

Original Research Article

How do growth and nutrition explain social inequalities in lung function in children with cystic fibrosis? A longitudinal mediation analysis using interventional disparity effects with time-varying mediators and intermediate confounders.

Daniela K. Schlüter¹, Ruth H. Keogh², Rhian M. Daniel³, Schadrac C. Agbla⁴, David Taylor-Robinson¹

¹Department of Public Health, Policy and Systems, University of Liverpool, Liverpool, UK

²Department of Medical Statistics and Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK

³Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK

⁴Department of Health Data Science, University of Liverpool, Liverpool, UK

Corresponding author

Daniela K. Schlüter
Department of Public Health, Policy and Systems
University of Liverpool
Liverpool
UK
e-mail: d.k.schlueter@liverpool.ac.uk

Sources of Financial Support

DKS and DTR were supported by the Strategic Research Centre “CF-EpiNet: Harnessing data to improve lives” funded by the Cystic Fibrosis Trust. DTR is funded by the MRC on a Clinician Scientist Fellowship (MR/P008577/1). RHK is supported by a UKRI Future Leaders Fellowship (MR/S017968/1). RMD is supported by a Sir Henry Dale Fellowship from the Wellcome Trust and The Royal Society (107617/Z/15/Z).

Conflicts of interest

None declared

Process by which someone else could obtain the data and computing code

The data used in this study are from the UK CF Registry and are available from the UK CF Registry on reasonable request. The computing code can be found in the supplementary material.

Acknowledgements

We thank the UK CF Registry team, and the UK CF centres and clinics for submitting data to the Registry. Special thanks to the people with cystic fibrosis and their families who have agreed for their UK CF Registry data to be used for research.

Word count: 3975

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Background

Deprivation is associated with poorer growth, worse lung function and shorter life expectancy in children with cystic fibrosis (CF). While early growth is associated with lung function when first measured at around age 6, it is unclear whether improving early growth in the most disadvantaged children would reduce inequalities in lung function.

Methods

We used data from children born 2000-2010 and followed up to 2016 in the UK CF Registry. To estimate the association between deprivation and lung function at around age six, and the causal contribution of early weight trajectories, we extended the mediation analysis approach based on interventional disparity effects to the setting of a longitudinally measured mediator. We adjusted for baseline confounding by sex, birthyear and genotype and accounted for time-varying intermediate confounding by lung infection.

Results

853 children were included in the study, including 165 and 172 children from the least and most deprived population quintiles, respectively. The average difference in lung function between the least and most deprived quintile of children, was 4.51 percent of predicted forced expiratory volume in one second (95% CI: 1.08 - 7.93). We estimated this would be reduced to 3.97 percentage points (95% CI: 0.57 – 7.38) if early weight trajectories in the most deprived children were shifted to the distribution observed in the least disadvantaged children.

Conclusion

Socio-economic conditions are strongly associated with lung function for children with CF which we estimated would only be marginally reduced if early weight trajectories could be improved for the most disadvantaged children.

Keywords: cystic fibrosis, health inequalities, health disparities, mediation analysis, interventional disparity measures, interventional effects

Introduction

Cystic fibrosis (CF) is a serious inherited condition. 1 in 2,500 babies born in the UK have CF, with over 10,000 people living with CF in the UK today. CF is caused by variants of the CF transmembrane conductance regulator (CFTR) gene, resulting in the build-up of thick, sticky mucus in various organs, in particular those of the respiratory and digestive systems. Most people with CF die prematurely from their disease through respiratory failure.

Growth and nutrition in early life are associated with lung function at age six for children with CF and interventions aimed at improving growth in early life may improve later pulmonary function¹. The respiratory system develops from shortly after conception until adolescence and matures most rapidly in size and intricacy in the first three years. Therefore, adequate nutrition during this phase is deemed essential. In children with CF diagnosed within 6 months after birth and who were malnourished at diagnosis, catch-up growth by age two is associated with improved lung function at age six, while maintenance of growth from ages two to six has no strong additional effect on pulmonary function².

Children with CF from socio-economically disadvantaged areas in the UK have worse outcomes than those from more affluent areas³⁻⁵. Socio-economic inequalities in weight and nutritional status have been reported from infancy and inequalities in lung function are evident as soon as this can be routinely measured at around age six³. Although there is some evidence from other countries that inequalities may widen with age^{6,7}, no evidence has emerged that this is the case in the UK⁸. Therefore, understanding the pathways that lead to inequalities in lung function at age six is crucial to identifying potential targets for interventions that could reduce inequalities in CF across the lifecourse.

Achieving and maintaining healthy growth and nutrition is difficult for many people with CF due to fat malabsorption, increased energy requirements, reduced oral energy intake and psychosocial factors⁹. Thus dietary management has been found to be a difficult task for children with CF and

their parents¹⁰, further compounded by socio-economic adversity. Studies in the USA have shown that parental CF-specific nutritional knowledge is a strong predictor of nutritional outcomes¹¹ and also strongly associated with parental income and level of education¹². In addition, there has recently been an increased awareness of high levels of food insecurity in people with CF in the USA^{12,13}. Patterns are likely similar in the UK although there are currently no data on the number of children with CF that live with food insecurity.

In this study we examined how growth and nutrition in the first six years of life can explain inequalities in lung function (when first measured around age six) between groups with different levels of deprivation in the UK CF population. More specifically, we were interested in estimating the association between deprivation at birth and lung function at around age six, and the extent to which this could be reduced if we were to intervene to improve the weight trajectories in the most deprived children such that they have the same distribution as those in the least deprived children.

To estimate the effects of interest, we build on recent developments in the causal inference literature that have led to the definition of “interventional disparity effects”¹⁴ and extend this approach to the mediation setting with a longitudinally measured mediator.

Methods

Data Source

We used UK CF Registry data, which collects data at approximately annual clinic visits when a patient is clinically stable. Records date back to the 1990s and are estimated to capture 99% of the current UK CF population¹⁵. Our data included follow-up to the end of 2016. Annual review visits include evaluation of clinical and nutritional status, lung function, and microbiology of respiratory tract secretions since the last annual review. Demographic data, including socio-economic status, are also recorded.

Setting and Participants

We included children who were born after 2000, diagnosed by newborn screening, had at least one lung function measure between ages ≥ 6 and < 9 years and at least one weight measurement between birth and < 7 years of age, i.e. before or at the same time as the lung function measure. We excluded individuals who did not have complete data on sex, genotype, year of birth and socio-economic conditions (SECs).

Outcome, exposure, mediator, and confounders

Our outcome of interest was the first available lung function measurement given by percent of predicted forced expiratory volume in one second (%FEV1) based on GLI reference equations, which adjust for sex, age, height and ethnicity¹⁶(continuous variable).

Our exposure of interest was SECs at birth. Measures of individual SECs are not available in the UK CF Registry. Therefore, we used the Index of Multiple Deprivation (IMD), a small area deprivation measure, based on first home-address postcode recorded in the registry, as proxy for SECs at birth. The IMD is a measure of relative deprivation which combines information on income, employment, health, education, access to services, crime/safety and environment/housing^{17–20}. The domains vary between England, Wales, Scotland and Northern Ireland. In order to obtain a comparable measure of deprivation across countries, we calculated country-specific IMD z-scores⁵. All children were included in the study but our main interest was the difference in outcome between children in the highest and lowest quintile of deprivation (see Supplementary Material for details). We therefore converted z-scores to quintiles.

Our mediator of interest was early weight measured at annual review visits from birth up to but excluding age 7. We used weight-for-age z-scores based on the UK WHO reference²¹ as a continuous variable.

Confounders were identified based on our assumed causal relationships between SECs, weight, lung function and their common causes (Figure 1). We included sex (male/female), birthyear (continuous)

and genotype (categorical: number of F508del alleles). Respiratory infection may be a time-varying confounder of the mediator-outcome relationship affected by the exposure³. We only considered infection by *Pseudomonas aeruginosa*, the most common cause of infection in the UK CF population under study²². At any annual review visit, it is recorded whether an infection with *P. aeruginosa* had been present since the last annual review visit.

Statistical methods

Rationale and effects of interest

We were interested in estimating the association between SECs at birth (high versus low) and lung function at around age six and the extent to which it can be explained by early weight trajectories. Figure 1 illustrates the assumed causal relationships between SECs, weight, first lung function measurement and their common causes, including baseline and time-varying confounders of the mediator-outcome relationship that are also affected by exposure, and latent variables, which are assumed present but not observed.

Deprivation at birth as measured by IMD is multi-factorial and determined by a number of historical factors, and it is difficult to conceive of a single intervention (specifically one that would satisfy the consistency condition²³⁻²⁵) that could plausibly be implemented for subsets of the CF population and could shift this exposure. However, one can conceive of interventions that could be implemented to influence weight gain for children with CF, which is down-stream of the effect of deprivation at birth and also related to lung function at around age six. This motivates defining our effects of interest in terms of interventional disparity effects which are a variant of interventional effects^{26,27}. They are based on the ideas of counterfactual disparity measures by VanderWeele & Robinson²⁸ and Naimi et al²⁹ and were formally defined in the context of multiple mediators by Micali et al¹⁴. Interventional disparity effect measures define the indirect effects in terms of the effect on the expected outcome in the exposed group if we could intervene on a mediator so as to shift the mediator distribution in the exposed to that in the unexposed; direct effects are defined in terms of the association that

would remain between exposure and expected outcome if we could intervene on a mediator so as to shift the mediator distribution in the exposed to that in the unexposed. In our context, this enables us to estimate how the association between deprivation at birth and lung function at around age six would change if we could intervene on weight trajectory up to age six. In contrast to interventional effects, interventional disparity measures do not give a causal interpretation to the effect of the exposure on the outcome.

We further extend this approach by considering a hypothetical intervention on a longitudinal mediator trajectory. This is similar to the extension of interventional effects to incorporate time-varying exposures and mediators^{30,31}.

Let X be the exposure of interest, here SECs at birth. We are interested in the contrast between those born in the least and most deprived quintiles of the population; let $X = 0$ indicate that children were born in the least deprived quintile and $X = 1$ indicate that children were born in the most deprived quintile. Let Y be the outcome (first lung function measurement at age [6-9]) and \mathbf{C} the vector of baseline covariates (sex, year of birth, genotype). Let M_t be a time-varying mediator (weight) and L_t a time-varying confounder of the mediator-outcome relationship affected by the exposure (infection), where t denotes age. The complete set of mediator measurements up to age T , when the outcome is measured, is denoted $\bar{M} = (M_1, \dots, M_T)$ and similarly $\bar{L} = (L_1, \dots, L_T)$. Let \bar{M}_c^x denote a random draw from the distribution of mediator trajectories in those where $X=x$ and $\mathbf{C}=\mathbf{c}$. For justification of conditioning the draws on \mathbf{C} , see Supplementary Material. The *interventional disparity measure direct effect* (IDM-DE) is the difference in the expected lung function between children from the least and most deprived quintiles, if the weight trajectories of all children were randomly drawn from the distribution of those observed in the least deprived children within strata of baseline confounders, standardized to the distribution of baseline confounders across both groups of children. Formally:

$$IDM - DE = \sum_c \{E[Y(\bar{M}_c^0)|X = 0, \mathbf{C} = \mathbf{c}] - E[Y(\bar{M}_c^1)|X = 1, \mathbf{C} = \mathbf{c}]\}P(\mathbf{C} = \mathbf{c})$$

The *interventional disparity measure indirect effect (IDM-IE)* is the difference in expected lung function amongst the most deprived children if we were to randomly draw their weight trajectories from the distribution of these in least deprived children versus if we were to randomly draw their weight trajectories from the distribution of these in the most deprived children, again standardized to the distribution of baseline confounders across both groups of children. Formally:

$$IDM - IE = \sum_c \{E[Y(\overline{M}_c^0)|X = 1, \mathbf{C} = \mathbf{c}] - E[Y(\overline{M}_c^1)|X = 1, \mathbf{C} = \mathbf{c}]\}P(\mathbf{C} = \mathbf{c})$$

Finally, the *total adjusted association (Adj-TA)* is the difference in expected lung function between the least and most deprived children standardized to the distribution of baseline confounders across the two groups of children. Formally:

$$\begin{aligned} Adj - TA &= \sum_c \{E[Y|X = 0, \mathbf{C} = \mathbf{c}] - E[Y|X = 1, \mathbf{C} = \mathbf{c}]\}P(\mathbf{C} = \mathbf{c}) \\ &= IDM - DE + IDM - IE \end{aligned}$$

In all the estimands corresponding integrals and densities are used for continuous \mathbf{C} .

The IDM-IE captures the entire effect of deprivation on lung function via weight including any effect from deprivation to infection to weight to lung function. However, it does not include any effect of deprivation on lung function via infection if this is not also mediated by weight. For justification of this see Supplementary Material.

Assumptions required for estimation using registry data

We estimate the estimands specified above using UK CF Registry data. This requires several assumptions: 1. no interference 2. consistency and 3. no unmeasured confounding of the mediator-outcome relationship. For explanations of assumptions 1. and 2. in this context, see Supplementary Material. The third assumption of ‘no unmeasured confounding of the mediator- outcome relationship’ includes both baseline and time-dependent confounders. We include sex, genotype and year of birth as baseline mediator-outcome confounders, and *P. aeruginosa* infection as a time-

dependent confounder. However, we are not able to control for development in utero, which we believe could confound the relationship between weight and lung function (Figure 1) but which we do not have proxy measures of in the registry data. We will assess the implications of this in the discussion.

As we do not aim to infer a causal effect of the exposure on the outcome, the assumptions required for identifiability of these effects are weaker than for other causal mediation estimands. In particular, we do not have to assume no unmeasured confounding of the exposure-outcome relationship.

Assuming all 3 conditions are met, we can identify the effects of interest from the observed data (Supplementary Material).

Estimation

Full details of the estimation procedure are given in the Supplementary Material. We provide a brief overview here. We discretized time and assumed that one weight and infection measure per year were recorded as is the guideline in the UK. Our discrete time intervals ran from one birthday up to but not including the next birthday. For example, measurements recorded after a child's sixth birthday but before they turned seven, were labelled as measurements at age 6. We used Monte Carlo simulation to estimate the effects of interest^{14,27}. This required a number of parametric models to be specified: (1) for the expected outcome Y conditional on exposure, mediator at ages 0-6, baseline confounders and time-varying intermediate confounders at ages 0-6; (2) for the distribution of the intermediate confounder L_t at each age 0-6 conditional on exposure, baseline confounders, mediator value at the previous age and previous intermediate confounder value; and (3) for the distribution of the mediator M_t at ages 0-6 conditional on exposure, previous mediator value, baseline confounders and intermediate confounder value at the same age. These models were fitted to data from all children in the study. We used non-parametric bootstrap (1,000 bootstrap samples) for variance estimation. To reduce Monte-Carlo error, all the simulations were

performed on a 1,200-time expanded dataset, where the 1,200 was chosen to render negligible the Monte-Carlo error in the results at the chosen number of decimal places.

Duplicate entries and missing data

In some cases, children had multiple measurements in a year, for example if they moved between clinics. In these cases, we selected the data from the visit closest to the child's next birthday that had complete data on weight and infection status.

We used multiple multivariate imputation by chained equations³² to impute data if weight and infection status were missing for any years (10 imputed datasets). We applied the estimation algorithm to each imputed dataset and pooled point estimates across all 10 datasets. To estimate the 95% confidence intervals, we applied standard multiple imputation procedures³³ using the variance estimated across bootstrap samples as approximation of the within imputation variance³⁴. See supplementary materials for full details.

Sensitivity Analyses

Lung function is measured with considerable error, due to natural day-to-day variation and measurement error. In a sensitivity analysis we used all available lung function data up to age 12 for individuals in the study population and fitted a linear mixed effects model with random intercept and random slope adjusted for sex, genotype, year of birth and deprivation quintile. Based on this model we predicted lung function at age six for all individuals which we used as an alternative outcome.

Due to the relatively small sample size, the parametric models specified for the estimation procedure did not include interaction terms or higher order terms but we conducted a sensitivity analysis, where we included interaction terms between deprivation and all covariates in the parametric models (1)-(3) described above.

Results

Study population

853 children met the eligibility criteria (Figure 2) including 165 and 172 children in the least and most deprived quintiles respectively (Table 1).

695 individuals had at least one year from age 0-6 without a review visit, including 135 individuals in the least, and 139 individuals in the most deprived quintile (Table 1). An additional 87 annual review visits did not record weight, including 18 review visits of children in the least deprived quintile and 17 review visits of children in the most deprived quintile. Amongst the observed data, children in the most deprived quintile were on average slightly lighter than those in the least deprived quintile across almost all ages 0-6 (Table 1, Figure 3). The proportion of children ever diagnosed with *Pseudomonas* was similar across deprivation quintiles; the same was the case for mean age at first infection.

There was a social gradient in lung function at first measurement; mean lung function was 93.11 %FEV1 (SD: 13.87) and 88.56 %FEV1 (SD: 17.57) in the least and most deprived quintiles, respectively (mean difference 4.55). A Welch two sample t-test showed that this is a statistically significant difference (95% confidence interval for difference: 1.16–7.9, p-value<0.01).

Estimated total adjusted association and interventional disparity effects

The estimated total adjusted association (Adj-TA), given as the average difference in lung function between the least and most deprived quintile of children, was 4.51 percentage points (95% CI: 1.08-7.93) (Table 2). The estimated interventional disparity measure indirect effect (IDM-IE) via weight trajectories at ages 0-6 was 0.53 percentage points (95% CI: -0.64-1.71). This is the estimated amount by which lung function would increase on average in the most deprived children if their early weight could be improved to have the same distribution as that of the least deprived children. The estimated interventional disparity measure direct effect (IDM-DE) was 3.97 percentage points (95% CI: 0.57–7.38). The difference in lung function at first measurement between the least and the

most deprived children would therefore be on average 3.97 percentage points (95% CI: 0.57–7.38) even if the weight at ages 0-6 was comparable between the least and the most deprived quintiles of children.

Sensitivity analyses

The Adj-TA, IDM-DE and IDM-IE from the sensitivity analyses are given in Table 2. The results are comparable to the main results.

Discussion

We assessed the relationship between socio-economic circumstances at birth and lung function at first measurement in children with CF in the UK, and the extent to which early inequalities in lung function can be explained by differences in weight trajectories from birth. Using modern methods for causal inference in longitudinal observational data, we estimated that inequalities in early lung function would only be marginally reduced (3.97 %FEV1 (95% CI: 0.57–7.38) versus 4.51 %FEV1 (95% CI: 1.08-7.93)) if early weight trajectories could be improved for the most disadvantaged children.

To address this question, we extended the framework of interventional disparity effects to the setting of a longitudinal mediator and time-varying intermediate confounder. This allowed us to separate the association between deprivation and lung function into the effects via distinct, longitudinally measured pathways, but without giving a causal interpretation to the effect of the exposure. It therefore requires weaker assumptions than more commonly used mediation methods, and may help gain insights in other areas of health inequalities research and social epidemiology where the consistency and no-unmeasured confounding (of the exposure and mediator and or/outcome) assumptions are often not satisfied^{29,35}.

Strength and Limitations

A strength of this study is that it uses the UK CF Registry data which collects data on 99% of the current CF population. We excluded just under 200 children because of missing data; the majority

(156) had no lung function measurement primarily because their first lung function measurement had not yet been taken at the end of our follow-up in 2016. Other reasons for missing data were incomplete recordings of postcode which meant that deprivation data could not be linked, and no annual review data before age 7. As the main source of missingness was missing of the outcome by 2016, we are confident that the exclusion of individuals with missing data did not lead to bias and that our study population is representative of the UK CF population of children born 2000-2010.

The data used in this study are recorded systematically at clinical review visits for the purposes of reporting and research and are therefore more reliable than many observational datasets.

Measurement error in lung function is, however, nevertheless likely to be non-negligible. This error is non-differential and would therefore not be expected to lead to biased estimates³⁶. We also conducted a sensitivity analysis where we used longitudinally collected lung function data to get an estimate of the error-free measure at age six. As expected, the results agreed with the results from our main analysis.

The major limitation of this study is that our results rely on untestable assumptions, namely, “no interference”, “consistency” and “no unmeasured confounding of the mediator-outcome relationship”. In our setting the “no interference” assumption states that the lung function of one child is not affected by the weight trajectory of another. This is plausible especially as children with CF are encouraged to not meet in person to prevent cross-infections. The second assumption, consistency, makes assumptions about the hypothetical intervention that would shift the weight distribution in the most deprived children. We can imagine here a potentially complex intervention that could combine financial support with aspects of higher levels of training, engagement and support from dieticians for children growing up in disadvantaged circumstances and their families. The consistency assumption states that this intervention would not change the outcome for those children whose weight trajectory was set to the trajectory that was actually observed. However, it is conceivable that stronger engagement and more support for more deprived families would improve

several other factors which may lead to improved lung function, such as clinic attendance and self-confidence in other aspects of CF care as well as reduced parental stress and improved mental health. Therefore, although we did not find any evidence that a hypothetical intervention on weight would reduce inequalities in lung function significantly, it may have positive effects beyond any direct effects on weight. Finally, we had to assume no unmeasured confounding of the mediator-outcome relationship. According to our assumed causal relationships (Figure 1), it is plausible that development in-utero impacts early weight and lung development and there is some evidence that supports that birthweight (but not gestational age) is associated with lung function at age six^{37,38}. Unfortunately, the CF registry has only recently started collecting data on birthweight and there was no proxy in the dataset that could be used to assess the impact of ignoring development in-utero in our analysis. However, as the effect of growth in-utero would have positive effects on early growth and lung function, then any hypothetical intervention on weight after birth would likely have even less of an impact than our estimated effect.

Our analysis further relies on the correct specification of the regression models used in the estimation procedure. Due to the small sample size, we did not include interactions and higher order terms in our models for the main analysis. However, we did conduct a sensitivity analysis in which we included interactions between deprivation and all covariates in the models. This did not substantially change our results although the uncertainty around estimates was increased.

Comparison with other studies and implications for further research and clinical practice

Weight in the first years of life has been shown to be strongly associated with later lung function in children with CF^{1,2,39-42} which has led to the development of CF clinical practice guidelines for nutrition management⁴³. However, this is the first study that we are aware of that has looked at the role of early weight in the generation of socio-economic inequalities in lung function in CF. Although optimal nutrition is undoubtedly important, we found that reducing inequalities in early weight would likely only have a marginal effect on inequalities in later lung function.

Other factors that may lie on the pathway from SECs to lung function have previously been studied, including the prescription of antibiotics³, exposure to environmental tobacco smoke^{44,45}, and age at diagnosis⁵. However, it remains unclear what drives inequalities in early lung function and further research is needed. There is anecdotal evidence that the regularity of clinic attendance may display a social pattern due to barriers that more deprived families face regarding transportation and the ability to take time off work. Also, engagement with the clinical care team and health literacy of parents from more disadvantaged backgrounds may be lower than in more advantaged families and may have a role to play^{46,47}. In addition, environmental factors and living conditions need further consideration.

Conclusion

There is a strong association between socio-economic conditions at birth and early lung function for children with CF which we estimated would only be marginally reduced if early weight trajectories could be improved for the most disadvantaged children. Further research is needed to identify modifiable pathways to inequalities for children with CF such that the underlying factors can be addressed. In the meantime, action to address poverty and improve social conditions for disadvantaged children growing up with CF is required to address health inequalities.

The methods developed and applied here may also help to give insights into the generation of inequalities in other conditions.

References

1. Konstan MW, Butler SM, Wohl MEB, et al. Growth and nutritional indexes in early life predict pulmonary function in cystic fibrosis. *J Pediatr*. 2003;142(6):624-630. doi:10.1067/mpd.2003.152
2. Lai HJ, Shoff SM, Farrell PM. Recovery of birth weight z-score within two years of diagnosis is positively associated with pulmonary status at age six years in children with cystic fibrosis. *Pediatrics*. 2009;123(2):714-722. doi:10.1542/peds.2007-3089

3. Taylor-Robinson DC, Smyth RL, Diggle PJ, Whitehead M. 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):121-128. doi:10.1016/S2213-2600(13)70002-X
4. Barr HL, Britton J, Smyth AR, Fogarty AW. Association between socioeconomic status, sex, and age at death from cystic fibrosis in England and Wales (1959 to 2008): cross sectional study. *BMJ*. 2011;343:d4662. doi:10.1136/bmj.d4662
5. Schlüter DK, Southern KW, Dryden C, Diggle P, Taylor-Robinson D. Impact of newborn screening on outcomes and social inequalities in cystic fibrosis: a UK CF registry-based study. *Thorax*. 2020;75(2):123-131. doi:10.1136/thoraxjnl-2019-213179
6. Taylor-Robinson DC, Thielen K, Pressler T, et al. Low socioeconomic status is associated with worse lung function in the Danish cystic fibrosis population. *Eur Respir J*. 2014;44(5):1363-1366. doi:10.1183/09031936.00063714
7. Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med*. 2001;163(6):1331-1337. doi:10.1164/ajrccm.163.6.9912100
8. Taylor-Robinson D, Diggle P, Smyth R, Whitehead M. Health Trajectories in People with Cystic Fibrosis in the UK: Exploring the Effect of Social Deprivation. In: Burton-Jeangros C, Cullati S, Sacker A, Blane D, eds. *A Life Course Perspective on Health Trajectories and Transitions*. Springer; 2015. Accessed May 25, 2021. <http://www.ncbi.nlm.nih.gov/books/NBK385365/>
9. Nutrition Working Group. *Nutritional Management of Cystic Fibrosis*.; 2016:60. <https://www.cysticfibrosis.org.uk/sites/default/files/2020-12/Nutritional%20Management%20of%20cystic%20fibrosis%20Sep%2016.pdf>
10. Savage E, Callery P. Weight and energy: parents' and children's perspectives on managing cystic fibrosis diet. *Arch Dis Child*. 2005;90(3):249-252. doi:10.1136/adc.2004.056804
11. Anthony H, Paxton S, Bines J, Phelan P. Psychosocial predictors of adherence to nutritional recommendations and growth outcomes in children with cystic fibrosis. *J Psychosom Res*. 1999;47(6):623-634. doi:10.1016/s0022-3999(99)00065-3
12. McDonald CM, Christensen NK, Lingard C, Peet KA, Walker S. Nutrition Knowledge and Confidence Levels of Parents of Children With Cystic Fibrosis. *ICAN Infant Child Adolesc Nutr*. 2009;1(6):325-331. doi:10.1177/1941406409355192
13. Brown PS, Durham D, Tivis RD, et al. Evaluation of Food Insecurity in Adults and Children With Cystic Fibrosis: Community Case Study. *Front Public Health*. 2018;6. doi:10.3389/fpubh.2018.00348
14. Micali N, Daniel RM, Ploubidis GB, De Stavola BL. Maternal Prepregnancy Weight Status and Adolescent Eating Disorder Behaviors: A Longitudinal Study of Risk Pathways. *Epidemiol Camb Mass*. 2018;29(4):579-589. doi:10.1097/EDE.0000000000000850
15. Taylor-Robinson D, Archangelidi O, Carr SB, et al. Data Resource Profile: The UK Cystic Fibrosis Registry. *Int J Epidemiol*. 2018;47(1):9-10e. doi:10.1093/ije/dyx196

16. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40(6):1324-1343. doi:10.1183/09031936.00080312
17. English indices of deprivation. GOV.UK. Accessed May 27, 2021. <https://www.gov.uk/government/collections/english-indices-of-deprivation>
18. Welsh Index of Multiple Deprivation. Accessed May 27, 2021. <https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation>
19. SpatialData.gov.scot SG. Scottish Index of Multiple Deprivation (SIMD) 2016. Published January 15, 2021. Accessed May 27, 2021. <https://data.gov.uk/dataset/a448dd2a-9197-4ea0-8357-c2c9b3c29591/scottish-index-of-multiple-deprivation-simd-2016>
20. Northern Ireland Multiple Deprivation Measures 2017. GOV.UK. Accessed May 27, 2021. <https://www.gov.uk/government/statistics/northern-ireland-multiple-deprivation-measures-2017>
21. Wright CM, Williams AF, Elliman D, et al. Using the new UK-WHO growth charts. *BMJ*. 2010;340:c1140. doi:10.1136/bmj.c1140
22. UK Cystic Fibrosis Trust. *Annual Data Report 2010.*; 2011. Accessed July 6, 2021. <https://www.cysticfibrosis.org.uk/sites/default/files/2020-12/2010%20Registry%20Annual%20Data%20Report.pdf>
23. Hernán MA, Robins JM. *Causal Inference: What If?* Boca Raton: Chapman & Hall/CRC; 2020.
24. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674-680. doi:10.1016/j.annepidem.2016.08.016
25. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes* 2005. 2008;32 Suppl 3:S8-14. doi:10.1038/ijo.2008.82
26. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiol Camb Mass*. 2014;25(2):300-306. doi:10.1097/EDE.0000000000000034
27. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiol Camb Mass*. 2017;28(2):258-265. doi:10.1097/EDE.0000000000000596
28. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiol Camb Mass*. 2014;25(4):473-484. doi:10.1097/EDE.0000000000000105
29. Naimi AI, Schnitzer ME, Moodie EEM, Bodnar LM. Mediation Analysis for Health Disparities Research. *Am J Epidemiol*. 2016;184(4):315-324. doi:10.1093/aje/kwv329
30. VanderWeele T, Tchetgen ET. Mediation Analysis with Time-Varying Exposures and Mediators. *Harv Univ Biostat Work Pap Ser*. Published online March 4, 2014. <https://biostats.bepress.com/harvardbiostat/paper168>

31. Lin SH, Young J, Logan R, Tchetgen Tchetgen EJ, VanderWeele TJ. Parametric Mediation g-Formula Approach to Mediation Analysis with Time-varying Exposures, Mediators, and Confounders. *Epidemiol Camb Mass*. 2017;28(2):266-274. doi:10.1097/EDE.0000000000000609
32. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(1):1-67. doi:10.18637/jss.v045.i03
33. Rubin DB. Multiple Imputation After 18+ Years. *J Am Stat Assoc*. 1996;91(434):473-489. doi:10.2307/2291635
34. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med*. 2018;37(14):2252-2266. doi:10.1002/sim.7654
35. Rehkopf DH, Glymour MM, Osypuk TL. The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose is Not a Rose. *Curr Epidemiol Rep*. 2016;3(1):63-71. doi:10.1007/s40471-016-0069-5
36. VanderWeele T. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press; 2015.
37. Darrah R, Nelson R, Damato EG, Decker M, Matthews A, Hodges CA. 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
38. Atteih SE, Raraigh KS, Blackman SM, Cutting GR, Collaco JM. Predictive effects of low birth weight and small for gestational age status on respiratory and nutritional outcomes in cystic fibrosis. *J Cyst Fibros Off J Eur Cyst Fibros Soc*. 2020;19(6):888-895. doi:10.1016/j.jcf.2020.02.003
39. Peterson ML, Jacobs DR, Milla CE. Longitudinal changes in growth parameters are correlated with changes in pulmonary function in children with cystic fibrosis. *Pediatrics*. 2003;112(3 Pt 1):588-592. doi:10.1542/peds.112.3.588
40. Pedreira C c., Robert R g. d., Dalton V, et al. Association of body composition and lung function in children with cystic fibrosis. *Pediatr Pulmonol*. 2005;39(3):276-280. doi:10.1002/ppul.20162
41. Konstan MW, Morgan WJ, Butler SM, et al. Risk Factors For Rate of Decline in Forced Expiratory Volume in One Second in Children and Adolescents with Cystic Fibrosis. *J Pediatr*. 2007;151(2):134-139.e1. doi:10.1016/j.jpeds.2007.03.006
42. 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):1081-1088.e1. doi:10.1016/j.jpeds.2015.07.044
43. Stallings VA, Stark LJ, Robinson KA, Feranchak AP, Quinton H. Evidence-Based Practice Recommendations for Nutrition-Related Management of Children and Adults with Cystic Fibrosis and Pancreatic Insufficiency: Results of a Systematic Review. *J Am Diet Assoc*. 2008;108(5):832-839. doi:10.1016/j.jada.2008.02.020
44. Ong T, Schechter M, Yang J, et al. Socioeconomic Status, Smoke Exposure, and Health Outcomes in Young Children With Cystic Fibrosis. *Pediatrics*. 2017;139(2). doi:10.1542/peds.2016-2730

45. Oates GR, Baker E, Rowe SM, et al. Tobacco smoke exposure and socioeconomic factors are independent predictors of pulmonary decline in pediatric cystic fibrosis. *J Cyst Fibros Off J Eur Cyst Fibros Soc.* 2020;19(5):783-790. doi:10.1016/j.jcf.2020.02.004
46. Jackson AD, Kirwan L, Gibney S, Jeleniewska P, Fletcher G, Doyle G. Associations between health literacy and patient outcomes in adolescents and young adults with cystic fibrosis. *Eur J Public Health.* 2020;30(1):112-118. doi:10.1093/eurpub/ckz148
47. Mercer SW, Higgins M, Bikker AM, et al. General Practitioners' Empathy and Health Outcomes: A Prospective Observational Study of Consultations in Areas of High and Low Deprivation. *Ann Fam Med.* 2016;14(2):117-124. doi:10.1370/afm.1910

Tables

Table 1: Demographics and clinical characteristics of the study population stratified by deprivation quintile.

	1=least deprived	2	3	4	5=most deprived	Overall
n	165	164	180	172	172	853
Male (%)	95 (57.6)	76 (46.3)	89 (49.4)	79 (45.9)	94 (54.7)	433 (50.8)
F508 class (%)						
Heterozygous	72 (43.6)	74 (45.1)	72 (40.0)	67 (39.0)	67 (39.0)	352 (41.3)
Homozygous	77 (46.7)	74 (45.1)	91 (50.6)	90 (52.3)	86 (50.0)	418 (49.0)
other	16 (9.7)	16 (9.8)	17 (9.4)	15 (8.7)	19 (11.0)	83 (9.7)
Cohort 2007-2010 (%)	107 (64.8)	110 (67.1)	120 (66.7)	115 (66.9)	115 (66.9)	567 (66.5)
First measured lung function mean (SD*)	93.11 (13.87)	91.39 (14.01)	90.91 (14.97)	89.04 (16.68)	88.56 (17.57)	90.58 (15.56)
Age at first lung function measure (%)						
6	138 (83.6)	142 (86.6)	152 (84.4)	136 (79.1)	143 (83.1)	711 (83.4)
7	24 (14.5)	18 (11.0)	23 (12.8)	32 (18.6)	25 (14.5)	122 (14.3)
8	3 (1.8)	4 (2.4)	5 (2.8)	4 (2.3)	4 (2.3)	20 (2.3)
Ever diagnosed with pseudomonas (%)	98 (59.4)	105 (64.0)	108 (60.0)	101 (58.7)	95 (55.2)	507 (59.4)
Age at first pseudomonas diagnosis mean (SD)	2.77 (1.57)	3.23 (1.72)	3.00 (1.80)	2.97 (1.67)	2.94 (1.66)	2.99 (1.69)
Number of years with missing weight data (%)						
0	25 (15.2)	25 (15.2)	30 (16.7)	31 (18.0)	28 (16.3)	139 (16.3)
1	65 (39.4)	62 (37.8)	60 (33.3)	56 (32.6)	61 (35.5)	304 (35.6)
2	44 (26.7)	51 (31.1)	38 (21.1)	41 (23.8)	43 (25.0)	217 (25.4)
3	20 (12.1)	12 (7.3)	27 (15.0)	24 (14.0)	23 (13.4)	106 (12.4)
4-6	11 (6.7)	14 (8.5)	25 (13.9)	20 (11.6)	17 (9.9)	87 (10.2)
Weight z-scores mean (SD)						
at age [0 - 1)	-0.35 (1.17)	-0.54 (1.52)	-0.50 (1.47)	-0.45 (1.39)	-0.66 (1.37)	-0.49 (1.38)
at age [1 - 2)	0.32 (0.88)	0.40 (1.05)	0.40 (1.01)	0.27 (0.89)	0.08 (1.14)	0.30 (1.01)
at age [2 - 3)	0.10 (0.92)	0.26 (0.95)	0.37 (0.92)	0.17 (0.88)	0.14 (1.11)	0.21 (0.96)
at age [3 - 4)	0.11 (0.92)	0.11 (0.97)	0.18 (0.97)	0.09 (0.94)	-0.04 (1.07)	0.09 (0.97)
at age [4 - 5)	-0.17 (0.97)	-0.00 (1.08)	0.03 (1.02)	-0.07 (0.93)	-0.17 (1.12)	-0.08 (1.03)
at age [5 - 6)	-0.07 (0.95)	-0.08 (1.04)	-0.03 (0.98)	-0.15 (1.03)	-0.14 (1.11)	-0.09 (1.02)
at age [6 - 7)	-0.08 (0.96)	0.01 (1.12)	-0.00 (0.98)	-0.03 (1.02)	-0.18 (1.18)	-0.06 (1.05)

* Standard deviation.

Table 2: Estimated total adjusted association and interventional disparity measure direct and indirect effects. In sensitivity analysis 1 we used predicted lung function at age 6 rather than observed lung function as outcome; in sensitivity analysis 2 we included interaction terms between deprivation and all covariates in the models for weight, infection and lung function.

	Main analysis	Sensitivity analysis 1	Sensitivity analysis 2
	Point estimate (95% confidence interval)		
Total adjusted association	4.51 (1.08 – 7.93)	4.32 (1.67 – 6.98)	4.40 (0.95 – 7.85)
Interventional disparity measure indirect effect (IDM-IE)	0.53 (-0.64 – 1.71)	0.47 (-0.46 – 1.41)	0.23 (-1.60 – 2.05)
Interventional disparity measure direct effect (IDM-DE)	3.97 (0.57 – 7.38)	3.85 (1.24 – 6.45)	4.17 (0.43 – 7.92)

Figures

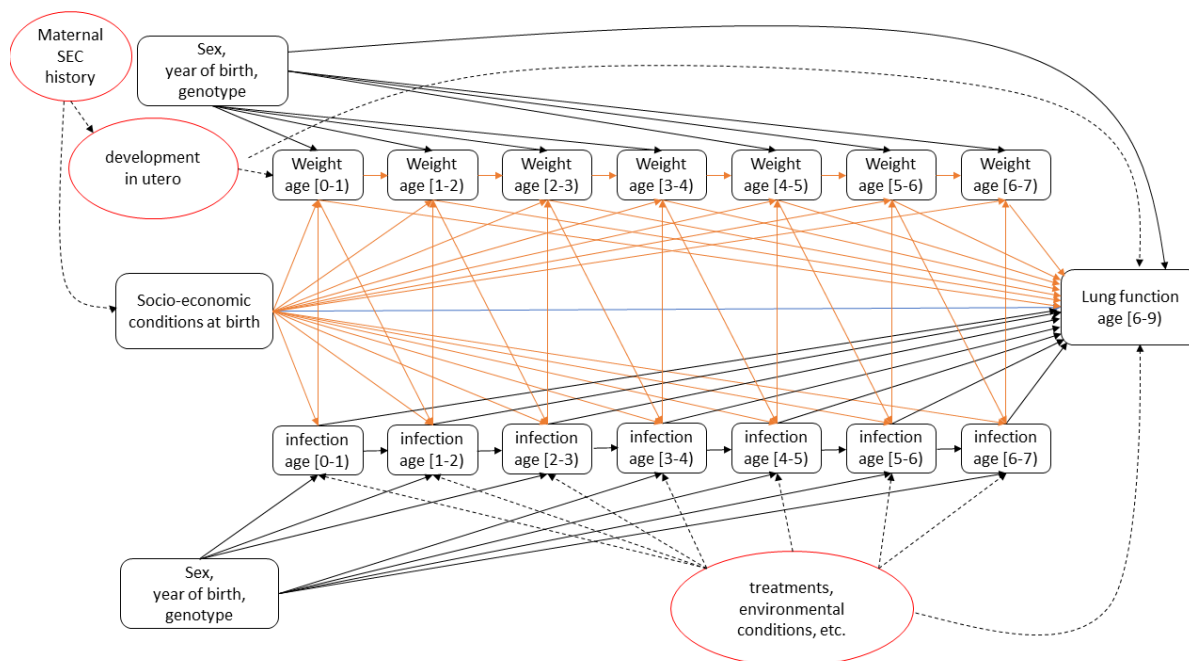


Figure 1: Assumed causal relationships between socio-economic conditions at birth, early weight, first lung function measurement and their common causes including baseline and time-varying confounders of the mediator-outcome relationship that are also affected by exposure, and latent variables, which are assumed present but not observed (red ellipses).

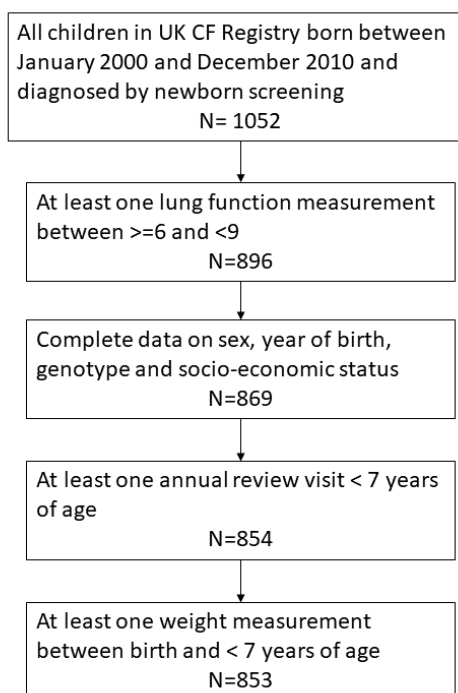


Figure 2: Derivation of the study population

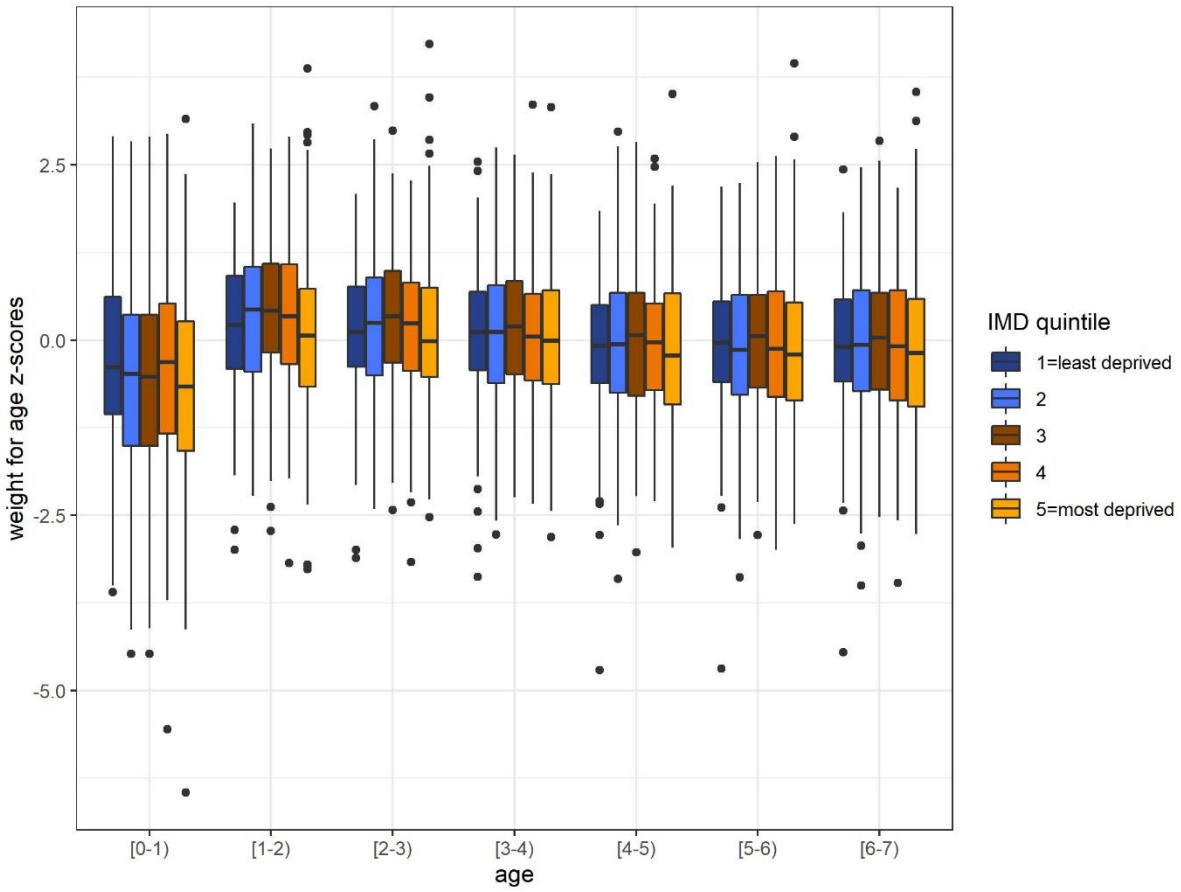


Figure 3: Weight for age z-scores by age as given in Table 1.