1 7-14]. A higher prevalence of preterm birth in babies with CF has been suggested as a possible reason [7 13] whereas other studies have suggested that even in babies born at term, those with CF are lighter than those without CF [1] raising the possibility of prenatal and genetic influences on growth. A limitation of these studies has been that they have not been population-based, and have only captured small samples of children with CF from specific regions or seen at individual care centres.

In children with CF, as in the general population, those from socially disadvantaged backgrounds are smaller and lighter compared to children from more affluent families, and have worse health outcomes and survival. In children with CF these associations are evident at diagnosis, and they appear to track forward influencing subsequent outcomes [15]. The socio-economic differences in early growth in children with CF may be a consequence of social differences in birthweight in CF children, as well as differential weight loss by socio-economic status in CF children prior to diagnosis in those that are not diagnosed by new-born screening.

In this study, we aimed to clarify the association of CF with preterm birth and birthweight, and establish whether these associations vary by childhood socio-economic conditions (SECs), at population level in Denmark and Wales. We further aimed to assess whether any differences in birthweight observed were due to differences in gestational age between CF and non-CF babies, or whether there are separate biological processes linked to CF that lead to reduced birthweight.

METHODS

Study design, setting, data sources and participants

We undertook a population level linkage study of all singleton children born in Denmark between 1980 and 2010; and in Wales between 1998 and 2015.

In Denmark we accessed the data from the Danish Medical Birth Register [16] linked to the Danish National Patient Register (DNPR) and to data on highest obtained educational level in Statistics Denmark for all children and their parents. Linkage was based on the Central Person Registry Number (CPR).

In Wales, we analysed data from the Secure Anonymised Information Linkage (SAIL) databank [17 18]. We accessed the National Community of Child Health (NCCH) Database and used Anonymized Linkage Fields to link to the Congenital Anomaly Register and Information Service (CARIS), Patient Episode Database for Wales (PEDW), Welsh Longitudinal General Practice data (WLGP), the Annual District Birth Extracts (ADBE - also known as the Office of National Statistics (ONS) birth register), and the Welsh Demographics Service Dataset (WDSD).

Multiple births, individuals with missing birth weight, gestational age, parity, sex or deprivation score were excluded from the analysis. We also excluded individuals with birthweight less than 100g or greater than 7kg or gestational age less than 21 weeks or greater than 45 weeks, on the basis that these may be data entry errors. See Supplementary Material Section S.1 for more details.

Outcome, exposure and covariates

The main outcomes of interest were birthweight and gestational age; and the main exposure of interest was being coded as having CF compared to not having CF in the linked administrative datasets in each country. Selection of covariates in the analysis was informed a priori using a directed acyclic graph (Figure 1), and this was used to inform similar, but separate, analyses in Denmark and Wales based upon data availability. Further information on the datasets and data cleaning are provided in the Supplementary Material section S.1.

Danish analysis

In Denmark CF diagnosis was extracted from the Danish National Patient Registry (DNPR) [19] which includes hospital in- and outpatient data, identified by ICD 8 code 273 in years prior to 1994 and ICD 10 code E84 thereafter. Birthweight, gestational age, date of birth, sex, parity (from 1996 onwards), mode of delivery, maternal age at birth, and maternal smoking (from 1996 onwards) during pregnancy were obtained from the Danish Medical Birth Register. A binary variable indicating first-born status was derived from the available parity variable post 1996; parity for births prior to 1996 was estimated as the number of previous live-births and still births which reached gestational age greater 22 weeks. Additional available covariates in Denmark were pre-eclampsia or eclampsia during index pregnancy (ICD 8 code 637, ICD 10 code O14 and O15), and diabetes during index pregnancy (ICD 8 code 250, ICD 10 code O24) which were obtained from the Danish National Patient Register. Highest maternal educational level at birth (grouped as ISCED levels 1 to 6 and obtained from the Education Register) was used as a measure of childhood SECs at birth.

Welsh analysis

In Wales we obtained information on diagnosis of CF from the Congenital Anomaly Register and Information Service (CARIS), Patient Episode Database for Wales (PEDW) and Welsh Longitudinal General Practice data (WLGP), identified by ICD 10 code E84 and READ codes 1264., 66k.., 66k0., 9No7., C370. We obtained birth weight, week of birth of the baby, and sex from the Annual District Birth Extracts (ADBE). Other covariates in the analysis included a binary variable indicating first-born status (derived from the number of previous live births given in NCCH); information on the mother's smoking status in the year previous to birth and the mother's week of birth (through WLGP). The mother's age at birth was calculated as difference in days between week of birth of the baby and week of birth of mother divided by 365.25. Deprivation quintiles of small area of residence based on the Welsh Index of Multiple Deprivation (WIMD) was obtained for the Lower Super Output Area (LSOA) of the mother's postcode from the Welsh Demographics Service Dataset (WDSD) and used as a measure of SECs at birth. The LSOAs from the 2001 census were used together with the WIMD scores from 2008.

Statistical Analysis

We used causal mediation methods [20] to estimate the direct effect of CF on birthweight after adjustment for gestational age, and the indirect effect of CF on birthweight due to any effect on gestational age, which in turn influences birthweight (see Figure 1).

We modelled the joint distribution of birth weight and gestational age using a two-component mixture of bivariate linear regression models for the two countries separately. Finite mixture models have previously been used for modelling of gestational age and birthweight in order to capture the skewed shape of their distributions [21 22]. For parameter estimation, we factorised the joint distribution into the distribution of gestational age multiplied by the distribution of birthweight conditional on gestational age.

We initially included CF, sex, parity, and socio-economic status as explanatory variables for both gestational age, and birthweight conditional on gestational age. The significance of each of the covariates was tested by backward elimination based on the likelihood ratio criterion with a cut-off for significance of $p<0.05$. The direct effects of CF on birthweight and gestational age after adjustment for possible confounders were estimated in both sub-populations. The indirect effect of CF on birthweight through gestational age was estimated by the product of the effect of CF on gestational age and the effect of gestational age on birthweight [23]. In the context of linear regression models for both the outcome and the mediator variable, as is the case here, this estimator has been shown to be

identical to the average causal mediation effect or natural indirect effect if identifiability assumptions hold [24 25] (see Supplementary material section S.6 for the assumptions). We estimated the total effects of CF on birthweight in both subpopulations as the sum of the direct and indirect effects [24]. Effects of CF on birthweight and gestational age at a whole population level were estimated as a weighted average of the effects in the subpopulations. 95% confidence intervals for all effect sizes were estimated by non-parametric bootstrap based on 999 samples. We carried out the analysis in R version 3.3.1 using the package FlexMix [26] for fitting the models.

Robustness tests and additional analyses

We performed multiple sensitivity analyses to assess the robustness of our results. We allowed for interaction terms between CF and any of the other covariates and assessed their significance using the likelihood ratio test with cut-off for significance of $p<0.05$. We repeated the analyses: allowing for an interaction term between CF and gestational age; and including just one child per mother to remove any biases that could be introduced due to correlations between outcomes in children from the same mother. In order to assess the influence of unmeasured confounders in our main analysis we repeated our analyses in sub-populations in which we have additional data on age of the mother at birth and smoking status in the year prior to birth; in Denmark, we additionally included data on mode of delivery, pre-eclampsia or eclampsia, and diabetes during pregnancy. In order to assess the impact of potential misclassification of CF cases, we repeated the analyses using stricter classification criteria for CF. In Denmark we classified those as CF that had a CF code in the Danish National Patient Register and that had been seen at a hospital more than once. In Wales we selected only those individuals that were coded as CF in CARIS. In an additional post-hoc analysis using our final model, we estimated the probability for a CF baby compared to a non-CF baby to be born with low birthweight, since this may also be of clinical interest. See Supplementary Material section S.8 for details.

Ethics and Information Governance

We use anonymised data in this study, therefore specific ethics approval was not needed. Approval for the Welsh analysis was granted from the Health Information Research Unit (HIRU) Information Governance Review Panel (IGRP). Use of the de-identified Danish linked register data from Statistics Denmark was approved by the Danish Data Protection Agency.

RESULTS

Population Characteristics

Figure 2 shows the derivation of the final study populations. In Denmark and Wales 597 out of 1,736,186 children and 255 out of 442,409 had CF codes, respectively. The baseline characteristics of both study populations are given in Table 1.

Table 1: Demographics of the study population.

	Wales		Denmark	
	Non-CF population	CF population	Non-CF population	CF population
N	442,409	255	1,736,186	597
Birth weight in g (median [IQR])	3400.00 [3060.00, 3740.00]	3200.00 [2845.00, 3567.50]	3500.00 [3150.00, 3850.00]	3300.00 [2900.00, 3700.00]
Low birth weight =yes (%)	24,035 (5.4)	28 (11.0)	72,480 (4.2)	71 (11.9)
Gestational age in weeks (median [IQR])	40.00 [39.00, 41.00]	39.00 [38.00, 40.00]	40.00 [39.00, 41.00]	39.71 [38.00, 40.43]
Preterm=yes (%)	25,770 (5.8)	24 (9.4)	87,317 (5.0)	76 (12.7)
Sex =male (%)	215,238 (48.7)	116 (45.5)	891,336 (51.3)	320 (53.6)
First born =yes (%)	205,805 (46.5)	122 (47.8)	523,457 (30.1)	179 (29.8)
Mother's age at birth (median [IQR])	27.94 [23.67, 32.35]	27.06 [23.07, 31.78]	29.00 [25.00, 32.00]	28.00 [24.00, 31.00]
Mother smoked year prior pregnancy				
No (%)	79,208 (17.9)	48 (18.8)	911,436 (52.5)	239 (40.0)
Yes (%)	42,449 (9.6)	20 (7.8)	286,993 (16.5)	72 (12.1)
NA (%)	320,752 (72.5)	187 (73.3)	537,757 (31.0)	286 (47.9)
WIMD quintile[1]				
1 (%)	109,986 (24.9)	68 (26.7)	NA	NA
2 (%)	98,595 (22.3)	58 (22.7)	NA	NA
3 (%)	87,144 (19.7)	36 (14.1)	NA	NA
4 (%)	76,076 (17.2)	48 (18.8)	NA	NA
5 (%)	70,608 (16.0)	45 (17.6)	NA	NA
Maternal education[2]				
Primary or lower secondary (%)	NA	NA	490,215 (28.2)	218 (36.5)
Upper secondary or post secondary non tertiary (%)	NA	NA	733,037 (42.2)	233 (39.0)
Tertiary education or not known (%)	NA	NA	512,934 (29.5)	146 (24.5)
Diabetes =no (%)	NA	NA	1,734,670 (99.9)	590 (98.8)
Pre-eclampsia =no (%)	NA	NA	1,733,619 (99.9)	*
Mode of delivery				
Vaginal (%)	NA	NA	1,249,816 (72.0)	402 (67.3)
Elective caesarean section (%)	NA	NA	79,127 (4.6)	21 (3.5)
Emergency caesarean section (%)	NA	NA	83,866 (4.8)	21 (3.5)
NA (%)	NA	NA	323,377 (18.6)	153 (25.6)

[1] Welsh Index of Multiple Deprivation where 1=most deprived and 5=least deprived

[2] Original levels are grouped together as follows "Primary education, first stage of basic education" and "Lower secondary education, second stage of basic education" are grouped into "Primary or lower secondary education"; "Upper secondary education" and "Post-secondary, non tertiary education" are grouped into "Upper secondary or post secondary non tertiary"; "First stage of tertiary education", "Second stage of tertiary education", and "Not known" are grouped into "Tertiary education or not known".

* omitted due to cells containing five or fewer individuals or because it would allow the calculation of a cell with five or fewer individuals.

In Denmark 11.9% of the CF population had low birthweight (<2.5kg) compared to 4.2% in the non-CF population; in Wales 11% compared to 5.4% had low birthweight. In Denmark 12.7% of the CF coded babies were preterm (<37 weeks) compared to 5% in the non-CF population. In Wales 9.4% of the CF coded babies were preterm compared to 5.8% in the non CF population. Socio-economic position, first born status, gender, age of the mother at birth, smoking in the year prior to birth, mode of delivery and rates of diabetes and pre-eclampsia were distributed similarly in the CF and non-CF coded populations.

The effect of cystic fibrosis on birth weight and gestational age

The estimated whole population effects are discussed in this section; for details on the mixture sub-populations and the estimated effects at the level of the sub-populations see Supplementary Material section S.2 and S.4, respectively. All explanatory variables included in the model were significant at the 5% level (see Supplementary Material section S.3). Tables 2 and 3 give the estimated effect sizes of the explanatory factors on gestational age and birthweight, respectively.

Table 2: Estimated effects on gestational age in weeks.

	Wales	Denmark
CF (95% CI)	-0.57 (-0.83 , -0.33)	-0.41 (-0.55 , -0.25)
First born = yes (95% CI)	0.16 (0.15 , 0.17)	0.03 (0.02 , 0.03)
Sex =male (95% CI)	-0.06 (-0.07 , -0.05)	-0.04 (-0.05 , -0.04)
WIMD (95% CI) [1]		
1	-0.11 (-0.13 , -0.09)	NA
2	-0.05 (-0.07 , -0.03)	NA
3	-0.02 (-0.04 , 0)	NA
4	0.01 (-0.01 , 0.03)	NA
Maternal education (95% CI)		
Lower secondary education, second stage of basic education	NA	0.01 (-0.02 , 0.05)
Upper secondary education	NA	0.15 (0.12 , 0.18)
Post secondary non tertiary education	NA	0.28 (0.09 , 0.46)
First stage of tertiary educ., not adv. research qualification	NA	0.23 (0.2 , 0.26)
Second stage of tertiary educ., adv. research qualification	NA	0.28 (0.21 , 0.35)
Not known	NA	0.23 (0.15 , 0.32)
[1] Welsh Index of Multiple Deprivation where 1=most deprived and 5=least deprived		

Table 3: Estimated effects on birth weight in grams

	Wales	Denmark
Gestational age (95% CI) (g per week)	139.58 (137.86 , 141.21)	141.42 (140.85 , 142.2)
CF direct (95% CI)	-127.24 (-182 , -72.61)	-116.09 (-152.53 , -80.5)
CF indirect (95%CI)	-82.85 (-122.22 , -44.91)	-62.71 (-84.47 , -39.96)
CF total (95% CI)	-210.08 (-281.97 , -141.5)	-178.8 (-225.43 , -134.47)
First born (95% CI)	-115.21 (-117.84 , -112.54)	-113.57 (-115.38 , -112.45)
Sex =male (95% CI)	133.18 (130.34, 135.91)	133.92 (132.49 , 135.34)
WIMD (95% CI) [1]		

1	-114.34 (-118.59 , -109.94)	NA
2	-71.09 (-75.33 , -66.99)	NA
3	-39.73 (-44.15 , -35.34)	NA
4	-19.24 (-24.01 , -14.68)	NA
Maternal education (95% CI)		
Lower secondary education, second stage of basic education	NA	0.54 (-7.08 , 8.48)
Upper secondary education	NA	86.16 (78.48 , 93.76)
Post secondary non tertiary education	NA	113.15 (71.9 , 155.69)
First stage of tertiary educ., not adv. research qualification	NA	124.25 (116.21 , 131.96)
Second stage of tertiary educ., adv. research qualification	NA	142.84 (124.75 , 159.22)
Not known	NA	-32.11 (-52.59 , -10.55)
[1] Welsh Index of Multiple Deprivation where 1=most deprived and 5=least deprived		

Findings in the Danish Population

In the Danish population babies with a CF diagnosis were on average born about 3 days earlier than non-CF babies (-0.41 weeks; 95%CI -0.55 weeks to -0.25 weeks). Parity and sex had negligible effects on gestational age whereas gestational age increased by about 2 days when comparing babies born to mothers with primary education and those born to mothers with second stage tertiary education (see Table 2).

In total, CF babies in the Danish population were 178.8g (95%CI 134.47g to 225.43g) lighter than non-CF babies. Approximately 35% of the effect of CF on birthweight was mediated by gestational age. First born and female babies were significantly lighter than not first born and male babies and birthweight increased by almost 150g when comparing babies from mothers with primary education to those born to mothers with second stage tertiary education (see Table 3).

Findings in the Welsh Population

In the Welsh population CF coded babies were born approximately 4 days earlier than non CF babies (-0.57 weeks; 95%CI -0.83 weeks to -0.33 weeks). First-borns were found to be born slightly earlier than non-first-borns. Sex had a negligible effect on gestational age. The estimated increase in gestational age for babies from the least compared to the most deprived mothers was less than one day. See Table 2 for details.

CF coded babies in Wales were estimated to be born 210.08g (95%CI 141.5g to 281.97g) lighter than non-CF babies with approximately 39% of the effect being due to reduced gestation. First-borns and

females were significantly lighter than non-first-born and male babies. Birthweight increased by approximately 115g when comparing the most to the least deprived quintile of babies (see Table 3).

Robustness tests and additional analyses

Although no interaction effects between CF and any of the covariates were statistically significant ($p<0.05$), we found that the interaction between CF and first-born status was only marginally above the threshold level in both populations (see Supplementary Material section S.5).

We found a significant interaction term between CF and gestational age in the birthweight sub-model for the Danish population. However, the estimated direct and indirect effects of CF did not change markedly when this interaction term was introduced (see Supplementary Material section S.6 for details). Results were similar in analyses including only one baby from each mother in both countries (Supplementary Material section S.7). Repeating the analysis with additional covariates or using a stricter definition of CF cases did not markedly change our results (Supplementary Material S.7).

On the basis of simulations using our final model we found that the probability for CF babies to be born with low birthweight was between 1.3 and 1.8; and between 1.2 and 2.1 times that of non-CF babies in Wales and Denmark, respectively, and depended and sex, first-born status, deprivation and gestational age. The ratio between the probabilities of being born LBW for CF and non-CF babies increased with increasing gestational age. Being female, first born or from a more a deprived area increased the probability of being born with low birthweight only slightly and did not significantly affect the ratios between CF and non-CF babies. See Supplementary Material section S.8 for the effect of covariate values on the probabilities for a CF baby to be born with low birthweight and the estimated ratios of the probabilities for CF compared to non-CF babies.

DISCUSSION

In a whole population linkage study in Wales and Denmark, we showed that babies coded as having CF are born on average about 200g lighter than non-CF babies. This is a large difference at a population level, whereby the total effect of CF on birthweight is similar to the impact of maternal smoking during pregnancy [27]. Babies with CF are born on average about half a week earlier, but this only accounts for around 40% of the total effect of CF on birthweight, suggesting a significant direct biological impact of CF on intrauterine growth. Babies from socially disadvantaged backgrounds are significantly lighter than those from more affluent /more educated families, but this effect was the same in CF and non-

CF babies. Similarly, female babies are born significantly lighter than male babies and first-borns are born lighter than non-first-borns.

Comparison with other studies

Our large study looking at births from two separate countries corroborates previous smaller studies that suggested a difference in birthweight between CF and non-CF babies [1 7-14]. Festini and colleagues [7] compared perinatal data from 70 children with CF in Tuscany, Italy, to regional population samples over an 11 year period. The authors found that overall babies with CF were born 246.2g lighter (95%CI 129.8g to 362.5g) than non-CF babies. In babies born at term (>37weeks gestation) the difference was 205.7g (95%CI 95.4g to 315.9g). In another study Darrah and colleagues [1] compared the birth weight of 79 patients with CF born at term with pancreatic insufficiency and cared for at the Cystic Fibrosis Center at Rainbow Babies and Children's Hospital in Cleveland, Ohio between 1975 and 2005 to the national average. In the study population male CF babies weighed an average 3,239.80g (standard deviation= 367.83g, N=40) compared to a national average of 3,530.20g; female CF babies weighed 3,142.94g (standard deviation =422.75g, N=39) compared to a national average of 3,399.19g. Recently Ramos and colleagues [13] conducted a study in Washington State comparing the birthweight of 170 babies with CF to that of 3400 non-CF babies matched by birth year and born between 1996 and 2013. They found mean birthweights of 3031g (standard deviation=759g) and 3387g (standard deviation = 581 g) for babies with and without CF, respectively. Earlier studies had found comparable results [10-12].

Our results from Denmark and Wales show that the difference in birthweight between CF and non-CF babies is about 200g, similar to the estimates from these smaller studies. We extend these findings to assess the contribution of gestational age to explaining these effects. Our study highlights for the first time that there is a biological effect of CF on birthweight distinct from an impact of CF on gestational age. Different plausible biological explanations have been suggested, one of which being an association between CFTR mutation and reduced levels of insulin-like growth factor 1 (IGF1). This association has been demonstrated in CF pig models, with reduced levels of insulin-like growth factor 1 (IGF1) associated with reduced bone length and bone mineral content at birth [28]. The same study also showed that new-born humans with CF also have reduced levels of IGF1. In addition, a study in CF mice found strong correlations between IGF1 levels and weight [29]. However, the study also showed that growth deficiency in CF mice was evident late in gestation whereas, in contrast to the pig model, IGF1 levels were comparable to those in control mice prenatally and at birth but reduced at 3 weeks. It is therefore unlikely that reduced IGF1 levels are the sole cause for reduced weight at birth in CF. The study was carried out using CF mice as well as gut-corrected CF mice that did not present intestinal obstruction. Results were equivalent in both types of CF mice with similar growth retardation

compared to control, thus also ruling out a prominent role of intestinal obstruction in growth inhibition in CF mice. Further investigation into a possible role of the placenta in CF-related growth deficiency showed that although aquaporin expression was altered in CF mice, there was no evidence of an effect on placental fluid exchange[29]. Further research is needed to fully understand the effect of CFTR mutations on placental function to assess its involvement in prenatal development in CF. Further studies are also needed to assess the impact of CF on different aspects of growth, for instance birth length, which is not routinely collected in population registry data.

Clinical implications

More research is needed to assess the prognostic value of birthweight for subsequent outcomes in CF. Nutritional status and growth in the early years are closely linked to lung function, which subsequently influences survival in CF [2-5]. In a study of 79 CF patients birthweight has been shown to be associated with pulmonary function at age 6 and age 10 with FEV_1% increasing by an estimated 1% per additional 100g birthweight at age 6 [1].

Within the CF study population, we found an average difference in birthweight of over 100g between babies from the least and the most deprived families in both Wales and Denmark. Children with CF from the most deprived areas in the UK have been found to weigh less, be shorter, have a lower body mass index, be more likely to have chronic Pseudomonas aeruginosa infection and have lower lung function compared to CF children from the least deprived areas [15]. Similar results have also been found in the United States [31]. Our study suggests that these differences may to some extent be explained by differences in birthweight. This indicates that social inequalities in CF outcomes may start in the intrauterine period.

Strengths and limitations

A key strength is that our study made use of routinely collected data, which led to an unselected population-based cohort of around 2.2 million babies with 852 CF cases. A further strength is that we were able to carry out the analysis in two countries with remarkably consistent findings in both national populations. A wide range of information is collected in the Welsh and Danish registry linkages, allowing us to adjust for a range of variables in our analysis. A further strength compared to previous studies is that we have used modern methods for causal mediation analysis to better understand the role of gestational age in the pathway from CF to low birthweight. Misclassification of CF cases is a potential limitation of our analysis. CF is a lifelong condition that requires intensive healthcare support over the patient's life, and all patients can be expected to receive hospital-based care at some point. It is therefore likely that all true CF cases are captured in both countries. In Wales

we were able to identify cases on the basis of both hospital in-and outpatient data and GP records. In addition, a universal new-born screening program was introduced in the UK (but not in Denmark) in 2007 leading to early diagnosis of CF in Wales and capture in the CARIS dataset. Any misclassification of CF cases identified in routine administrative datasets is expected to be independent of birthweight and gestational age and would, as such, lead to a conservative estimate of the effects on birthweight and gestational age. Robustness tests, selecting cases most likely to be true CF cases showed similar results to our main analysis.

In order to be able to estimate the direct and indirect effects of CF on birthweight we made four assumptions to ensure identifiability of the natural indirect effect (see Supplementary Material section S.6). If these assumptions do not hold there is the potential to introduce bias in the estimated effects, which is a well-known problem in causal inference (see for example [32 33]). In order to minimise the risk of biased estimates we included many of the well-established factors that affect birthweight and gestational age [34] in the analysis and conducted robustness tests in which we included further variables that were only available in sub-sets of the study population. Our results did not change, increasing our confidence that omission of these variables in the main analysis did not affect our findings markedly. However, we were not able to adjust for variables such as pre-pregnancy BMI, and weight gain during pregnancy, which are not collected in the registry data that could potentially lead to mediator-outcome confounding. Similarly we were not able to adjust for ethnicity, which may not only be a potential confounder but which may also plausibly modify the relationship between CF and birthweight. The non-white populations of both Wales and Denmark are small (7% and 10%, respectively) and they may therefore be ill-suited to give insights into CF-ethnicity interactions. Since non-white ethnicity is associated with worse outcomes in CF [35-37], it may, however, be of interest to explore this in future studies in other populations. One further point to consider is that gestational age is prone to measurement error or interval censoring. This may have the effect of reducing the association between gestational age and birthweight and therefore the indirect effect of CF while increasing the direct effect [38]. Due to the size of our study population we do not believe that this is a major concern in this study; it should; however, be taken into account when interpreting the findings.

Implications for research and practice

Birthweight and gestational age are not currently collected in most CF Registry datasets, including the UK, US and Danish databases. Based on our study results, we recommend the addition of birthweight to the list of variables collected in CF registries. This will allow further longitudinal studies to understand the association of birthweight with trajectories of nutritional status and lung function and its association with survival. Further research is also needed to understand the biological mechanisms

that lead to reduced intrauterine growth in CF compared to non-CF babies. Finally, our study has shown the merits of data linkage, without which we would not have been able to adjust for all the important confounders and identify the direct and indirect effects. In order to further our understanding of disease progression in CF, linkage of CF registries to other routinely collected data sources should be considered.

In conclusion, our findings from the analysis of national populations in Wales and Denmark suggests that CF has a significant effect on birthweight, and that this is only partially due to an effect on gestational age.

ROLE OF THE FUNDING SOURCE

The work was supported by the Strategic Research Centre "EpiNet: Harnessing data to improve lives" funded by the UK Cystic Fibrosis Trust. DTR was also funded by the MRC on a Clinician Scientist Fellowship (MR/P008577/1). The funder was not involved in the study design, data collection, data analysis, data interpretation, or in the writing of the report.

CONTRIBUTORS

DKS and DTR conceived the original idea for this study. DKS, RG, AAdam, AAkbari, MH, SP, AMNA, PJD and DTR designed the study. DKS, PJD and DTR developed the analysis plan. RG and AAdam extracted the data and prepared the datasets. DKS analysed the data and conducted the literature searches. SBC and TP helped identify previous work and gave the clinical interpretation. DKS and DTR wrote the first draft of the paper. All authors were involved in interpreting the findings and revising drafts and agreeing the final version.

DECLARATIONS OF INTEREST

No conflicts of interest exist.

ACKNOWLEDGEMENTS

This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) system, which is part of the national e-health records research infrastructure for Wales. We would like to acknowledge all the data providers who make anonymised data available for research.

We acknowledge the support from The Farr Institute of Health Informatics Research. The Farr Institute is supported by a 10-funder consortium: Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute of Health Research, the

National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), the Wellcome Trust, (MRC Grant Nos: CIPHER MR/K006525/1, HeRC MR/K006665/1)

REFERENCES

1. Darrah R, Nelson R, Damato EG, et al. Growth Deficiency in Cystic Fibrosis Is Observable at Birth and Predictive of Early Pulmonary Function. *Biol Res Nurs* 2016;18(5):498-504. doi: 10.1177/1099800416643585
2. Milla CE. Nutrition and lung disease in cystic fibrosis. *Clin Chest Med* 2007;28(2):319-30. doi: 10.1016/j.ccm.2007.02.006 [published Online First: 2007/05/01]
3. Snell GI, Bennetts K, Bartolo J, et al. Body mass index as a predictor of survival in adults with cystic fibrosis referred for lung transplantation. *J Heart Lung Transplant* 1998;17(11):1097-103. [published Online First: 1998/12/17]
4. Szwed A, John A, Gozdzik-Spychalska J, et al. Survival of Patients with Cystic Fibrosis Depending on Mutation Type and Nutritional Status. *Adv Exp Med Biol* 2017 doi: 10.1007/5584_2017_66 [published Online First: 2017/07/20]
5. Sanders DB, Fink A, Mayer-Hamblett N, et al. Early Life Growth Trajectories in Cystic Fibrosis are Associated with Pulmonary Function at Age 6 Years. *J Pediatr* 2015;167(5):1081-8 e1. doi: 10.1016/j.jpeds.2015.07.044 [published Online First: 2015/09/06]
6. Patel L, Dixon M, David TJ. Growth and growth charts in cystic fibrosis. *J R Soc Med* 2003;96 Suppl 43:35-41.
7. Festini F, Taccetti G, Repetto T, et al. Gestational and neonatal characteristics of children with cystic fibrosis: a cohort study. *J Pediatr* 2005;147(3):316-20. doi: 10.1016/j.jpeds.2005.04.031
8. Haeusler G, Frisch H, Waldhor T, et al. Perspectives of longitudinal growth in cystic fibrosis from birth to adult age. *Eur J Pediatr* 1994;153(3):158-63.
9. Ranganathan SC, Bush A, Dezateux C, et al. Relative ability of full and partial forced expiratory maneuvers to identify diminished airway function in infants with cystic fibrosis. *Am J Respir Crit Care Med* 2002;166(10):1350-7. doi: 10.1164/rccm.2202041
10. Hsia DY. Birth weight in cystic fibrosis of the pancreas. *Ann Hum Genet* 1959;23:289-99. [published Online First: 1959/07/01]
11. Boyer PH. Low birth weight in fibrocystic disease of the pancreas. *Pediatrics* 1955;16(6):778-84. [published Online First: 1955/12/01]
12. Muller AE, Thamm B, Lietz T, et al. Cystic fibrosis: a cause of reduced birth weight? *Eur J Pediatr* 1999;158(3):264. [published Online First: 1999/03/27]
13. Ramos KJ, Sack CS, Mitchell KH, et al. Cystic Fibrosis is Associated with Adverse Neonatal Outcomes in Washington State, 1996-2013. *J Pediatr* 2017;180:206-11 e1. doi: 10.1016/j.jpeds.2016.09.069 [published Online First: 2016/10/30]
14. Leung DH, Heltshe SL, Borowitz D, et al. Effects of Diagnosis by Newborn Screening for Cystic Fibrosis on Weight and Length in the First Year of Life. *JAMA Pediatr* 2017;171(6):546-54. doi: 10.1001/jamapediatrics.2017.0206 [published Online First: 2017/04/25]
15. Taylor-Robinson DC, Smyth RL, Diggle PJ, et al. The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: a longitudinal study. *Lancet Respir Med* 2013;1(2):121-8. doi: 10.1016/S2213-2600(13)70002-X
16. Knudsen LB, Olsen J. The Danish Medical Birth Registry. *Dan Med Bull* 1998;45(3):320-3. [published Online First: 1998/07/24]

17. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;9:3. doi: 10.1186/1472-6947-9-3 [published Online First: 2009/01/20]
18. Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9:157. doi: 10.1186/1472-6963-9-157 [published Online First: 2009/09/08]
19. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health* 2011;39(7 Suppl):30-3. doi: 10.1177/1403494811401482 [published Online First: 2011/08/04]
20. VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction: Oxford University Press 2015.
21. Gage TB. Classification of births by birth weight and gestational age: an application of multivariate mixture models. *Ann Hum Biol* 2003;30(5):589-604.
22. Schwartz SL, Gelfand AE, Miranda ML. Joint Bayesian analysis of birthweight and censored gestational age using finite mixture models. *Stat Med* 2010;29(16):1710-23. doi: 10.1002/sim.3900
23. Baron RM, Kenny DA. The Moderator Mediator Variable Distinction in Social Psychological-Research - Conceptual, Strategic, and Statistical Considerations. *J Pers Soc Psychol* 1986;51(6):1173-82. doi: Doi 10.1037/0022-3514.51.6.1173
24. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Stat Sci* 2010;25(1):51-71. doi: 10.1214/10-Sts321
25. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface* 2009;2:457-68.
26. Leisch F. FlexMix: A general framework for finite mixture models and latent classs regression in R. *Journal of Statistical Software* 2004;11(8)
27. Abraham M, Alramadhan S, Iniguez C, et al. A systematic review of maternal smoking during pregnancy and fetal measurements with meta-analysis. *PLoS One* 2017;12(2):e0170946. doi: 10.1371/journal.pone.0170946 [published Online First: 2017/02/24]
28. Rogan MP, Reznikov LR, Pezzulo AA, et al. Pigs and humans with cystic fibrosis have reduced insulin-like growth factor 1 (IGF1) levels at birth. *Proc Natl Acad Sci U S A* 2010;107(47):20571-5. doi: 10.1073/pnas.1015281107
29. Darrah R, Bederman I, Vitko M, et al. Growth deficits in cystic fibrosis mice begin in utero prior to IGF-1 reduction. *PLoS One* 2017;12(4):e0175467. doi: 10.1371/journal.pone.0175467 [published Online First: 2017/04/07]
30. Hortencio TD, Nogueira RJ, Marson FA, et al. [Factors impacting the growth and nutritional status of cystic fibrosis patients younger than 10 years of age who did not undergo neonatal screening]. *Rev Paul Pediatr* 2015;33(1):3-11. doi: 10.1016/j.rpped.2014.11.004 [published Online First: 2015/02/15]
31. Schechter MS, Shelton BJ, Margolis PA, et al. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med* 2001;163(6):1331-7. doi: 10.1164/ajrccm.163.6.9912100 [published Online First: 2001/05/24]
32. VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on Intermediates in Perinatal Epidemiology. *Epidemiology* 2012;23(3):507-07. doi: 10.1097/EDE.0b013e31824fceb8
33. Imai K, Keele L, Tingley D, et al. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *Am Polit Sci Rev* 2011;105(4):765-89. doi: 10.1017/S0003055411000414
34. Goldenberg RL, Culhane JF, Iams JD, et al. Epidemiology and causes of preterm birth. *Lancet* 2008;371(9606):75-84. doi: 10.1016/S0140-6736(08)60074-4 [published Online First: 2008/01/08]

35. Hamosh A, FitzSimmons SC, Macek M, Jr., et al. Comparison of the clinical manifestations of cystic fibrosis in black and white patients. *J Pediatr* 1998;132(2):255-9. [published Online First: 1998/03/20]
36. O'Connor GT, Quinton HB, Kahn R, et al. Case-mix adjustment for evaluation of mortality in cystic fibrosis. *Pediatr Pulmonol* 2002;33(2):99-105. [published Online First: 2002/01/22]
37. McColley SA, Schechter MS, Morgan WJ, et al. Risk factors for mortality before age 18 years in cystic fibrosis. *Pediatr Pulmonol* 2017;52(7):909-15. doi: 10.1002/ppul.23715 [published Online First: 2017/04/25]
38. Hoyle RH, Kenny DA. Sample size, reliability, and tests of statistical mediation. In: Hoyle RH, ed. Statistical strategies for small sample research. Thousand Oakes, CA: Sage 1999:195-222.