Economics and Human Biology

The effects in-utero exposure to influenza on mental health and mortality risk throughout the life-course.

--Manuscript Draft--

Manuscript Number:	EHB-D-20-00160R1
Article Type:	Full Length Article
Keywords:	Influenza; Prenatal Exposure; Mental Health, Mortality
Corresponding Author:	Alex Turner, PhD University of Manchester Manchester, UNITED KINGDOM
First Author:	Alex Turner, PhD
Order of Authors:	Alex Turner, PhD
	Eleonora Fichera, PhD
	Matt Sutton, PhD
Abstract:	Studies examining the later-life health consequences of in-utero exposure to influenza have typically estimated effects on physical health conditions, with little evidence of effects on mental health outcomes or mortality. Previous studies have also relied primarily on reduced-form estimates of the effects of exposure to influenza pandemics, meaning they are unlikely to recover effects of influenza exposure at an individual-level. This paper uses inverse probability of treatment weighting and “doubly-robust” methods alongside rare mother-reported data on in-utero influenza exposure to estimate the individual-level effect of in-utero influenza exposure on mental health and mortality risk throughout childhood and adulthood. We find that in-utero exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms, and increases in depressive symptoms in mid-life for males. There is also evidence that in-utero influenza exposure is associated with substantial increases in mortality, although these effects are primarily driven by a 75% increase in the probability of being stillborn, with limited evidence of additional survival disadvantages at later ages. The potential for mortality selection implies that estimated effects on mental health outcomes are likely to represent a lower bound.
Suggested Reviewers:	Mara Violato, PhD Associate Professor, University of Oxford mara.violato@dph.ox.ac.uk Expert in applied health econometrics, including research on the early-life determinants of (child) mental health.
Response to Reviewers:	

Dear Prof. Chatterji,

We would like to thank you for inviting us to resubmit our manuscript titled "The effects in-utero exposure to influenza on mental health and mortality risk throughout the life-course" for publication in Economics & Human Biology.

We thank the reviewers for their kind comments on the interest and importance of our study and also thank them for their very useful and constructive comments. We have taken account of the comments and amended the manuscript accordingly. We believe the suggested changes have helped strengthen our paper.

Below we have detailed how we have responded to each of the reviewer comments.

With kind regards,
The authors

Reviewer #1 comments

1. This paper uses longitudinal data from the National Child Development Study (NCDS) to measure the effects of influenza exposure in utero on mental health and mortality during child- and adulthood. The NCDS has three useful features: First, it includes self-reported information on influenza infections during pregnancy (reported by the mother). Second, with a focus on individuals born in the week of March 2nd 1958, the sample population has been exposed to the 1957 Avian Flu pandemic with presumably higher rates of maternal infection than during non-pandemic years. Third, useful for this study, the NCDS provides detailed measures of mental health. The authors find strong increases in mortality, mainly driven by increases in still birth. Moreover, there are positive impacts on several measures of mental health. The paper is very concise and well-motivated, it is a great read. I think it passes the bar as is but I believe there is a way to make the paper simpler, more transparent, and potentially much stronger.

Response:

We thank the reviewer for their positive comments on our manuscript

2. The main observation is that there is actually no systematic selection into influenza infections. The descriptive statistics table shows very small and I assume in many cases not statistically significant differences between mothers with and without pregnancy infections. Also, for many indicators, selection is actually reversed with infected mothers having better slightly characteristics (e.g., minority status, height, co-morbidities, previous stillbirth). This strongly indicates, that we can probably learn a lot from simple OLS regressions alone and we don't need to convince the reader of the more complicated methods to make the analysis relevant. Since OLS is the most efficient, it might likely result in somewhat more precise estimates. My other suggestion how to make effects stronger is to pool mental health outcomes across different ages. Finally, there is the question how exactly to include the mortality analysis in the narrative. The impact of maternal influenza infections on miscarriages and still births is somewhat established and currently the main motivation of the paper is the focus on mental health. I would keep the focus on mental health but then I would arrange the abstract accordingly. Currently, the abstract reads as if mortality was the main finding and mental health more of a secondary outcome.

Response:

We thank the reviewer for these constructive suggestions on how we might re-structure our analysis. We have re-arranged the abstract so that the effects on mental health are the main focus. We have also adapted our description of the mortality results, which makes clearer our contribution of being able to examine impacts on mortality throughout childhood and adulthood as well as at birth. The other changes we have made in response to these comments are detailed below.

3. This is how I would restructure the analysis:

   Include the summary statistics table (A2) in the main paper and add two columns: the difference and the p-value of the difference. Maybe mark in italics or bold those differences that go in the reverse direction. Point out in the discussion that there is no systematic pattern of socio-economic selection into maternal influenza.

Response:

Table 1 includes data on the standardised differences in covariates between the exposed and unexposed groups. As we believe this provides very similar but more interpretable information to the summary statistics, then we have chosen not to include the summary statistics in the main text to avoid information overload. For assessing statistically significant differences in covariates, we prefer to interpret the probit regression estimates and the balancing regressions, as they adjust for differences in other covariates when examining statistically significant differences in covariates across groups. However, following the

suggestion of the reviewer we have marked standardised differences that go in the reverse direction in table 1 in bold. We have also added further information in the appendix showing the standardised differences in covariates and the probit regression estimates for all ages, not just the largest estimation sample (age 7), as this shows that for a small subset of covariates (primarily number of siblings) the differences in covariates are slightly larger in estimation samples at older ages. This is likely driven by attrition.

We have also made clear in the discussion that there is little systematic pattern of socio-economic selection into maternal influenza.

2. Start the analysis with a simple OLS using raw differences in outcomes, then add a row controlling for maternal characteristics, then the row(s) with the propensity score results. Keeping them in the main table still allows you to feature this nice technique prominently, but in this order you emphasize that your data does not even necessarily require this technique.

Response:

Following the advice of the reviewer, we have now moved the OLS regressions showing raw (unweighted) differences in outcomes between the exposed and unexposed groups from Appendix 10 to the main text. Consistent with the larger differences in covariates at older ages, there is a greater difference between the adjusted and unadjusted effects on mental health at these ages. However, unadjusted and adjusted effects on mental health in childhood are almost identical, and adjustment for covariates increases the effect on mortality. In addition, our further robustness checks show that, conditional on other covariates, removing any single or group of covariates has very little impact of effects at all ages. Therefore, we believe that our results are unlikely to be explained by omitted variable bias.

We have chosen not to present results using OLS to adjust for differences in covariates. A major reason why we chose IPTW was due to its benefits over regression adjustment (Stuart, 2010). This was primarily due to it involving estimation of a range of diagnostics to show balance on covariates following adjustment and that there is sufficient overlap in the distribution of covariates between treatment and control groups. Regression adjustment tends to perform poorly when there is insufficient overlap as it involves extrapolation. It also allowed us to finalise the study design prior to analysing the effect of flu exposure on the outcome. To clarify why chose this approach, we have now added a description of these benefits to the methods section.

However, to exploit the benefits of regression adjustment, our sensitivity analysis (Section 5.1) includes the implementation of IPTW combined with regression adjustment, and we find that the main results are robust to the addition of regression adjustment.

Stuart, E., 2010. Matching methods for causal inference: A review and a look forward. Stat. Sci. 25, 1–21. doi:10.1214/09-STS313.Matching

3. Show mental health outcomes at younger and at older ages but also show results for pooled outcomes across ages, clustering at the individual level.

Response:

We thank the reviewer for this very interesting suggestion. If we are interpreting this comment correctly, the reviewer is suggesting that we construct a longitudinal dataset with age representing the time element, and use mental health across all ages as a single outcome. We agree with the reviewer that this would provide an opportunity to increase the power of the analysis. However, the mental health outcomes differ across the childhood periods (the Rutter index) and the adult periods (the Malaise index). The Malaise index at age 50 differs from the index at other adult ages, with the former being the 9-item version and the latter being the 24-item version. These measures also differ considerably both in what aspects of mental health they capture (the Rutter index being much broader in scope) and in their distributional characteristics (the Rutter index being approximately normally distributed and the Malaise index being

count distributed – see Appendix 2). For these reasons we don't believe it would be appropriate to combine mental health across ages into a single outcome in this sample.

However, we have added to the Discussion that there is a potential that our analysis may not be powered to detect effects on mental health at older ages, and that re-examination of this question using longitudinal data on consistent measures of mental health across all ages would be an interesting extension of this study.

4. Then add the balancing regressions and other checks to show one more time that selection indeed does not matter.

Response:

Thank you. We have maintained the implementation of these checks, and placed them after the previous analysis as you have suggested.

5. For mortality, instead of measuring accumulated mortality it would be more informative to measure look at mortality at different age ranges individually. My guess is that the mortality impact across all ages shown in the current version of the paper is just driven by still births.

Response:

We thank the reviewer for identifying this possibility, and after conducting additional analysis we find that our results are consistent with mortality effects at later ages being primarily driven by effects at birth. After conditioning on survival beyond birth, effects at later ages are not statistically significant, although remain positive. These results are included as Figure A.5 in Appendix A7 and are now highlighted in the main text in section 5.2.1:

"Further analysis suggests that most of the mortality effect occurs before birth, with effects at later ages being positive but losing statistical significance after conditioning on survival beyond birth." We have also highlighted this result in the abstract.

However, we have chosen to still provide impacts on accumulated mortality we feel this is most appropriate way to highlight how the risk of mortality selection differs across the life-course.

Reviewer #2 comments

The manuscript has the potential to become a contribution to a large but ever so relevant topic; the long-term consequences of influenza illness.

Response:

We thank the reviewer for their kind words on our manuscript.

1. I find it odd that the authors emphasize the result from what essentially is a sensitivity analysis (section 5.2 - Sources of bias) as their primary finding. Yet nowhere in the outcomes section is this variable mentioned, and I was unable to find it in the descriptive tables. In other words, I do not know the prevalence of this in the data. Overall, the description of the data is frustratingly insufficient. What proportion of the individuals from the original sample die before age 7/11 and so forth and how does this compare to life table estimates?

Response:

We thank the reviewer for their comment. In response, we have reorganised the text in the abstract, introduction, and discussion to ensure that the effects on mental health is seen as the main finding.

Although effects on mortality are used as a sensitivity analysis, given that previous studies have not been able to examine impacts of prenatal influenza exposure on mortality risk from birth to adulthood, we still believe this is an important finding and so we have chosen to still highlight these results in the summary of findings.

We agree that more descriptive information on the mortality measures should have been provided. In section 5.2 we have now added a figure which shows how the % of NDCS cohort members who are dead increases as the cohort ages for both the treatment and control group. Although life tables for 1958 were not accessible, we find that the proportion of cohort members who die in childhood and are comparable to the infant mortality estimates from that period. Details of this comparison have been added to section 5.2.

2. The authors appear to believe that the 1957 pandemic that they study the consequences of was more similar to the 1918 pandemic than it was. In the introduction section, they state that "although influenza mortality during seasonal influenza seasons is typically concentrated in the very young, the very old, or those with pre-existing conditions, evidence from influenza pandemics, particularly the 1918 Spanish Flu, indicate that a large proportion of excess pandemic deaths were concentrated amongst young and otherwise healthy individuals", claiming that this has the potential to aid identification of the effects of IU influenza exposure. The 1957 pandemic, however, did not have an unusual age pattern, instead being the typical U-shape, without elevated mortality among young adults.

Response:

Thank you for highlighting this mistake, and we recognise that characteristics of other pandemics will not necessarily translate into those of the 1957 pandemic. The aim of the text was to demonstrate that cases were not necessarily restricted to those who were unhealthy (e.g. the old), and that cases being prevalent in generally healthy populations increases the likelihood that maternal influenza was independent of maternal characteristics in our sample. Therefore, we have removed the sentence highlighted by the reviewer, and replaced it with a description of studies, identified from Vynnycky and Edmunds (2008), that show a large proportion of 1957 pandemic cases(identified from a large general practice in South-east London (Woodall et al., 1958), a sample of individual in Sheffield (Clarke et al., 1958), and a single GP practice in Wales (Ministry of Health, 1960)) were concentrated in young adult populations.

Clarke, S.K.R., Heath, R.B., Sutton, R.N.P., Stuart-Harris, C.H., 1958. SEROLOGICAL STUDIES WITH ASIAN STRAIN OF INFLUENZA A. Lancet 271, 814–818. doi:10.1016/S0140-6736(58)91739-2

Ministry of Health, 1960. Reports on Public Health and Medical Subjects. The influenza pandemic in England and Wales 1957–58. London.

Vynnycky, E., Edmunds, W.J., 2008. Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol. Infect. 136, 166–179. doi:10.1017/S0950268807008369

Woodall, J., Rowson, K., McDonald, J., 1958. Age and Asian influenza, 1957. Br. Med. J. 2, 1316–1318.

3. Does it make sense to make such a fuss about the methods used, primarily the treatment weighing, if the weighing turns out to essentially be unnecessary?

Response:

Although there are only small differences between exposed and non-exposed groups on the majority of covariates, results in section 4.1 suggest that are differences on some covariates. We believe that we should use the most appropriate methods to adjust for these residual differences, and to examine how robust results are to alternative ways of adjusting for these differences. We feel it is important to do this irrespective of whether these adjustments ultimately impact the overall conclusions of the study. We have now added further reasoning for our chosen approaches in the methods section.

4. The authors are inconsistent about sample selection choices and without proper motivation. When it comes to the Rutter items, less than 0.3 percent of the observations are missing. In order to limit sample loss, the authors set values to the mean values of the population (the appropriateness of this is debatable). For other measures, such as adult mental health, the authors drop missing observations. I'm left wondering why imputation is required in one case, but not in the other?

Response:

We apologies for any unclearness in the description of how missing responses in the mental health measures were handled. For the Rutter index, the 0.3% of observations alluded to by the reviewer are those with missing responses for 4 or more (out of 14) behaviours. We do not impute this data and treat the Rutter index as missing for these observations. We only impute data for observations (10% at age 11) where responses to 3 or fewer of the 14 behaviours are missing, as not doing so would lead to large amounts of sample loss (and therefore loss of power to detect meaningful effects) when information on a large amount of behaviours is available.

For these observations, missing responses for behaviours are not imputed by mean values of the population, but rather with mean values of the non-missing behaviours for the same individual. We believe this is a better approach than using population averages. We have now clarified the description in the manuscript.

For adult mental health measures, this type of item-level missingness was less prevalent. Where missing data is present for Malaise symptoms for a given individual, they are typically missing for all symptoms. Responses are available for some but not all Malaise symptoms for only 1.5% of respondents at age 23, 1.3% at age 33, 0.1% at age 42, and 0.2% of respondents at age 50. Consequently, imputation of the type implemented for the Rutter index would not result in a large increase in estimation samples. We have now noted this in the manuscript.

5. The authors state that 17,415 individuals completed the NCDS, yet later that 18,558 individuals completed an interview for the NCDS - how does this add up?

Response:

The NCDS added non-British-born children (who were born in the same week as the original NCDS cohort) during the interviews at childhood ages. This is now made clear in both the description of the NCDS in Section 2.1, and Section 3.2 where we describe the derivation of the estimation samples.

6. The authors state that the Rutter scale is a precursor to the SDQ, "one of the gold-standard instruments (...)". From this, does it automatically follow that it's equally appropriate to use? I feel this equivalency reasoning is a recurring theme in the article (see previous comment about the 1918 flu and how that experience is extrapolated on the 1957 one), and I would need more convincing.

Response:

We agree that the validity of Rutter scale doesn't necessarily follow from the validity of the SDQ. We have therefore removed this text. However, the Rutter scale is widely-used to measure mental health in childhood and has been shown to have good psychometric properties. Details of these properties have now been added to the main text in Section 2.3.

Title: The effects *in-utero* exposure to influenza on mental health and mortality risk throughout the life-course.

Authors and affiliations:

Alex J. Turner[a*], Eleonora Fichera[b], Matt Sutton[a]

[a]Health Organisation, Policy and Economics (HOPE) group, Centre for Primary Care & Health Services Research, The University of Manchester, Manchester, United Kingdom, M13 9PL.

[b]Department of Economics, University of Bath, Bath, United Kingdom, BA2 7JP

[*]Corresponding author

Email addresses & phone numbers:

Alex J. Turner: alexander.turner@manchester.ac.uk; +44 (0)161 275 1139

Eleonora Fichera: ef404@bath.ac.uk

Matt Sutton: matt.sutton@manchester.ac.uk

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interest statement: Declarations of interest: none

Original publication statement: This is original unpublished work and has not being submitted for publication elsewhere.

Ethics: Ethical approval was not required for this study.

Abstract:

Studies examining the later-life health consequences of *in-utero* exposure to influenza have typically estimated effects on physical health conditions, with little evidence of effects on mental health outcomes or mortality. Previous studies have also relied primarily on reduced-form estimates of the effects of exposure to influenza pandemics, meaning they are unlikely to recover effects of influenza exposure at an individual-level. This paper uses inverse probability of treatment weighting and "doubly-robust" methods alongside rare mother-reported data on *in-utero* influenza exposure to estimate the individual-level effect of *in-utero* influenza exposure on mental health and mortality risk throughout childhood and adulthood. ~~We find that that *in-utero* influenza exposure is associated with substantial increases in mortality, driven primarily by a 75% increase in the probability of being stillborn. There is~~We find ~~also~~ ~~evidence~~ that *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms, and increases in depressive symptoms in mid-life for males. There is also evidence that *in-utero* influenza exposure is associated with substantial increases in mortality, although these effects are primarily driven by a 75% increase in the probability of being stillborn, with limited evidence of additional survival disadvantages at later ages. The potential for mortality selection implies that estimated effects on mental health outcomes are likely to represent a lower bound.

Key words: Influenza; Prenatal exposure; Mental Health, Mortality

1. Introduction

Mental health disorders account for 13 percent of the global disease burden (Collins et al., 2011) and are the second leading cause of years lived under disability worldwide (Ferrari et al., 2017). It is also estimated that mental health problems are associated with a 4% reduction in GDP across 28 European Union countries, driven by increases in direct spending on healthcare, greater spending on social security programmes, and lower rates of employment and in-work productivity (OECD, 2018). Approximately half of mental disorders in adults begin before the mid-teenage years (Kessler et al., 2007; World Health Organization, 2013). Given this, a well-documented upward trend in child and adolescent mental health problems is worrying (Collishaw, 2015; McManus et al., 2019; Sadler et al., 2018). Understanding the determinants of mental health problems and multiple stages of the life-course is therefore vital.

A large body of literature examines how health shocks occurring in the prenatal (*in-utero*) period may impact the development of disease in later life (Almond et al., 2018; Currie and Almond, 2011). However, this literature has focused almost solely on physical health problems, with limited, although growing, evidence on how *in-utero* shocks may impact mental health (Currie, 2020). Studies assessing effects on mental health have typically identified mental health problems using mental health-related hospital admissions (Brown et al., 2000) or drug prescriptions for mental disorders (Persson and Rossin-Slater, 2018), reducing their ability to detect effects on milder mental health symptoms and to capture problems faced by individuals not seeking treatment or not having access to healthcare ~~treatment~~.

Stemming from Almond (2006)'s seminal work, exposure to influenza has emerged as one of the most widely-examined markers of the *in-utero* environment. Influenza is highly prevalent, infecting 5% - 20% of the population in the seasonal influenza season each year, and has significant effects on both mortality and morbidity (Cassini et al., 2018; Iuliano et al., 2018), particularly for pregnant women (Campion et al., 2014; Jamieson et al., 2006). A large number of studies have established effects of exposure to influenza *in-utero* on the prevalence of diabetes (Almond and Mazumder, 2005; Garthwaite, 2008; Lin and Liu, 2014), cardiovascular diseases (Garthwaite, 2008; Mazumder et al., 2010), metabolic and kidney diseases (Garthwaite, 2008; Lin and Liu, 2014), respiratory problems (Lin and Liu, 2014), and the incidence of strokes (Almond and Mazumder, 2005).[1] However, there is currently no evidence on how influenza exposure may impact the development of mental health problems. This is surprising given several hypothesised mechanisms linking the *in-utero* environment to these outcomes (Schlotz and Phillips, 2009). This includes hypothermia and inflammation which may directly damage the foetal brain during its early stages of development (Rasmussen et al., 2008) and indirect damage to the brain through under-nutrition (Colombo et al., 2004; Gale et al., 2008; Parsons et al., 2008), driven by influenza suppressing appetite, interfering with the absorption of nutrients such as fats and proteins, and through fever-induced increases in the rates of nutrient loss and energy consumption (Tomkins et al., 1994).

Previous studies are also limited by being unable to examine the impacts of *in-utero* exposure to influenza at an individual level. Long-run effects on offspring have primarily been examined by studying the impacts of influenza pandemics, making use of temporal and/or geographical variation in pandemic intensity under the assumption that influenza cases during pandemics are randomly assigned. However, a drawback of these empirical strategies is that

[1] A large number of studies also demonstrate an effect of *in-utero* influenza exposure on labour market and human capital outcomes, such as childhood cognition (Kelly, 2011), educational attainment (Almond, 2006; Garthwaite, 2008; Lin and Liu, 2014; Neelsen and Stratmann, 2012; Nelson, 2010; Percoco, 2016), literacy (Nelson, 2010), income (Almond, 2006; Nelson, 2010) and employment (Nelson, 2010), and weaker effects on family formation (Fletcher, 2018).

in-utero exposure to an influenza pandemic, or exposure to a high-virulence area, is not equivalent to *in-utero* exposure to influenza itself. Even in areas where the virulence of a pandemic is high, only a fraction of pregnant women will contract influenza. Also, given that influenza is still a prevalent disease even in non-pandemic periods, some pregnant mothers will contract influenza even in areas where the severity of the pandemic is low or even non-existent. As a result, reduced-form effects of pandemic exposure will not reflect consistent estimates of the average causal effect of *in-utero* exposure to influenza on later-life outcomes.

Schwandt (2017) uses individual-level data in a sibling fixed-effects design to study the effects of seasonal influenza on health at birth and labour market outcomes. However, influenza cases are identified using admissions to hospital for influenza-like illnesses, meaning cases which aren't severe enough to warrant hospital attendance are not recorded, and so mothers with these milder symptoms will be incorrectly identified as not contracting influenza. Estimates may therefore not test the effects of *in-utero* exposure to influenza versus no exposure, but the effect of being exposed to particularly severe cases of influenza compared to not contracting influenza or contracting a form of influenza with milder symptoms.

This study addresses these limitations using data from the 1958 National Child Development Study (NCDS); a rare dataset which collects longitudinal information on mental health symptoms of a cohort of individuals in both childhood and adulthood, and also includes mother-reported information on a range of diseases contracted during pregnancy. As data on influenza exposure is self-reported and mental health outcomes are collected via validated scales rather than recorded diagnoses, milder cases of influenza and mental health problems can be detected. Importantly, the NCDS follows all children in Great Britain born from 2nd to 9th March 1958, meaning that mothers of the NCDS cohort were pregnant during the end of the 1957 Asian Flu epidemic, which reached its peak around June when 95% of the NCDS cohort were between 16 and 25 weeks into gestation.[2] ~~Although influenza mortality during seasonal influenza seasons is typically concentrated in the very young, the very old,~~Throughout the 1957 epidemic in the UK, ~~or those with pre-existing conditions, evidence from influenza pandemics, particularly the 1918 Spanish Flu,~~ there was evidence~~indicate~~ that a large proportion of ~~excess~~ influenza cases ~~pandemic deaths~~ were concentrated amongst young ~~and otherwise healthy individuals~~adults, and therefore not necessarily concentrated among the old and less-healthy (Clarke et al., 1958; Ministry of Health, 1960; Vynnycky and Edmunds, 2008; Woodall et al., 1958). This has the potential to aid identification of the effects of *in-utero* influenza exposure estimated through a simple comparison of outcomes for exposed and non-exposed groups, as exposure is less likely to be correlated with (unobserved) markers of maternal health. We find evidence of this in our data, with mothers reporting having contracted and not contracting influenza during pregnancy being similar on markers of maternal health and socio-demographic characteristics.

However, to adjust for small remaining differences in characteristics of those exposed and not-exposed to influenza, we estimate the effects of *in-utero* exposure to influenza using inverse propensity of treatment weighting and "doubly-robust" inverse-probability-weighted regression adjustment methods to ensure mothers contracting influenza in pregnancy are identical in observed characteristics to those not contracting influenza.

In a further contribution, we use linked data on NCDS and administrative death records to examine whether *in-utero* exposure to influenza impacts mortality risk at different stages of the lifecourse. This allows an assessment of whether effects on mental health are likely to be affected by mortality selection, but also addresses the paucity of

[2] See Kelly (2011) for further details on the 1957 Asian Flu epidemic in Great Britain.

studies examining the later-life mortality effects of *in-utero* influenza. Fletcher (2019) finds that exposure to the 1918 Spanish Flu did not lead to increases in overall or cause-specific mortality, but uses only death information for individuals with data on education and labour market outcomes in adulthood, meaning impacts on mortality risk in early-life and how this risk changes as individuals age could not be examined.

Our results indicate that ~~*in-utero* influenza is associated with a significant reduction in survival rates, primarily driven by effects on the rate of stillbirths. Results also suggest that~~ *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms of mental health, and increases in depressive symptoms in mid-life for males. Findings are robust to the methods used to adjust for observed covariates, and tests indicate they are likely to be unaffected by attrition bias. ~~We also estimate bounds for treatment effects under survival bias, which suggest that estimated effects are likely conservative.~~ We also demonstrate that, consistent with small differences in observed child and family background characteristics between exposed and non-exposed groups, estimates are stable to the removal of these characteristics from the covariate set, providing evidence that omitted variable bias is unlikely to be impacting results greatly. However, we find that *in-utero* influenza is associated with a significant reduction in survival rates, primarily driven by effects on the rate of stillbirths, suggesting effects on mental health may be impacted by survival bias. Estimating bounds for treatment effects under survival bias, we find that estimated mental health effects are likely to be conservative.

The remainder of this article is organised as follows. The data is described in Section 2. Section 3 outlines the empirical strategy. The main findings are presented in Section 4. Robustness checks are presented in Section 5. Section 6 is devoted to the discussion and concluding remarks.

2. Data

2.1. National Child Development Study (NCDS)

Our primary data source is the National Child Development Study (NCDS). The NCDS is a longitudinal study which follows a cohort of approximately 17,000 individuals born in England, Scotland and Wales between the 2nd and 9th of March, 1958. The NCDS began with the Perinatal Mortality Survey (PMS), which collected data via questionnaires completed by mothers and midwifes in attendance at delivery, on factors associated with stillbirth and infant death. Population coverage was high, with the PMS accounting for 98% of all live and still births occurring in that week (Butler and Bonham, 1963). There currently exists data from nine subsequent interviews which took place at ages 7 (in 1965), 11 (1969), 16 (1974), 23 (1981), 33 (1991), 42 (1999/2000), 46 (2004/05), 50 (2008/09) and 55 (2013/14), which gathered information on a range of health, education, social and economic outcomes. For interviews during childhood ages, the NCDS also traced immigrants born in reference week and added them to the sample (Power and Elliott, 2006). Data at ~~these~~ follow-ups was drawn from a variety of sources including the cohort members themselves, their parents and partners, local authority medical officers, and schools. 17,415 individuals completed the PMS, with the number completing falling as the cohort aged.

We also use information on deaths of NCDS cohort members identified from the National Health Service Central Register (NHSCR), NCDS death cards, NCDS interviewers, and relatives and friends of cohort members (Johnson and Brown, 2015), as well as secure access data on the exact month and year of death (Institute of Education, 2015a).

2.2. *In-utero* influenza exposure

The key feature of the NCDS is that mothers were asked whether they experienced a range of health conditions during pregnancy. In the PMS, NCDS mothers were asked "Were any of the following abnormalities or illnesses, or any other condition, encountered in pregnancy?" of which influenza was one of the conditions listed. We construct a binary variable, flu, which is equal to one if the mother of an NCDS cohort member reports having contracted influenza during pregnancy, and zero otherwise.

2.3. Outcomes

We measure childhood mental health at ages 7 and 11 using indicators from the Rutter Behaviour Child Scale A. ~~(Rutter et al., 1970).~~ The scale has good psychometric properties, with estimates of inter-rater reliability of r=0.64 and re-test reliability of ~~,~~ r=0.74 (Rutter et al., 1970). ~~The Rutter scale is the pre-cursor to the Strengths and Difficulties Questionnaire, widely regarded as one of the gold standard instruments for the measurement of social and emotional well-being in children (Wolpert et al., 2009).~~ When cohort members were aged 7 and 11, parents of each cohort member were asked to state the frequency that they observed their child engaging in range of negative behaviours on a scale of "Never", "Sometimes", "Frequently". The complete scale includes 31 behaviours, but only 14 are used in the NCDS[3] (Centre for Longitudinal Studies, 2012). Responses for each behaviour are recorded as 0, 1 and 2 respectively, and are summed to create an overall index ranging from 0 (perfect mental health) to 28 (worst mental health). Approximately 6% and 10% of parents complete only a subsection of the Rutter items at the age 7 and 11 follow-ups, respectively (Table A1, Appendix A1). Where parental responses are missing for~~to~~ four or more ~~items~~ of the 14 behaviours ~~are missing~~ (<0.3%), the index is set to missing. To limit sample loss, for individuals with missing responses for ~~those with~~ less than four ~~missing responses~~behaviours, missing values for these ~~responses~~ behaviours were imputed with the average response to the non-missing ~~responses~~ behaviours for that same individual[4].

To examine effects on different aspects of mental health, we also construct separate Rutter indices for externalising and internalising symptoms. Externalising symptoms relate to an inability to regulate behaviour, with symptoms similar to those associated with attention deficit hyperactivity disorder (ADHD). Internalising symptoms relate to problems regulating emotions and mood, with symptoms usually indicative of disorders relating to depression and anxiety (Klein et al., 2009). Rutter items were assigned as externalising/internalising using categorisations derived using principal components analysis in Blanden et al., (2007) and Klein et al., (2009).

Adult mental health is measured using Rutter's Malaise Inventory (Rutter et al., 1970) at ages 23, 33, 42, and 50. The inventory asks cohort members to self-report the presence of a variety of symptoms of anxiety and depression, through questions such as "Do you often feel depressed?', "'Do you feel tired most of the time?", "Are you easily upset or irritated?", and "Do you often have bad headaches?". The NCDS reports the 24-item measure at ages 23, 33 and 42, and the shorter 9-item version at age 50. Yes/No responses to each item are coded as 1/0, and summed

[3] In the NCDS, parents are asked whether their child: "Is squirmy or fidgety", "Destroys own or others' belongings (e.g. tears or breaks)", "Fights with other children", "Worries about many things", "Prefers to do things on his/her own rather than with others", "Is irritable, often flies off the handle", "Is miserable or tearful", "Has twitches or mannerisms of the face, eyes or body", "Sucks thumb or finger during the day", "Bites nails", "Is disobedient at home", "Has difficulty in settling to anything for more than a few moments", "Is upset by new situation, by things happening for first time", "Is bullied by other children".

[4] Being exposed to influenza in-utero was also not a significant predictor of the Rutter index being partially imputed, when all other covariates were controlled for, giving us confidence that the imputation method is unlikely to affect results substantially.

to construct a count of depressive symptoms, consistent with evidence that the inventory measures a single underlying factor of distress (Hirst, 1983; Hirst and Bradshaw, 1983; McGee et al., 1986). Due to low levels of item-specific missing data for adults[5], no imputation was conducted for this measure.

The distributions of each of the mental health outcomes in childhood and adulthood are shown in Figure A1 and Figure A2 (Appendix A2).

2.4. Covariates

Child and family background characteristics are derived primarily from the PMS. Gender and ethnicity are measured using binary indicators for being male and an ethnicity minority, respectively[6]. Parental socioeconomic status is measured using the social class of the cohort members' father and mother as well as mother's education. Social class is measured in five categories based on parental occupation prior to pregnancy: Class I (professional occupations), Class II (managerial and lower professional occupations), Class III (manual or non-manual skilled occupations), Class IV (semi-skilled occupations), and Class V (unskilled occupations). For mother's social class, classes I and II are not separable and so are combined, and an additional category is included for the 57% of NCDS mothers who are not employed and therefore cannot be assigned to a social class based on their occupation[7]. Maternal education is measured using a binary variable equalling one if the mother was educated past the minimum school leaving age of 15.

Maternal and paternal age at birth is measured using categorical variables with 5-year age intervals to allow for non-linearity in its effects. For mother's age, these range between age 20 and under to age 40+, and for father's age, between age 20 and under and age 50+[8]. Mother's height is measured in inches. Where information on paternal variables are missing due to an absent father-figure, we include a separate category for missing information for all father variables (father's age and father's social class)[9].

Maternal health investment is measured using maternal smoking and weight prior to pregnancy, and engagement with maternity services. Maternal smoking is included as a categorical variable ranging from non-smoker to 20+ cigarettes per day. Mothers' weight is measured using their body mass index (BMI), and were categorised as being a normal/healthy weight, underweight, overweight or obese, using validated cut-offs (Centers for Disease Control and

[5] At ages 23, 33, 42 and 50, responses are missing for a subset of Malaise symptoms for only 1.5%, 1.3%, 0.1%, and 0.2% of observations, respectively, and so imputation of the type implemented for the Rutter index would not result in a large increase in estimation samples.

[6] The low prevalence of ethnic minorities in this historical sample prohibited the use of a more granular measure of ethnicity. This measure was derived from secure access data on the deaths of NCDS cohort members as self-reported ethnicity was not provided until the age 11 follow-up and so would be missing for cohort member dying prior to this follow-up.

[7] This is consistent with the low female labour force participation rate prevalent in the UK throughout the 1950s (Sprahue, 1988).

[8] A missing value was assigned to maternal age for a cohort member whose mother was recorded as age 8.

[9] We do this because we believe dropping observations with missing data relating to these variables could result in a selected sample. The alternative would be dropping these variables as controls. However, we feel this will likely generate omitted variable bias. For example, given that only 40% of NCDS mothers report working, mother's social class is unlikely to be representative of the socioeconomic status of the family, and so the inclusion of father's social class is required to ensure effects of socioeconomic status are adequately captured.

Prevention, 2015)[10]. An indicator for whether the mother visited a neo-natal unit in the first six weeks of pregnancy is used to willingness to engage with health services[11].

To measure maternal health, we construct a binary co-morbidity indicator which equals one if a mother reported suffering from diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy. As a further proxy of a mother's health, as well as the health of their offspring, we measure NCDS mother's experience in previous pregnancies using separate binary indicators for previous abortions, premature births, large births, stillbirth/neo-natal deaths, and birth-related complications. This is under the assumption the negative birth outcomes in pervious pregnancies is indicative of poor genetic health endowments.

Housing density and the number of older siblings are included, due to evidence that they are key determinants of influenza spread (Woodall et al., 1958). Housing density is measured using a categorical variable describing the average number of persons per room in the cohort member's place of residence. The number of older siblings is first derived from self-reported sibling data at age 16. Given this generates missing data for those who left the NCDS or die prior to this age, we replace missing values with a proxy which combines PMS data on parity (the number of previous pregnancies an NCDS mother has experienced prior to the birth of the NCDS cohort member), and the number of previous still births and neo-natal deaths. The number of live older siblings is proxied as the difference between these measures[12].

We supplement these individual-level characteristics with area-level measures of population health which may determine disease spread and proxy maternal health at an individual-level. Area-level variables were matched to NCDS cohort members via identifiers of their local authority of birth[13], which were obtained under special licence (Institute of Education, 2015b). To proxy the underlying prevalence of influenza, we construct a rate of the annual number of pneumonia notifications occurring between the months of September and November averaged across the two years prior to the pregnancy period of the NCDS cohort (1955 and 1956), per 100,000 persons[14]. These were derived from weekly data on pneumonia notifications by local authority, obtained from the Registrar General's Returns for England and Wales, and of Scotland (Registrar General for England and Wales, 1957; Registrar General for Scotland, 1957)[15].

Local authority stillbirth rates were derived from the Registrar General's Returns as a ratio of the total number of stillbirths in 1956 to the total number of births in the same year. Population density, measured as the population in thousands per square kilometre, was constructed using population figures from the 1956 Registrar General's Returns and local authority area data from UKBORDERS. Finally, further controls are derived from the 1951 Census county reports, which provided statistics based on the total population of each local authority. These include the

[10] In the PMS, NCDS mothers report weight in categories of <7 stone, 7-8 stone, 8-9 stone, 9-10 stone, 10-11 stone, 11-12 stone, 12-13 stone, 13-14 stone, 14-15 stone, and >15 stone. To construct BMI, lower and upper limits are taken for the <7 stone and >15 categories, respectively, and mid-points are taken for all other categories. These are then converted to pounds, before being divided by the square of mother's height and multiplied by 703.

[11] This was preferred to a measure of total number of visits as this is more likely increased by contacting influenza. This visit is highly correlated with the total number of visits.

[12] Although we cannot correct for the number of miscarriages here, we still believe this to be good quality proxy. This is supported by the 89% agreement rate between the proxy measure and the self-reported measure in the sample of cohort members for which self-reported data is available.

[13] In 1958, local authorities numbered 173 and contained an average of approximately 300,000 inhabitants.

[14] Official notification of influenza cases by physicians was not required by law in Britain during the study period, and Pneumonia is clinically related to influenza, with approximately 1 in 417 cases of influenza resulting in pneumonia (Hunter and Young, 1971).

[15] We thank Dr Elaine Kelly for the provision of this data.

proportion of the male working population in an unskilled occupation and the proportion of men leaving school aged 16 or older. We also include indicators for each of Great Britain's 11 regions.

3. Methods

3.2. Empirical strategy

A primary concern with estimating the impact of *in-utero* exposure to influenza by comparing outcomes of exposed and non-exposed groups is that mothers who contract influenza may differ in characteristics that predict outcomes. A particular concern here is differences in underlying maternal health. Due to evidence of substantial intergenerational transmission of health even in early life (Coneus and Spiess, 2012), maternal and offspring health are likely to be positively correlated, meaning that if mothers in poorer health are at increased risk of contracting influenza, any negative effects of *in-utero* influenza exposure on mental health outcomes could be inflated.

However, due to previous evidence of high prevalence of influenza even in prime-age healthy adults during pandemic periods (Woodall et al., 1958), it is unclear *a priori* whether a maternal health gradient in influenza exposure should be expected in this sample. Summary statistics for child and family background characteristics available in the NCDS provides evidence against this (Table A2, Appendix A3), with the 13% of mothers reporting contracting influenza during preganacy experiencing similar (and marginally lower) rates of co-morbidity and previous stillbirth/neonatal death compared to the non-exposed group. This provides evidence that confounding due to unobserved markers of maternal health is unlikely to subantially bias simple differences in outcomes across groups.

However, to adjust for residual differences in characteristics across exposed and non-exposed groups, we identify the effects of *in-utero* exposure to influenza using inverse probability of treatment weighting (IPTW), which estimates causal effects under the assumption of no unobserved confounding. IPTW estimates treatment effects by re-weighting the sample based on the probability of treatment (here, the probability of influenza exposure) such that characteristics of exposed and non-exposed cohort members are identical, on average, in observed characteristics. Higher weights are placed on non-exposed cohort members which are more similar in characteristics to exposed cohort members.

IPTW was preferred to recently-developed matching methods such as coarsened exact matching, which have the potential to generate greater balance on covariates (Iacus et al., 2011), but can result in a substantial unmatched sample when the majority of covariates are binary or categorical. IPTW also has benefits over regression adjustment, as the latter relies heavily on extrapolation when there is insufficient overlap in distribution of covariates between exposed and non-exposed group (Stuart, 2010), and tends to preform poorly in these situations (Dehejia and Wahba, 2002, 1999). IPTW offers straightforward diagnostics to assess whether adjsutments can be made without extrapolation. A further benefit of IPTW is that that, similar to randomised experiments, the study design can be finalised prior to the analysis of exposure on the outcome (Rubin, 2001).

To implement IPTW, we first estimate the propensity score, $\rho = P(flu = 1|X)$, representing the probability of being exposed to influenza *in-utero* given values of the covariates (Rosenbaum and Rubin, 1983). Here, propensity

scores are derived as predicted probabilities from a probit model including all covariates described above[16]. Weights are then derived which equal the inverse of the probability of treatment that the cohort member received, such that they equal $\frac{1}{\rho}$ for the exposed and $\frac{1}{1-\rho}$ for the non-exposed (Rosenbaum, 1987).

The average treatment effect (ATE), which represents the effects of moving the entire population from non-exposed to exposed (Imbens, 2004), is then estimated as:

$$\frac{1}{n}\sum_{i}^{N}\frac{flu_i Y_i}{\rho} - \frac{1}{n}\sum_{i}^{N}\frac{(1-flu_i)Y_i}{1-\rho}$$

where N is the number of observations.

IPTW should ensure that the "balancing property" holds, such that the distribution of covariates are equivalent between the sample of exposed and non-exposed cohort members. We examine covariate balance in multiple ways (Austin and Stuart, 2015). First, we examine standardised differences in covariates between exposed and non-exposed groups, both prior to and following re-weighting the data, as well as examining whether re-weighting results in the ratio of variance of the covariates between groups converging to one (representing equal variances across groups). Finally, we formally test balance using a test derived in Imai and Ratkovic (2014) which treats restrictions imposed by balance as overidentifying restrictions.

IPTW also requires the validity of the overlap (or positivity) assumption, which states that each individual must have a positive probability of treatment. In cases where propensity scores are low, weights become large, which can cause IPTW estimates to become erratic due to an insufficient number exposed of observations to estimate treatment effects at the tails of the propensity score distribution (Cole and Hernán, 2008). To test for sufficient overlap, we compare kernal density plots of the propensity score for both exposed and non-exposed cohort members. The overlap assumption is violated when estimated densities have too much mass around 0 or 1 (Busso et al., 2014). We use heteroskedastic-robust standard errors throughout.

We estimate effects for the full sample and also seperately for males and female offspring, given evidence from previous stuides that the effects of *in-utero* shocks may differ by gender[17].

3.2. Derivation of the estimation samples

Due to substantial attrition in the NCDS, separate estimation samples were constructed for each outcome. The process describing the construction of these samples is depicted in Figure A3 (Appendix A4). In total, 18,558 individuals completed an interview for at least one wave of the NCDS. We began by dropping individuals with no PMS record, reducing the sample to 17,421 cohort members. This primarily included the non-British-born individuals added to the NCDS at the childhood follow-ups. We then excluded the 452 cases where data on individual-level exposure to influenza was missing[18]. For the age at which each outcome was recorded, we dropped individuals who failed to return a survey at this age. We then dropped individuals with missing data on the

[16] Results are unchanged when propensity scores are derived from a logit model.

[17] See for example Arthi (2018), Fletcher (2018), Havari and Peracchi (2017), Lin and Liu (2014), Lindeboom et al., 2010; Neelsen and Stratmann, 2012; Scholte et al., 2015; van den Berg et al., 2015; Yemelyanau et al., 2012).

[18] 426 of these relate to multiple births (212 twin-pairs and 4 sets of triplets), whose influenza data was not made available in the NCDS.

outcomes. We finally dropped observations with missing data on any of the control variables. Implementing this process resulted in final estimation samples of ranging between 12,291 (Rutter index at age 7) and 8,002 (Malaise at age 50).

4. Results

4.1. Covariate balance and overlap

Consistent with the summary statistics, probit estimates used to derive propensity scores (Table A4~~3~~, Appendix A6~~5~~)[19], suggest few differences in average levels of observed covariates between those exposed and not exposed to influenza *in-utero* in each estimation sample, with average marginal effects indicating that the majority of covariates, including parental socioeconomic status and maternal health, are not associated with the likelihood of influenza exposure *in-utero*[20]. Some patterns are present for maternal age and the number of siblings, where having a large number of siblings and having a younger mother are both associated with a statistically significant increase in the probability of influenza exposure, the former being consistent with children being a primary source of influenza spread. ~~-~~Rates of exposure were also slightly higher for cohort members whose mothers were underweight prior to pregnancy, and lower for cohort members born in local authorities with high stillbirth rates. The difference in covariates is slighltly more pronounced at adult ages, likely driven by the effects of sample attrition, with the largest differences being for the number of older siblings.

Despite small differences in mean levels of covariates prior to weighting, standardised differences and variance ratios for the unweighted and weighted data suggest that IPTW was successful in acheving greater covariate balance (Table 1; Table A3, Appendix A5). Following weighting, for the majority of covariates, standardized differnces in covariates between exposed and non-exposed groups were zero (to two decimal places), and variance ratios were extremely close to one, implying the mean values and the spread of the covariate distributions were approximately identical across groups. This is confirmed in formal balance tests, where balance could not be rejected in any of the outcome-specific estimation samples (Table 2).

Distributions of propensity scores for the exposed and non-exposed groups provide evidence in favour of the validity of the overlap assumption (Figure 2). Propensity scores were marginally larger on average in the group exposed to influenza *in-utero*, but densities of cohort members from exposed and non-exposed groups are similar at most parts of the propensity score distribution; consistent with the small unweighted differences in covariates across groups in Table 1. Numbers of cohort members with proponsity scores close to zero are also small.

~~4.2. ATE estimates~~

~~IPTW estimate of ATEs suggest that *in-utero* exposure to influenza led to small reductions in mental health in childhood at both ages 7 and 11 (Table 3). Reductions were smaller at age 7, where *in-utero* exposure led to a small but statistically insignificant 0.103 increase in the overall Rutter index, equivalent to 1.63% increase relative to the mean index for the non-exposed group. At age 11, the magnitude of the effect doubles to 0.194 (3.04%) and~~

[19] ~~For ease of presentation, the majority of~~Tab ~~balance statistics are only presented for the largest outcome-specific sample (Rutter indices at age 7).~~

[20] Statistically insignificant marginal effects don't seem to be driven by high levels of multicollinearity between covariates. The mean variance inflation factor (VIF) for the propensity score estimated model using ordinary least squares is 2.84, driven primarily by high values on ordered categorical variables where high VIFs are commonplace.

becomes statistically significant at a 5% level. At both ages, reductions in mental health were primarily driven by an increase in internalising behaviour (3.68% at age 11).

In utero influenza exposure also led to small increases in the number of depressive symptoms included in the 24-item Malaise inventory at ages 23, 33, and 42, ranging from an increase of 0.05 symptoms (1.3%) at age 42 to an increase of 0.08 symptoms (3.42%) at age 33, although these were not statistically significant (Table 1). At age 50, the magnitude rises to a 0.09 symptom increase (equivalent to 6.35% given the switch to the 9-item inventory at this age), but again this effect is not statistically significant. Statistical insignificance despite moderate effect sizes are likely to be driven by reduced sample sizes at later ages.

4.3. Heterogeneity by gender

Effects differ considerably by gender, with this heterogeneity varying by the age at which mental health is measured (Table A4, Appendix A6). At age 7, effects on the overall Rutter index and internalising and externalising subscales are stronger for males, but remain statistically insignificant for both genders. Effects on the Rutter index at age 11 are stronger for females, driven by a statistically significant increase in internalising symptoms of 7.72%. However effects on externalising symptoms remain stronger for males (4.57% vs -2.05%) although effects are again not statistically significant.

Effects of *in utero* influenza exposure on the Malaise inventory are not statistically significant for both genders at ages 23, 33, and 42, with some evidence of larger effects for females. However, at age 50, effects are much larger for males, with *in utero* influenza exposure resulting in a statistcially significant 0.176 (15.45%) increase in depressive symptoms.

Table 1: Covariate standardised differences and variance ratios before and after inverse probability of treatment weighting.

	Standardized differences		Variance ratio	
	~~Raw~~Unweighted	Weighted	~~Raw~~Unweighted	Weighted
Male gender	-0.019	0.001	1.001	1.000
Ethnic minority	-0.023	-0.009	0.797	0.917
Mother's occupational social class (Base: V: Unskilled)				
Unemployed	**-0.034**	0.001	1.016	1.000
I or II: Professional, managerial or technical	**0.010**	-0.007	1.044	0.970
III: Skilled manual/non-manual	**0.010**	0.001	1.014	1.001
IV: Partly skilled	**0.031**	0.001	1.087	1.004
Father's occupational social class (Base: V: Unskilled)				
I: Professional	-0.051	0.011	0.778	1.050
II: Managerial or technical	-0.099	-0.007	0.789	0.985
III: Skilled manual/non-manual	**0.077**	-0.002	0.970	1.001
IV: Partly skilled	**0.009**	0.009	1.021	1.021
Information missing	0.000	-0.010	1.000	0.953
Mother schooled past leaving age	-0.039	0.000	0.955	1.000
Mother's age in years (Base: 20 or under)				
21-25	0.057	0.003	1.045	1.003
26-30	-0.069	0.000	0.944	1.000
31-35	0.006	-0.010	1.010	0.984
36-40	**-0.006**	0.010	0.983	1.032
Over 40	**-0.035**	-0.006	0.768	0.961
Mother's height (inches)	**0.026**	-0.011	1.055	1.062
Father's age in years (Base: 20 or under)				
21-25	0.051	0.003	1.077	1.005
26-30	-0.003	0.001	0.998	1.000
31-35	-0.022	-0.001	0.973	0.998
36-40	-0.013	0.003	0.971	1.007
41-45	0.032	0.002	1.139	1.007
45-50	-0.025	-0.006	0.817	0.951
Over 50	-0.075	-0.008	0.276	0.900
Information missing	-0.005	-0.003	0.973	0.980
Mother's smoking prior to pregnancy (Base: Non-smoker)				
1-4 cigarettes per day	0.024	0.001	1.077	1.004
5-9 cigarettes per day	**-0.012**	0.005	0.975	1.011
10-14 cigarettes per day	**-0.015**	0.001	0.968	1.002
15-19 cigarettes per day	**-0.029**	0.004	0.844	1.024
20+ cigarettes per day	0.009	-0.001	1.057	0.995
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])				
Underweight [BMI<18.5]	**-0.071**	0.000	0.679	1.002
Overweight [25<BMI<30]	0.030	0.001	1.047	1.001
Obese [BMI>30]	0.039	-0.001	1.183	0.994
Antenatal visits within 7 weeks of pregnancy	**0.011**	-0.001	1.050	0.995
Any comorbidity†	**-0.014**	-0.008	0.952	0.972
Previous abortion	**-0.003**	0.005	0.993	1.011
Previous premature birth (<5lb,9oz)	0.040	0.007	1.127	1.022
Previous large birth (>8lb,14oz))	0.063	0.003	1.184	1.008
Previous stillbirth/neonatal death	**-0.012**	-0.001	0.951	0.996
Complication in previous pregnancy	0.038	0.000	1.116	1.000
Persons per room (Base: <1)				
1-1.5	-0.010	-0.005	0.984	0.991
1.5-2	0.002	0.006	1.006	1.016
>2	0.027	0.001	1.126	1.003
Number of older siblings (Base: No siblings)				
1	-0.017	-0.005	0.988	0.996
2	0.041	0.010	1.081	1.020
3	-0.010	0.004	0.969	1.013
4	0.034	-0.006	1.182	0.969
5	0.072	-0.001	1.612	0.996
6+	0.018	0.010	1.123	1.066

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.
†Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy

Highlighted in bold are the standardised differences which suggest that, compared to those who were not exposed, individuals exposed to influenza *in-utero* have levels of maternal characteristics that are less predictive of poor outcomes.

Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold
Formatted: Font: Bold

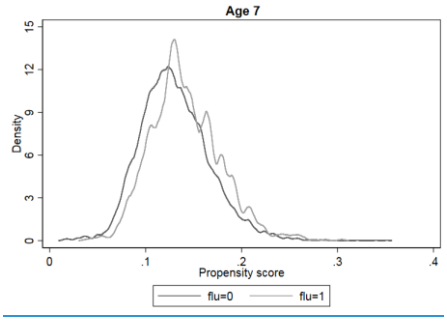
Table 1 (continued): Covariate standardised differences and variance ratios before and after inverse probability of treatment weighting.

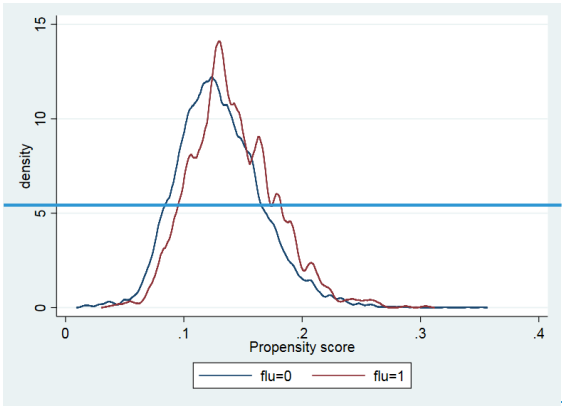
	Standardized differences		Variance ratio	
	~~Raw~~Unweighted	Weighted	~~Raw~~Unweighted	Weighted
Stillbirth rate (Still births/total births)	0.065	0.009	1.396	1.046
Proportion of male working population unskilled	-0.051	-0.002	0.948	1.014
Proportion of men leaving school aged 16 or older	-0.049	0.002	0.897	0.961
Population density (population in thousands/square km)	-0.055	-0.009	0.917	0.962
Pneumonia Notifications Per Thousand Population prior to pregnancy	-0.038	0.008	0.828	1.019
Region of birth (Base: Northern England)				
North West England	-0.066	-0.003	0.860	0.994
East & West Riding, England	0.023	-0.005	1.070	0.986
North Midlands, England	0.046	-0.001	1.156	0.996
Midlands, England	0.100	0.004	1.293	1.010
East England	0.020	0.002	1.061	1.006
South East England	-0.079	-0.012	0.867	0.979
South England	0.018	-0.008	1.073	0.970
South West England	-0.010	0.005	0.960	1.021
Wales	0.048	0.006	1.204	1.024
Scotland	-0.051	0.018	0.883	1.041

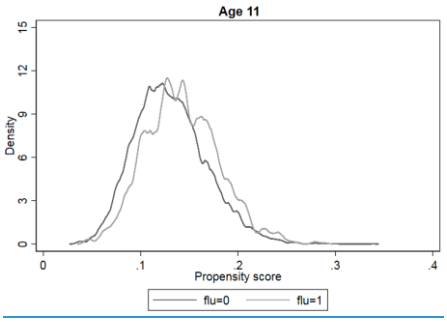
Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.

Table 2: Over-identification tests for covariate balance following inverse probability of treatment weighting for each outcome-specific estimation sample

	Chi-squared statistic	P-value
Rutter: age 7	48.181	0.951
Rutter: age 11	42.762	0.988
Malaise: age 23	51.180	0.910
Malaise: age 33	43.675	0.985
Malaise: age 42	39.818	0.996
Malaise: age 50	41.507	0.992







Formatted: Space Before: 0 pt, After: 0 pt

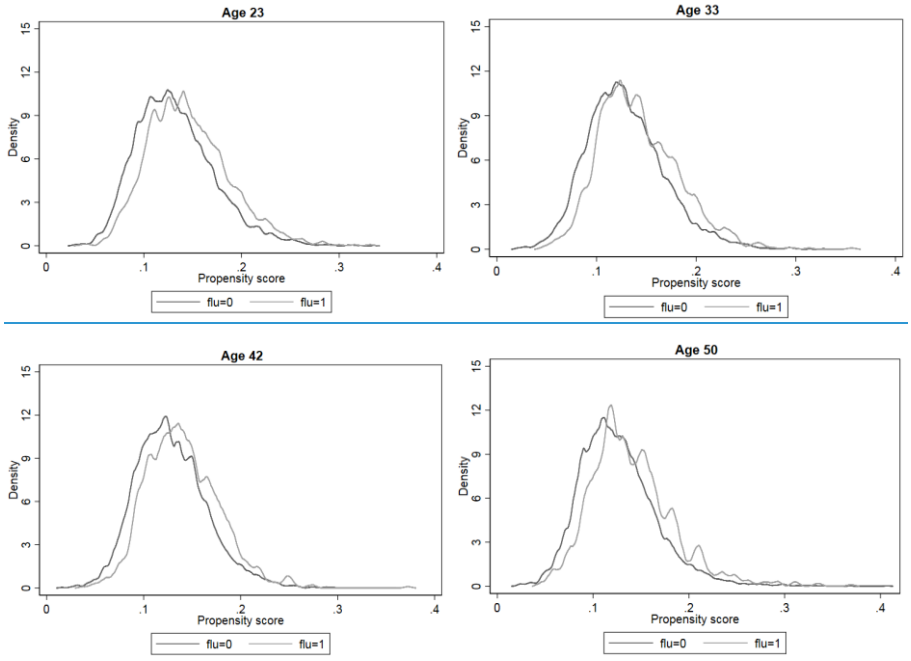


Figure 2: Kernel density plots of estimated propensity scores for both exposed and non-exposed cohort members in each estimation sample. Propensity scores were estimated from a probit regression ~~on the largest outcome-specific estimation sample (Rutter indices at age 7)~~.

4.2. Average Treatment Effect estimates

IPTW estimates of ATEs suggest that *in-utero* exposure to influenza led to small reductions in mental health in childhood at both ages 7 and 11 (Table 3). Reductions were smaller at age 7, where *in-utero* exposure led to a small but statistically insignificant 0.103 increase in the overall Rutter index; equivalent to 1.63% increase relative to the mean index for the non-exposed group. At age 11, the magnitude of the effect doubles to 0.194 (3.04%) and becomes statistically significant at a 5% level. At both ages, reductions in mental health were primarily driven by an increase in internalising behaviour (3.68% at age 11). Consistent with minimal differences in covariates, these estimates are almost identical to unweighted effects estimated using ordinary least squares.

In-utero influenza exposure also led to small increases in the number of depressive symptoms included in the 24-item Malaise inventory at ages 23, 33, and 42, ranging from an increase of 0.05 symptoms (1.3%) at age 42 to an increase of 0.08 symptoms (3.42%) at age 33, although these were not statistically significant (Table 1). At age 50, the magnitude rises to a 0.09 symptom increase (equivalent to 6.35% given the switch to the 9-item inventory at this age), but again this effect is not statistically significant. The lack of statistical significance despite moderate effect sizes is likely to be driven by reduced sample sizes at later ages. Differences between weighted and unweighted effects are larger in adulthood, consistent with greater differences in covariates prior to weighting.

4.3. Heterogeneity by gender

Effects differ considerably by gender, with this heterogeneity varying by the age at which mental health is measured (Table A5, Appendix A7). At age 7, effects on the overall Rutter index and internalising and externalising subscales are stronger for males, but remain statistically insignificant for both genders. Effects on the Rutter index at age 11 are stronger for females, driven by a statistically significant increase in internalising symptoms of 7.72%. However effects on externalising symptoms remain stronger for males (4.57% vs -2.05%) although effects are again not statistically significant.

Effects of *in-utero* influenza exposure on the Malaise inventory are not statistically significant for both genders at ages 23, 33, and 42, with some evidence of larger effects for females. However, at age 50, effects are much larger for males, with *in-utero* influenza exposure resulting in a statistcially significant 0.176 (15.45%) increase in depressive symptoms.

Table 3: Inverse probability weighted estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50

	Observations	Unweighted ATE	Potential outcome mean for the non-exposed group	Weighted ATE	Weighted ATE as % of non-exposed mean
Rutter: age 7					
Overall	12,291	0.103 (0.098)	6.312*** (0.035)	0.103 (0.095)	1.63%
Internalising	12,291	0.080 (0.061)	3.322*** (0.021)	0.076 (0.060)	2.29%
Externalising	12,291	0.026 (0.060)	2.980*** (0.021)	0.027 (0.058)	0.90%
Rutter: age 11					
Overall	11,430	0.189 (0.099)	6.355*** (0.035)	0.194* (0.098)	3.04%
Internalising	11,430	0.127* (0.061)	3.552*** (0.022)	0.131* (0.060)	3.68%
Externalising	11,430	0.060 (0.060)	2.796*** (0.021)	0.061 (0.058)	2.17%
Malaise: age 23	10,344	0.142 (0.085)	2.686*** (0.031)	0.054 (0.081)	1.99%
Malaise: age 33	9,318	0.156 (0.091)	2.372*** (0.033)	0.082 (0.089)	3.42%
Malaise: age 42	9,344	0.113 (0.110)	3.551*** (0.040)	0.046 (0.108)	1.29%
Malaise: age 50	8,002	0.133* (0.067)	1.449*** (0.023)	0.093 (0.066)	6.35%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.~~Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.~~

5. Robustness checks

5.1. Alternative methods for adjusting for residual differences in covariates

IPTW estimators model the probability of exposure, with no assumptions made about the functional form of the outcome model. The consistency of IPTW estimators relies solely on specifying the correct functional form of the probability of *in-utero* influenza exposure. To relax this assumption, we employ the inverse-probability-weighted regression adjustment (IPWRA) estimator which combines inverse-probability of treatment weighting with regression adjustment (Wooldridge, 2007). IPWRA is a member of a class of "doubly-robust" (DR) methods which are consistent when the functional form for either the probability of exposure or the outcome is correctly specified. IPWRA estimates are obtained via a four-step process (Wooldridge, 2010):

1. Inverse-probability of treatment weights are estimated as above.
2. A weighted regression of the outcomes on the set of covariates is performed seperately for the exposed and non-exposed groups, with weights equal to the inverse-probabiliy weights from step 1. Consistent with the outcome distributions (Appendix A2), we estimate linear models for the Rutter indices and Poisson models for the Malaise symptom counts.
3. Fitted values are generated for all cohort members from both regressions in step 2. These fitted values represent the predicted outcomes for each cohort member in each treatment state.
4. Cohort member-specific treatment effects are estimated as the difference between their predicted outcomes in each treatment state, and the ATE is estimated as average of these treatment effects across all cohort members.

IPWRA and IPW ATE estimates for the Rutter indices are very similar in magnitude, although the effect of influenza exposure on the overall Rutter index at age 11 becomes statistically insignificant at the 5% level (Table 4). Statistical significance of the effect on the internalising component of the Rutter index is maintained. For adult mental health outcomes, the use of IPWRA increases the magnitude of effects but they remain statistically insignificant at a 5% level.

Table 4: Inverse-probability-weighted regression adjustment estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50

	Observations	Potential outcome mean for the non-exposed group	ATE	ATE as % of non-exposed mean
Rutter: age 7				
Overall	12,291	6.313*** (0.035)	0.109 (0.095)	1.73%
Internalising	12,291	3.322*** (0.021)	0.081 (0.060)	2.44%
Externalising	12,291	2.980*** (0.021)	0.028 (0.059)	0.94%
Rutter: age 11				
Overall	11,430	6.355*** (0.035)	0.188 (0.098)	2.96%
Internalising	11,430	3.552*** (0.022)	0.130* (0.061)	3.66%
Externalising	11,430	2.796*** (0.021)	0.055 (0.058)	1.97%
Malaise: age 23	10,344	2.685*** (0.031)	0.061 (0.082)	2.27%
Malaise: age 33	9,318	2.372*** (0.033)	0.089 (0.090)	3.75%
Malaise: age 42	9,344	3.550*** (0.040)	0.063 (0.108)	1.77%
Malaise: age 50	8,002	1.449*** (0.023)	0.100 (0.067)	6.90%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

5.2. Sources of bias

5.2.1. Non-random sample selection

IPTW estimates could be impacted by two forms of non-random sample selection: mortality selection (survival bias) and survey non-response. Evidence suggests that it is infants at the bottom of the health distribution who die as a result of infection (Bozzoli et al., 2009). Any increase in mortality rates as a result of *in-utero* exposure to influenza therefore indicates that individuals in poor mental health may have been pruned from the exposed group, increasing mental health relative to the unexposed group. This subsequently leads to an under-estimation of the effects of influenza exposure.

~~To~~Figure 4 depicts how the percentage of NCDS cohort members who have died increases with age, and ~~examine effects on mortality,~~ Table 5 and Figure A4 (Appendix A~~7~~8) present IPTW estimates of the effect of *in-utero* influenza exposure on the probability of stillbirth/neo-natal death, death within 28 days of birth, and death by ages 7, 11, 16, 23, 33, 42, 46, and 50. For individuals not exposed to influenza *in-utero*, approximately 1.9% of individuals are stillborn, consistent with an infant mortality rate (age <1) of 2.4% in the UK for 1958 (CLOSER, 2020). Rates of death increase to 4.4% by age 7, 5.6% by age 33 and 7.8% by age 50.

A large effect of exposure is found at birth, where *in-utero* exposure to influenza increases the probability a stillbirth/neo-natal death by 1.2 percentage points. This is equivalent to 75% increase compared to the estimated stillbirth rate in the not-exposed group. This effect increases as the cohort ages, although not monotonically, to 1.7 percentage points by age 46. Effects are also stronger relative to unweighted estimated suggesting results are not explained by unobserved confounding. Further analysis suggests that most of the mortality effect occurs before birth, with effects at later ages being positive but losing statistical significance after conditioning on survival beyond birth (Figure A5, Appendix 8). These substantial effects on mortality suggest that survival bias may be causing IPTW estimates to be conservative estimates of the effects of *in-utero* influenza exposure on mental health, with this bias more pronounced at older ages.

Formatted: Font: 10 pt

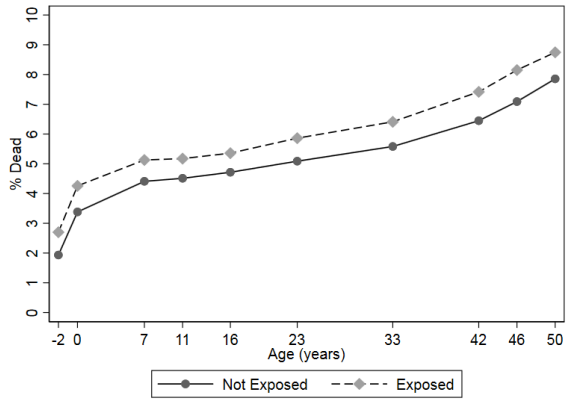


Figure 3: Proportion of NCDS cohort members who are dead at each age at follow-up. Age of -2 refers to birth, and age of 0 refers to 28 day post-birth.

Table 5: Unweighted and ~~I~~nverse probability weighted estimates of the ATE of *in-utero* exposure to influenza on the probability of death at birth, 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50.

	Observations	Unweighted	Potential outcome mean for the non-exposed group	ATE	ATE as % of non-exposed mean
Birth	15,019	0.010** (0.004)	0.016*** (0.001)	0.012** (0.004)	75.00%
28 days	15,019	0.012* (0.005)	0.029*** (0.001)	0.015** (0.005)	51.72%
Age 7	15,019	0.011* (0.005)	0.039*** (0.002)	0.013* (0.006)	33.33%
Age 11	15,019	0.010* (0.005)	0.040*** (0.002)	0.013* (0.006)	32.50%
Age 16	15,019	0.010 (0.005)	0.043*** (0.002)	0.013* (0.006)	30.23%
Age 23	15,019	0.012* (0.006)	0.046*** (0.002)	0.015* (0.006)	32.61%
Age 33	15,019	0.013* (0.006)	0.051*** (0.002)	0.016** (0.006)	31.37%
Age 42	15,019	0.013* (0.006)	0.059*** (0.002)	0.017* (0.007)	28.81%
Age 46	15,019	0.014* (0.006)	0.066*** (0.002)	0.017* (0.007)	25.76%
Age 50	15,019	0.012 (0.007)	0.074*** (0.002)	0.015* (0.007)	20.27%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
~~Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.~~

To examine the potential consequences of differential mortality rates on treatment effects for mental health outcomes, we use a bounding approach similar in spirit to Manski (1990) where "worst case" bounds are estimated by replacing missing values with the maximum or mimumum of the outcomes. Given the postive mortality effects found above, we depart slightly from this approach by assuming monotonicity (similar to assumptions made in the bounding approach in Lee (2009)), such that being exposed to influenza *in-utero* only increases mortality risk and not reduces it. Thus the lower bound is equivalent to the treatment effects from the main specification and the upper

bound equivalent to IPTW estimates after replacing missing values due to death with the maximum. In this most severe case, we cannot rule out substantial effects on mental health in childhood and adulthood, with effects on internalising symptoms at age 11 potentially as high as 8.37% relative to the non-exposed mean and effects on depressive symptoms at age 33 as high as 14.6% (Table A65, Appendix A98). As distributtions of the outcomes (Figures A1 and A2, Appendix A2) suggest outcome values at the maximum are extremely rare, we also estimate bounds instead replacing missing values with the sample 90th percentile. In this case the bounds are much tighter, but still don't rule out effects which are large in magnitude.

Survey non-completion is another possible source of sample selection bias. Even for the sample of individuals who survive, a proportion of NCDS mothers fail to complete the survey at the ages of interest. Consistent with arguments made in published studies using data from the NCDS (Dearden et al., 2002; Jones et al., 2011; Lindeboom et al., 2009), it is possible that the large number of characteristics available in the NCDS means that missing data can be assumed exogenous, conditional on covariates. A lack of correlation between *in-utero* exposure to influenza and non-completion would further reduce the likelihood that non-response is problematic. To test this, we estimate the effects of influenza exposure on the probability of response in each survey wave, controlling for all other covariates[21] and restricting the sample to cohort members who are alive at each wave. The latter allows us to separate the effects of non-response from that of mortality selection. This outcome incorporates both unit non-response, where NCDS mothers did not complete the entire survey wave, and outcome non-response. We find that effects of *in-utero* exposure on the probability of completion are small and not statistically significant at any survey wave (Table A76, Appendix A109), implying that survey non-completion is not a likely source of bias.

5.2.2. Unobserved confounding

As higlighted previously, consistency of IPTW estimates relies on the asumptions of no unobserved confounding, such that there exists no unobserved variables correlated with both influenza exposure and mental health. We therefore implement a series of tests to examine the robustness of results to unobserved confounding under the assumption that the relationship between *in-utero* influenza exposure and the unobservables can be recovered from the relationship between influenza exposure and observables (Altonji et al., 2011). Under this assumption, random assignment with respect to observed confounders implies random assignment with respect to unobserved confounders, and coefficient stability following the inclusion or removal of covariates provides evidence against unobserved confounding (Altonji et al., 2005).

First, we estimate a series of coefficient comparison tests, where we iteratively remove single, pairs, and triplets of all proxies of maternal health and re-estimate ATEs using IPTW. Tight clustering of ATEs around the baseline estimate provides evidence against unobserved confounding, by suggesting that the addition of single unobserved confounders may also not impact results greatly. These proxies include the co-morbidity indicator (which equals one if a mother reported suffering from other diseases during pregnancy) and indicators of mother's experience in previous pregnancies (premature births, large births, stillbirth/neo-natal deaths, and birth-related complications).

Histograms of re-estimated ATEs for child mental health outcomes (Figure 43) show that the removal of proxies result in only small changes in ATEs, with ATEs approximtely normally distributed around the ATEs estimated in

[21] Missing data on control variables from the PMS are implicitly assumed to be missing at random.

the main specification (shown by the vertical red lines). For some outcomes, the removal of covariates leads to a reduction in the ATE, suggesting that ATE from the main specification could represent a lower bound of the true ATE. Similar distributions are found for the ATEs for adult mental health outcomes (Figure 43), although there is some evidence that the removal of some covariates leads to an increase in the ATE. Differential non-response with respect to these covariates could be driving the larger effects of covariate-adjustment for these outcomes.

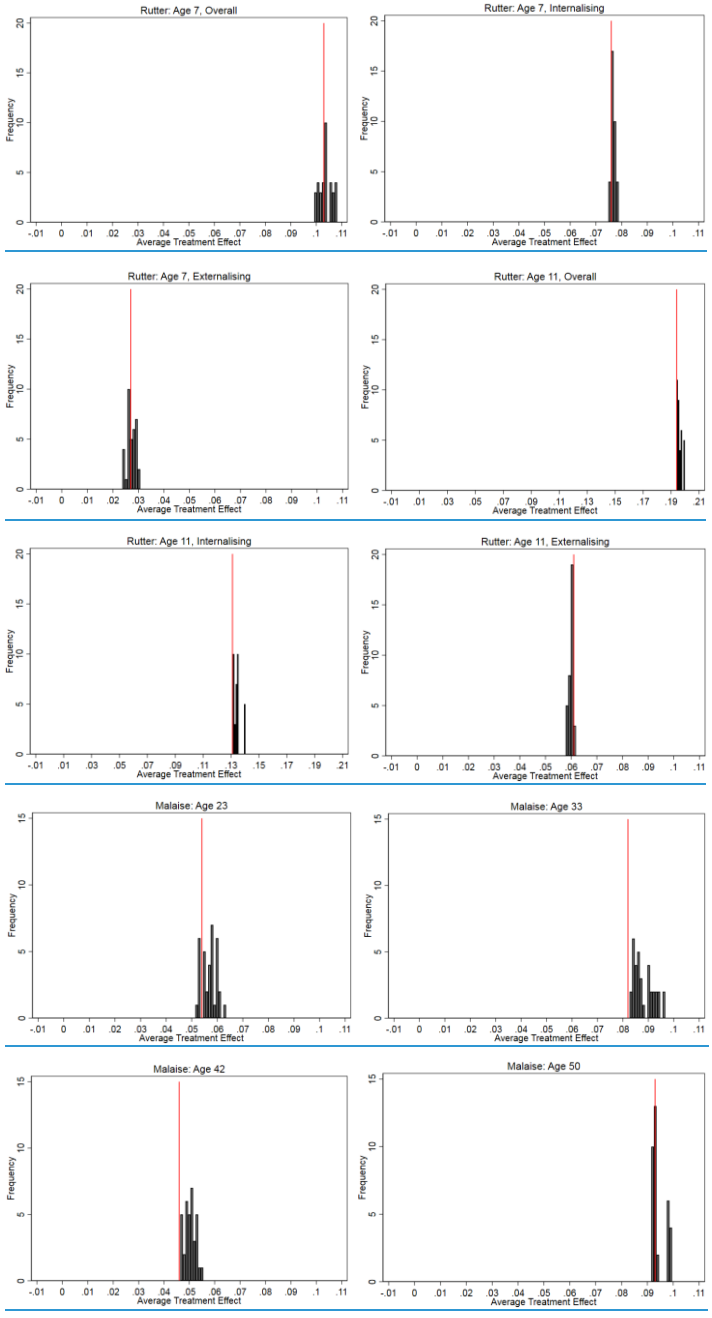
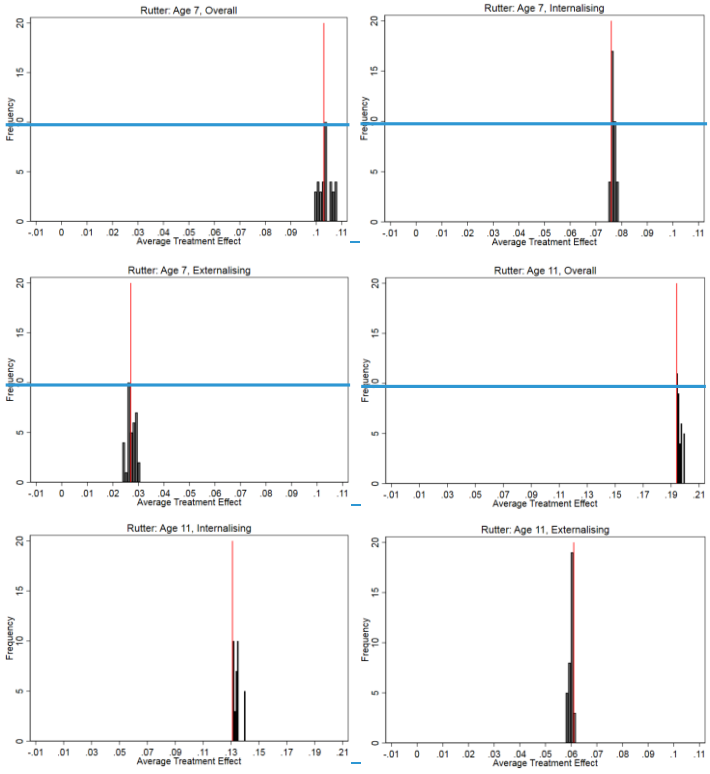


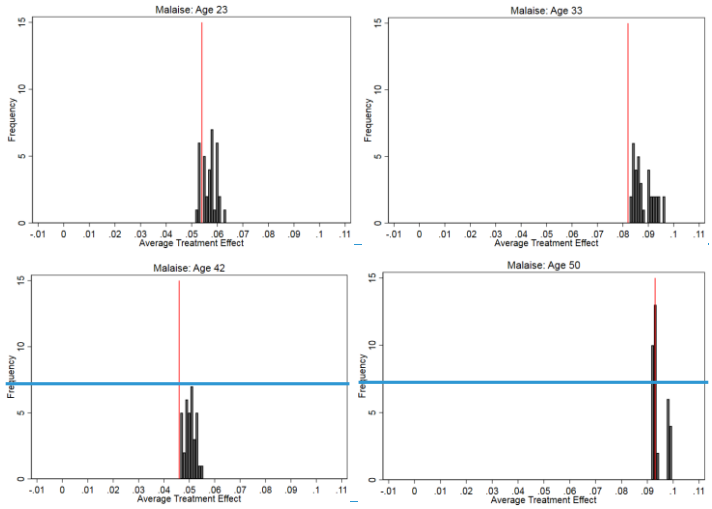
Figure 4: Histograms depicting variation in estimated average treatment effects due to the removal of single, pairs and triplets of proxies for genetic health endowments for mental health outcomes: Red vertical lines represent the average treatment effects with the full specification of covariates.

~~Comparisons of IPTW and unadjusted estimates of ATE corroborate these findings (Table A7, Appendix 10), showing that although ATE increases with the removal of covariates for adult mental health outcomes, unadjusted ATEs are almost identical to IPTW ATEs for child mental health outcomes and are slightly smaller than unadjusted ATEs for mortality outcomes.~~ The overall stability in ATEs is perhaps unsurprising given the small ~~unadjusted~~unweighted differences in covariates between those exposed and not-exposed to influenza *in-utero* shown in Table 1.

Second, we implement a series of balancing regressions where the main specification is re-estimated, but with proxies for genetic health endowments used as the outcome. This is preferred to traditional coefficient comparison tests in situations where proxy confounders could be measured with error (Pei et al., 2019). For each proxy, we remove the proxy from the covariate set, re-estimate propensity scores, and use IPTW to estimate the ATE of *in-utero* influenza on the proxy. Null effects on the proxy are suggestive of exogeneity. Results indicate that influenza exposure is not statistically significntly associated with any indicator of genetic health endowments after accounting for other covariates, and in some cases association between influenza exposure and markers of poor maternal health are negative (Table 6).

Results from both tests therefore suggest confounding from unobserved shared health endowments is unlikely to be biasing results considerably.





~~**Figure 3**: Histograms depicting variation in estimated average treatment effects due to the removal of single, pairs and triplets of proxies for genetic health endowments for mental health outcomes: Red vertical lines represent the average treatment effects with the full specification of covariates.~~

Formatted: Space Before: 12 pt

Table 6: Inverse-probability-weighted estimates of the ATE of *in-utero* exposure to influenza on maternal health characteristics in each estimation sample.

	Comorbidity	Previous premature birth	Previous large birth	Previous stillbirth/neo-natal death	Previous birth complication
Age 7 sample	-0.004	0.009	0.008	-0.005	0.008
	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)
Age 11 sample	-0.006	0.009	0.008	-0.003	0.005
	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)
Age 23 sample	-0.006	0.004	0.006	0.004	0.011
	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)
Age 33 sample	-0.006	0.005	0.010	0.002	0.014
	(0.008)	(0.008)	(0.008)	(0.006)	(0.009)
Age 42 sample	-0.008	0.002	0.004	0.002	0.013
	(0.007)	(0.007)	(0.008)	(0.006)	(0.009)
Age 50 sample	-0.006	0.000	0.010	0.002	0.007
	(0.008)	(0.008)	(0.008)	(0.007)	(0.010)

* p<0.05, ** p<0.01, *** p<0.001; robust standard errors in parenthesis

6. Discussion and Conclusion

A large body of research in economics has investigated the relationship between *in-utero* exposure to influenza and later-life health, human capital, and labour market outcomes. However, there is little evidence of effects on mental health outcomes despite numerous potential mechanisms. Evidence on mortality effects are also rare, with no study examining how mortality risk due to influenza exposure develops as an individual ages. In addition, the majority of studies estimate only reduced-form effects of exposure to influenza pandemics, with little evidence on the effects of influenza at an individual-level.

This paper adds to the existing literature by using unique survey data on mother-reported *in-utero* influenza exposure and longitudinal data on the mental health of their offspring to investigate impacts on ~~early death and~~ mental health outcomes recorded at multiple stages of the life-course and early death~~.~~ The data is also drawn from a cohort of individuals who were *in-utero* during the 1957 Asian Flu epidemic, increasing the likelihood of influenza exposure is independent of unobserved confounders. This is supported by the data, which shows that although influenza exposure was correlated with common predictors of disease spread (e.g. the number of siblings), exposed and non-exposed individuals were similar on markers of maternal health and there was no systematic pattern of socio-demographic ~~characteristics~~selection into maternal influenza. After adjusting for small residual differences between exposed and non-exposed groups using inverse probability weighting and "doubly-robust" inverse-probability-weighted regression adjustment, our findings suggest that *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms, and small increases in depressive symptoms in mid-life for males. We also find that *in-utero* influenza exposure is associated with a significant reduction in survival rates. This effect is driven primarily by a 75% increase in the relative risk of stillbirth, although we also find small increases in the magnitude of mortality effects in adulthood.

In addition, in a series of balancing regressions and coefficient comparison tests, we demonstrate limited sensitivity of results to removing maternal health and socio-demographic characteristics, particularly for child mental health and mortality outcomes. This, coupled with small unweighted differences in maternal health between exposed and non-exposed groups (with mothers of exposed individuals being in marginally better health on some markers) suggests that unobserved confounding is unlikely to be inflating the magnitudes of our estimates considerably. Nevertheless, the risk of some unobserved confounding remains.

Furthermore, substantial effects of *in-utero* exposure to influenza on early death indicate that estimated effects on mental health are conservative. This may explain the smaller and statistically insignificant effects we find on mental health in adulthood. Estimated bounds under survival bias could not rule out large effects on mental health outcomes. Also, as influenza cases are self-reported there is a potential for measurement error, which is likely to have further attenuated effects towards zero.

Our results are consistent with the limited number of studies finding mixed evidence on the effects of *in-utero* health shocks on mental health outcomes. Persson and Rossin-Slater (2018) find that exposure to maternal stress due to death of a close relative leads to increased use of ADHD medicine in childhood and anti-anxiety and depression medication in adulthood. Brown et al., (2000) find that exposure to the Dutch Winter Famine of 1944-45 led to an increase in the risk of admission for affective disorders. However, Maselko et al., (2015) find that a reduction in prenatal depression caused by a maternal cognitive behavioural therapy intervention did not lead to an increase in socio-emotional development of their offspring.

Our findings of generally stronger health effects of influenza exposure for males are consistent with evidence from other *in-utero* shocks. Havari and Peracchi (2017), Scholte et al. (2015), van den Berg et al. (2015), and Lindeboom et al. (2010) all find that the health effects of *in-utero* exposure to under-nutrition are stronger for males. Similarly, Arthi (2018) find that the effect of *in-utero* exposure to the Dust Bowl on physical disability is statistically significant only for males. These findings are also consistent with previous studies examining the effects of *in-utero* pandemic

exposure on human capital and family formation outcomes, with effects on educational attainment and the probability of marriage being stronger for males (Lin and Liu, 2014; Neelsen and Stratmann, 2012).

Our finding of a significant effect on stillbirth rates is consistent with the medical literature that states that influenza can lead to complications in late pregnancy which can subsequently lead to stillbirths (Liu et al., 2012; Pierce et al., 2011). We also find that the magnitude of the effect on the probability of death remained relatively constant up to age 50, indicating a persistence in the survival advantage for those not exposed to influenza *in-utero*, but no large additional survival advantage for those living beyond infancy. This is consistent with the foetal origins hypothesis which states that the health effects of *in-utero* adversity may remain latent until old age, when the onset of physical disease becomes prevalent (Barker, 1990), and corroborates findings in Fletcher (2019) who find no effects of exposure to the 1918 influenza pandemic on mortality by age 60 for individuals surviving to adulthood. This is also consistent with an evaluation of the long-term health effects of a neonatal program in Sweden, where the mortality advantage for the treated individuals began in infancy and remained constant throughout adulthood, before increasing again at age 70 (Bhalotra et al., 2015). A study examining impacts of influenza exposure on mortality at later ages is warranted once data on future NCDS survey waves become available.

One limitation of our paper is that, if the 1957 Asian Flu was responsible for the majority of the influenza cases in this cohort, our study will only examine the effects of *in-utero* exposure to influenza in a relatively small period of gestation (between 16 and 25 weeks). Therefore, given all NCDS cohort members were born in the same week, we are not able to examine whether there is heterogeneity in effects by trimester of exposure. Furthermore, a lack of statistically significant effects on mental health at some ages may be driven by a lack of power to detect these effects due to limited sample sizes. Examination of this question using longitudinal data on consistent measures of mental health across all ages would be an interesting extension of this study. ~~Furthermore~~Finally, external validity must be considered when interpreting results. The NCDS is a historical cohort and our study documents the effects of *in-utero* exposure to influenza in the 1950s. It is unclear whether exposure to influenza would have similar deleterious effects in more recent cohorts. However, the requirement for contemporaneously-measured exposure and outcomes means this limitation is unavoidable in studies such as this.

Despite these limitations, this paper presents findings which are relevant for policy-makers. In the US and UK, pregnant women are advised to have an influenza vaccine to protect against the risk of stillbirths and other pregnancy-related complications, with vaccination also being free-at-the-point-of-use in the UK (Centres for Disease Control and Prevention, 2013; NHS, 2015; Public Health England, 2015). However, recommendations are not ubiquitous elsewhere. Recommendations are still not in place in 5/31 countries in the European Economic Area (ECDPC, 2015), and a World Health Organisation (WHO) report suggests that 37/47 countries were not recommending influenza vaccinations for all pregnant women for the 2008-09 influenza season (World Health Organization, 2014). Furthermore, a 2011 survey found that under 51% of the 37 countries in the Western Pacific Region recommended vaccines for pregnant women (Dwyer et al., 2013).

In addition, even in countries where recommendations are in place, uptake of the influenza vaccines amongst pregnant women is still low. In the 2013-14 influenza season, uptake rates in the US stood at only 52% (Centers for Disease Control and Prevention, 2014), and in the UK, uptake rates fell from 47.2% to 45.2% between the 2017/18 and 2018/19 winter influenza seasons (Public Health England, 2019). A 2014 report found that influenza led to 1 in

11 of maternal deaths, and estimated that over half of influenza deaths could have been prevented with influenza vaccinations (Knight et al., 2014). Our findings, alongside those from previous studies, suggest the negative impacts of influenza during pregnancy extend beyond the mother to their unborn children. Given that current evidence indicates that influenza vaccines are successful in protecting mothers from influenza, and that there are no serious side-effects of these vaccines for either mother or infant (Mak et al., 2008), our results suggest that the roll-out of recommendations for influenza vaccinations for pregnant women across all countries should be explored and that interventions to improve uptake in countries where recommendations are in place are warranted.

7. Bibliography

Almond, D., 2006. Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population. J. Polit. Econ. 114, 672–712.

Almond, D., Currie, J., Duque, V., 2018. Childhood circumstances and adult outcomes: Act II. J. Econ. Lit. 56, 1360–1446. doi:10.1257/jel.20171164

Almond, D., Mazumder, B., 2005. The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data. Am. Econ. Rev. 95, 258–262.

Altonji, J.G., Conley, T.G., Elder, T.E., Taber, C.R., 2011. Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables. Mimeo.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. J. Polit. Econ. 113, 151–184. doi:10.1086/426036

Arthi, V., 2018. "The dust was long in settling": Human capital and the lasting impact of the American Dust Bowl. J. Econ. Hist. 78, 196–230. doi:10.1017/S0022050718000074

Austin, P.C., Stuart, E., 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat. Med. 34, 3661–3679. doi:10.1002/sim.6607

Barker, D., 1990. The fetal and infant origins of adult disease. BMJ Br. Med. J. 301, 1111.

Bhalotra, S., Karlsson, M., Nilsson, T., 2015. Infant health and longevity: evidence from a historical trial in Sweden (No. 2015–08), ISER Working Paper Series.

Blanden, J., Gregg, P., Macmillan, L., 2007. Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. Econ. J. 117, C43–C60. doi:10.1111/j.1468-0297.2007.02034.x

Bozzoli, C., Deaton, A., Quintana-Domeque, C., 2009. Adult height and childhood disease. Demography 46, 647–69. doi:10.1353/dem.0.0079

Brown, A.S., Van Os, J., Driessens, C., Hoek, H.W., Susser, E.S., 2000. Further evidence of relation between prenatal famine and major affective disorder. Am. J. Psychiatry 157, 190–195. doi:10.1176/appi.ajp.157.2.190

Busso, M., DiNardo, J., McCrary, J., 2014. New evidence on the finite sample properties of propensity score reweighting and matching estimators. Rev. Econ. Stat. 96, 885–897. doi:10.1162/REST_a_00431

Butler, N.R., Bonham, D.G., 1963. Perinatal Mortality. E & S Livingstone, Edinburgh.

Campion, E.W., Kourtis, A.P., Read, J.S., Jamieson, D.J., 2014. Pregnancy and Infection. N. Engl. J. Med. 23370, 2211–2218. doi:10.1056/NEJMra1213566

Cassini, A., Colzani, E., Pini, A., Mangen, M.-J.J., Plass, D., McDonald, S.A., Maringhini, G., van Lier, A., Haagsma, J.A., Havelaar, A.H., Kramarz, P., Kretzschmar, M.E., 2018. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. Eurosurveillance 23. doi:10.2807/1560-7917.es.2018.23.16.17-00454

Centers for Disease Control and Prevention, 2015. About Adult BMI [WWW Document]. URL http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/ (accessed 4.20.16).

Centers for Disease Control and Prevention, 2014. Influenza Vaccination Coverage Among Pregnant Women — United States, 2013–14 Influenza Season. Morb. Mortal. Wkly. Rep. 63, 816–821.

Centre for Longitudinal Studies, 2012. Measuring behaviour difficulties of children and teenagers.

Centres for Disease Control and Prevention, 2013. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2013–2014.

Clarke, S.K.R., Heath, R.B., Sutton, R.N.P., Stuart-Harris, C.H., 1958. SEROLOGICAL STUDIES WITH ASIAN STRAIN OF INFLUENZA A. Lancet 271, 814–818. doi:10.1016/S0140-6736(58)91739-2

CLOSER, 2020. Infant mortality rate (deaths under 1 year per 1,000 live births), 1930 - 2019 (United Kingdom) [WWW

Document]. URL https://www.closer.ac.uk/data/infant-mortality/ (accessed 4.15.21).

Cole, S.R., Hernán, M.A., 2008. Constructing inverse probability weights for marginal structural models. Am. J. Epidemiol. 168, 656–664. doi:10.1093/aje/kwn164

Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S., Bordin, I.A., Costello, E.J., Durkin, M., Fairburn, C., Glass, R.I., Hall, W., Huang, Y., Hyman, S.E., Jamison, K., Kaaya, S., Kapur, S., Kleinman, A., Ogunniyi, A., Otero-Ojeda, A., Poo, M.-M., Ravindranath, V., Sahakian, B.J., Saxena, S., Singer, P.A., Stein, D.J., Anderson, W., Dhansay, M.A., Ewart, W., Phillips, A., Shurin, S., Walport, M., Timmermans, H., 2011. Grand challenges in global mental health. Nature 475, 37–30. doi:10.1038/475027a.Grand

Collishaw, S., 2015. Annual research review: Secular trends in child and adolescent mental health. J. Child Psychol. Psychiatry Allied Discip. 56, 370–393. doi:10.1111/jcpp.12372

Colombo, J., Kannass, K.N., Shaddy, D.J., Kundurthi, S., Maikranz, J.M., Anderson, C.J., Blaga, O.M., Carlson, S.E., 2004. Maternal DHA and the development of attention in infancy and toddlerhood. Child Dev.

Coneus, K., Spiess, C.K., 2012. The intergenerational transmission of health in early childhood - Evidence from the German Socio-Economic Panel Study. Econ. Hum. Biol. 10, 89–97. doi:10.1016/j.ehb.2011.03.002

Currie, J., 2020. Child health as human capital. Health Econ. Online onl. doi:10.1002/hec.3995

Currie, J., Almond, D., 2011. Human capital development before age five, in: Card, D., Ashenfelter, T. (Eds.), Handbook of Labour Economics. North-Holland, Amsterdam.

Dearden, L., Ferri, J., Meghir, C., 2002. The effect of school quality on educational attainment and wages. Rev. Econ. Stat. 84, 1–20.

Dehejia, R., Wahba, S., 2002. Propensity Score Matching Methods for Non-experimental Causal Studies. Rev. Econ. Stud. 84, 151–161. doi:10.2139/ssrn.138259

Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. J. Am. Stat. Assoc. 94, 1053–1062.

Dwyer, D., Barr, I., Hurt, A., Kelso, A., Reading, P., Sullivan, S., Buchy, P., Yu, H., Zheng, J., Shu, Y., Wang, D., Lam, Aguon, A., Oliva, R.Q., Odagiri, T., Tashiro, M., Verasahib, K., Yusof, M.A., Nymadawa, P., Alexander, B., Gourinat, A.C., Grangeon, J.P., Jennings, L., Huang, S., Horwood, P., Lucero, M., Roque Jr., V., Lee Suy, L., Cardon, P., Tandoc 3rd, A., Olveda, R.M., Kang, C., Young-Joon, P., Cutter, J., Lin, R., Low, C., Mai le, T.Q., Balish, A., Kile, J., Mei, S., McFarland, J., Moen, A., Olsen, S., Samaan, G., Xiyan, X., Chea, N., Diorditsa, S., Feldon, K., Fox, K., Jamsran, M., Konings, F., Lewis, H.C., McPherson, M., Nilles, E., Olowokure, B., Partridge, J., 2013. Seasonal influenza vaccine policies, recommendations and use in the World Health Organization's Western Pacific Region, Western Pacific Surveillance and Response Journal. doi:10.5365/WPSAR.2013.4.1.009

ECDPC, 2015. Seasonal influenza vaccination in Europe: Overview of vaccination recommendations and coverage rates in the EU Member States for the 2012–13 influenza season. Stockholm.

Ferrari, A., Charlson, F., Norman, R., Patten, S., Freedman, G., Murray, C., Vos, T., Whiteford, H., 2017. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. PLoS Med. 10, 319–327. doi:10.1371/journal.pmed.1001547

Fletcher, J.M., 2019. Examining the long-term mortality effects of early health shocks. Appl. Econ. Lett. 26, 902–908. doi:10.1080/13504851.2018.1520960

Fletcher, J.M., 2018. The effects of in utero exposure to the 1918 influenza pandemic on family formation. Econ. Hum. Biol. 30, 59–68. doi:10.1016/j.ehb.2018.06.004

Gale, C.R., Robinson, S.M., Godfrey, K.M., Law, C.M., Schlotz, W., O'Callaghan, F.J., 2008. Oily fish intake during pregnancy - Association with lower hyperactivity but not with higher full-scale IQ in offspring. J. Child Psychol. Psychiatry Allied Discip. 49, 1061–1068.

Garthwaite, C., 2008. The effect of in-utero conditions on long term health: evidence from the 1918 Spanish flu pandemic.

Havari, E., Peracchi, F., 2017. Growing up in wartime: Evidence from the era of two world wars. Econ. Hum. Biol. 25, 9–32. doi:10.1016/j.ehb.2016.09.002

Hirst, M.A., 1983. Evaluating the Malaise Inventory - An item analysis. Soc. Psychiatry 18, 181–184. doi:10.1007/BF00583528

Hirst, M.A., Bradshaw, J.R., 1983. Evaluating the Malaise inventory: A comparison of measures of stress. J. Psychosom. Res. 27, 193–199. doi:10.1016/0022-3999(83)90022-3

Hunter, J., Young, J., 1971. Diffusion of influenza in England and Wales. Ann. Assoc. Am. Geogr. 61.

Iacus, S.M., King, G., Porro, G., 2011. Causal Inference without Balance Checking: Coarsened Exact Matching. Polit. Anal. 20, 1–24. doi:10.1093/pan/mpr013

Imai, K., Ratkovic, M., 2014. Covariate balancing propensity score. J. R. Stat. Soc. Ser. B Stat. Methodol. 76, 243–263. doi:10.1111/rssb.12027

Imbens, G., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. Rev. Econ. Stat. 86, 4–29.

Institute of Education, 2015a. National Child Development Study Deaths Dataset, 1958-2014: Special Licence Access (SN7717). Accessed via the UK Data Service.

Institute of Education, 2015b. National Child Development Study: Local Authority Data, 1958-1974: Special Licence Access (SN5744). Accessed via the UK Data Service.

Iuliano, A.D., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Gran, J.M., Schanzer, D., Cowling, B.J., Wu, P., Kyncl, J., Ang, L.W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., van Asten, L., Pereira da Silva, S., Aungkulanon, S., Buchholz, U., Widdowson, M.A., Bresee, J.S., Azziz-Baumgartner, E., Cheng, P.Y., Dawood, F., Foppa, I., Olsen, S., Haber, M., Jeffers, C., MacIntyre, C.R., Newall, A.T., Wood, J.G., Kundi, M., Popow-Kraupp, T., Ahmed, M., Rahman, M., Marinho, F., Sotomayor Proschle, C.V., Vergara Mallegas, N., Luzhao, F., Sa, L., Barbosa-Ramírez, J., Sanchez, D.M., Gomez, L.A., Vargas, X.B., Acosta Herrera, a. B., Llanés, M.J., Fischer, T.K., Krause, T.G., Mølbak, K., Nielsen, J., Trebbien, R., Bruno, A., Ojeda, J., Ramos, H., an der Heiden, M., del Carmen Castillo Signor, L., Serrano, C.E., Bhardwaj, R., Chadha, M., Narayan, V., Kosen, S., Bromberg, M., Glatman-Freedman, A., Kaufman, Z., Arima, Y., Oishi, K., Chaves, S., Nyawanda, B., Al-Jarallah, R.A., Kuri-Morales, P.A., Matus, C.R., Corona, M.E.J., Burmaa, A., Darmaa, O., Obtel, M., Cherkaoui, I., van den Wijngaard, C.C., van der Hoek, W., Baker, M., Bandaranayake, D., Bissielo, A., Huang, S., Lopez, L., Newbern, C., Flem, E., Grøneng, G.M., Hauge, S., de Cosío, F.G., de Moltó, Y., Castillo, L.M., Cabello, M.A., von Horoch, M., Medina Osis, J., Machado, A., Nunes, B., Rodrigues, A.P., Rodrigues, E., Calomfirescu, C., Lupulescu, E., Popescu, R., Popovici, O., Bogdanovic, D., Kostic, M., Lazarevic, K., Milosevic, Z., Tiodorovic, B., Chen, M., Cutter, J., Lee, V., Lin, R., Ma, S., Cohen, A.L., Treurnicht, F., Kim, W.J., Delgado-Sanz, C., de mateo Ontañón, S., Larrauri, A., León, I.L., Vallejo, F., Born, R., Junker, C., Koch, D., Chuang, J.H., Huang, W.T., Kuo, H.W., Tsai, Y.C., Bundhamcharoen, K., Chittaganpitch, M., Green, H.K., Pebody, R., Goñi, N., Chiparelli, H., Brammer, L., Mustaquim, D., 2018. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet 391, 1285–1300. doi:10.1016/S0140-6736(17)33293-2

Jamieson, D.J., Theiler, R.N., Rasmussen, S.A., 2006. Emerging Infections and Pregnancy. Emerg. Infect. Dis. 12, 1638–1643.

Johnson, J., Brown, M., 2015. National Child Development Study: User Guide to the Response and Deaths Datasets.

Jones, A.M., Rice, N., Rosa Dias, P., 2011. Long-Term Effects of School Quality on Health and Lifestyle: Evidence from Comprehensive Schooling Reforms in England. J. Hum. Cap. 5, 342–376. doi:10.1086/662441

Kelly, E., 2011. The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children. J. Hum. Resour.

Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Ustün, T.B., 2007. Age of onset of mental disorders: a review of recent literature. Curr. Opin. Psychiatry 20, 359–364.

Klein, J.M., Gonçalves, A., Silva, C.F., 2009. The Rutter Children Behaviour Questionnaire for teachers: from psychometrics to norms, estimating caseness. Psico-USF (Impresso) 14, 157–165. doi:10.1590/S1413-82712009000200004

Knight, M., Kenyon, S., Brocklehurst, P., Neilson, J., Shakespeare, J., Kurinczuk, J., 2014. Saving Lives, Improving Mothers' Care - Lessons learned to inform future maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-12. Oxford. doi:10.1021/cn400230x

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76, 1071–1102. doi:10.1111/j.1467-937X.2009.00536.x

Lin, M.J., Liu, E.M., 2014. Does in utero exposure to illness matter? The 1918 influenza epidemic in taiwan as a natural experiment. J. Health Econ. 37, 152–163.

Lindeboom, M., Llena-Nozal, A., van der Klaauw, B., 2009. Parental education and child health: evidence from a schooling reform. J. Health Econ. 28, 109–31. doi:10.1016/j.jhealeco.2008.08.003

Lindeboom, M., Portrait, F., van den Berg, G.J., 2010. Long-run effects on longevity of a nutritional shock early in life: The Dutch Potato famine of 1846-1847. J. Health Econ. 29, 617–629. doi:10.1016/j.jhealeco.2010.06.001

Liu, S., Wang, J., Yang, X., Chen, J., Huang, R., Ruan, B., He, H., Wang, C., Zhang, H., Sun, Z., Xie, L., Zhuang, H., 2012. Pandemic influenza A(H1N1) 2009 virus in pregnancy. Rev. Med. Virol. 1957. doi:10.1002/rmv.1712

Mak, T.K., Mangtani, P., Leese, J., Watson, J.M., Pfeifer, D., 2008. Influenza vaccination in pregnancy: current evidence and selected national policies. Lancet Infect. Dis. 8, 44–52. doi:10.1016/S1473-3099(07)70311-0

Manski, C.F., 1990. Nonparametric Bounds on Treatment Effects. Am. Econ. Rev. 80, 829–823. doi:10.2307/2006592

Maselko, J., Sikander, S., Bhalotra, S., Bangash, O., Ganga, N., Mukherjee, S., Egger, H., Franz, L., Bibi, A., Liaqat, R., Kanwal, M., Abbasi, T., Noor, M., Ameen, N., Rahman, A., 2015. Effect of an early perinatal depression intervention on long-term child development outcomes: follow-up of the Thinking Healthy Programme randomised controlled trial. The Lancet Psychiatry 0366, 1–9. doi:10.1016/S2215-0366(15)00109-1

Mazumder, B., Almond, D., Park, K., Crimmins, E.M., Finch, C.E., 2010. Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. J. Dev. Orig. Health Dis.

McGee, R., Williams, S., Silva, P.A., 1986. An evaluation of the Malaise inventory. J. Psychosom. Res. 30, 147–152. doi:10.1016/0022-3999(86)90044-9

McManus, S., Gunnell, D., Cooper, C., Bebbington, P.E., Howard, L.M., Brugha, T., Jenkins, R., Hassiotis, A., Weich, S., Appleby, L., 2019. Prevalence of non-suicidal self-harm and service contact in England, 2000–14: repeated cross-sectional surveys of the general population. The Lancet Psychiatry 6, 573–581. doi:10.1016/S2215-0366(19)30188-9

Ministry of Health, 1960. Reports on Public Health and Medical Subjects. The influenza pandemic in England and Wales 1957–58. London.

Neelsen, S., Stratmann, T., 2012. Long-run effects of fetal influenza exposure: Evidence from Switzerland. Soc. Sci. Med. 74, 58–66.

Nelson, R.E., 2010. Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic. Health Econ. 19, 1181–1192.

NHS, 2015. Flu, your pregnancy an you [WWW Document]. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448958/PHE_8879_Flu_Pregnancy_2014_DL__A5_leaflet_05_.pdf

OECD, 2018. Health at a Glance: Europe 2018. doi:10.1787/health_glance_eur-2018-en

Parsons, A.G., Zhou, S.J., Spurrier, N.J., Makrides, M., 2008. Effect of iron supplementation during pregnancy on the behaviour of children at early school age: long-term follow-up of a randomised controlled trial. Br J Nutr 99, 1133–1139. doi:S0007114507853359 [pii]\n10.1017/S0007114507853359 [doi]

Pei, Z., Pischke, J.S., Schwandt, H., 2019. Poorly Measured Confounders are More Useful on the Left than on the Right. J. Bus. Econ. Stat. 37, 205–216. doi:10.1080/07350015.2018.1462710

Percoco, M., 2016. Health Shocks and Human Capital Accumulation: The Case of Spanish Flu in Italian Regions. Reg. Stud. 50, 1496–1508. doi:10.1080/00343404.2015.1039975

Persson, P., Rossin-Slater, M., 2018. Family ruptures, stress, and the mental health of the next generation. Am. Econ. Rev. 108, 1214–1252. doi:10.1257/aer.20161124

Pierce, M., Kurinczuk, J.J., Spark, P., Brocklehurst, P., Knight, M., 2011. Perinatal outcomes after maternal 2009/H1N1 infection: national cohort study. BMJ 342, d3214. doi:10.1136/bmj.d3214

Power, C., Elliott, J., 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). Int. J. Epidemiol. 35, 34–41. doi:10.1093/ije/dyi183

Public Health England, 2019. Surveillance of influenza and other respiratory viruses in the UK Winter 2018 to 2019. London.

Public Health England, 2015. The national flu immunisation programme 2015/16.

Rasmussen, S.A., Jamieson, D.J., Bresee, J.S., 2008. Pandemic influenza and pregnant women. Emerg. Infect. Dis.

Registrar General for England and Wales, 1957. The Registrar General's Weekly Return for England and Wales. Births and Deaths, Infectious Diseases, and Weather. London.

Registrar General for Scotland, 1957. Weekly Return of the Registrar General Scotland. Births, Deaths, and Marriages. Infectious Diseases. Weather report. Edinburgh.

Rosenbaum, P., 1987. Model-based direct adjustment. J. Am. Stat. Assoc. 82, 387–394. doi:10.1080/01621459.1987.10478441

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika

70, 41–55.

Rubin, D.B., 2001. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Heal. Serv. Outcomes Res. Methodol. 2, 169–188. doi:10.1017/cbo9780511810725.030

Rutter, M., Tizard, J., Whitmore, K., 1970. Education, health and behaviour. Longman, London.

Sadler, K., Vizard, T., Ford, T., Goodman, A., Goodman, R., McManus, S., 2018. The Mental Health of Children and Young People in England 2017: Trends and characteristics. London.

Schlotz, W., Phillips, D.I.W., 2009. Fetal origins of mental health: Evidence and mechanisms. Brain. Behav. Immun. 23, 905–916.

Scholte, R.S., van den Berg, G.J., Lindeboom, M., 2015. Long-run effects of gestation during the dutch hunger winter famine on labor market and hospitalization outcomes. J. Health Econ. 39, 17–30. doi:10.1016/j.jhealeco.2014.10.002

Schwandt, H., 2017. The Lasting Legacy of Seasonal Influenza: In-utero Exposure and Labor Market Outcomes (No. 5/2017), COHERE discussion paper. doi:10.1017/CBO9781107415324.004

Sprahue, A., 1988. Post-war fertility and female labour force participation rates. Econ. J. 98, 682–700.

Stuart, E., 2010. Matching methods for causal inference: A review and a look forward. Stat. Sci. 25, 1–21. doi:10.1214/09-STS313.Matching

Tomkins, A., Murray, S., Rondo, P., Filteau, S., 1994. Impact of maternal infection on foetal growth and nutrition. SCN News 11, 18–20.

van den Berg, G.J., Pinger, P.R., Schoch, J., 2015. Instrumental Variable Estimation of the Causal Effect of Hunger Early in Life on Health Later in Life. Econ. J. n/a-n/a. doi:10.1111/ecoj.12250

Vynnycky, E., Edmunds, W.J., 2008. Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol. Infect. 136, 166–179. doi:10.1017/S0950268807008369

Woodall, J., Rowson, K., McDonald, J., 1958. Age and Asian influenza, 1957. Br. Med. J. 2, 1316–1318.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd ed. MIT Press, Cambridge, MA. doi:10.1016/B978-0-323-05712-7.00031-3

Wooldridge, J.M., 2007. Inverse probability weighted estimation for general missing data problems. J. Econom. 141, 1281–1301. doi:10.1016/j.jeconom.2007.02.002

World Health Organization, 2014. Evaluation of seasonal influenza vaccination policies and coverage in the WHO European Region: Results from the 2008/2009 and 2009/2010 influenza seasons. Copenhagen.

World Health Organization, 2013. Mental Health Action Plan 2013-2020 [WWW Document]. WHO Libr. Cat. DataLibrary Cat. Data. URL http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf

Supplementary appendices

Appendix A1: Item-level missing data for the Rutter index at ages 7 and 11

Table A1: Number of completed Rutter items at age 7 and age 11, conditional on mother completing at least one item

Age 7			Age 11		
Items completed	Frequency	Percentage	Items completed	Frequency	Percentage
1	1	0.01	1	0	<0.01
2	3	0.02	2	1	0.01
3	5	0.03	3	3	0.02
4	3	0.02	4	1	0.01
5	4	0.03	5	2	0.01
6	2	0.01	6	0	<0.01
7	3	0.02	7	2	0.01
8	5	0.03	8	4	0.03
9	2	0.01	9	6	0.04
10	13	0.09	10	10	0.07
11	23	0.16	11	45	0.33
12	111	0.76	12	184	1.33
13	782	5.35	13	1,139	8.25
14	13,651	93.45	14	12,408	89.88

All figures are to 2 decimal places

Appendix A2: Distributions of the mental health outcomes in childhood and adulthood

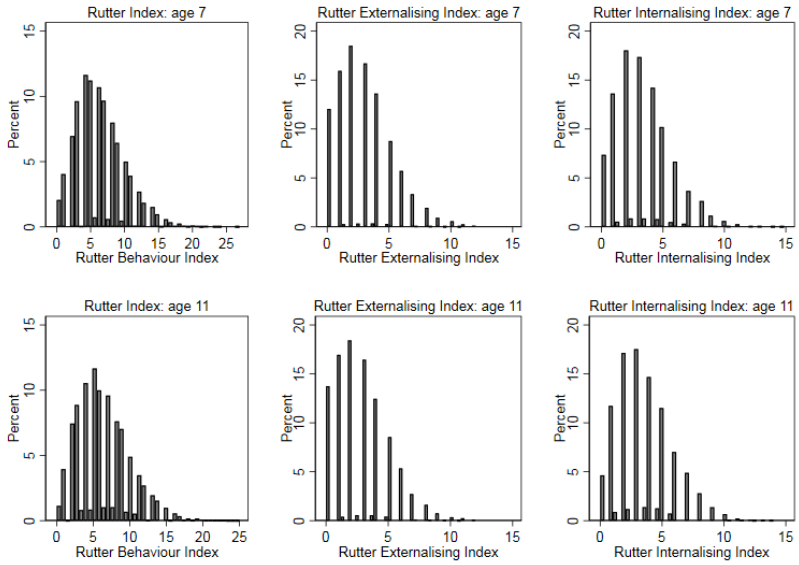


Figure A1: Distributions of the Rutter Behaviour Indices at ages 7 and 11

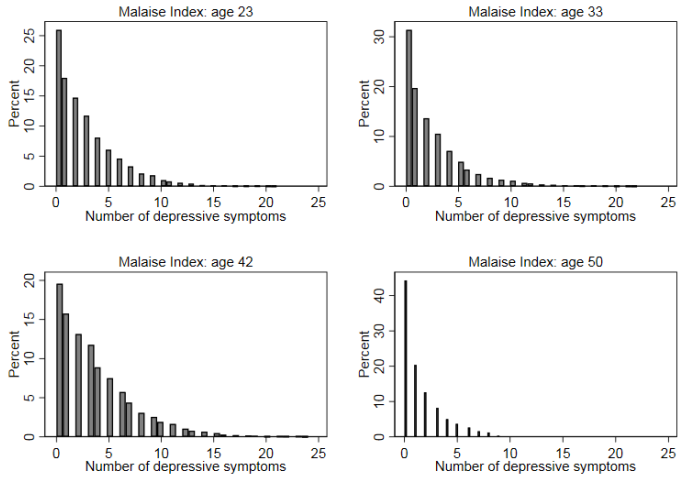


Figure A2: Distributions of the Malaise Index at ages 23, 33, 42 and 50. Derived from 24-item Malaise Inventory at ages 23, 33 and 42, and from the 9-item inventory at age 50.

Appendix A3: Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

Table A2: Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

	Not exposed to influenza *in-utero* (86.8%)	**Exposed to influenza *in-utero* *(13.2%)***
	N (%) or Mean (SD)	**N (%) or Mean (SD)**
Male gender	5490 (51.45%)	818 (50.49%)
Ethnic minority	116 (1.09%)	14 (0.86%)
Mother's occupational social class		
Unemployed	6592 (61.77%)	974 (60.12%)
I or II: Professional, managerial or technical	485 (4.55%)	77 (4.75%)
III: Skilled manual/non-manual	2314 (21.68%)	358 (22.10%)
IV: Partly skilled	1014 (9.50%)	169 (10.43%)
V: Unskilled	266 (2.49%)	42 (2.59%)
Father's occupational social class		
I: Professional	445 (4.17%)	52 (3.21%)
II: Managerial or technical	1420 (13.31%)	164 (10.12%)
III: Skilled manual/non-manual	6206 (58.16%)	1003 (61.91%)
IV: Partly skilled	1235 (11.57%)	192 (11.85%)
V: Unskilled	943 (8.84%)	145 (8.95%)
Information missing	422 (3.95%)	64 (3.95%)
Mother schooled past leaving age	2723 (25.52%)	386 (23.83%)
Mother's age in years		
20 or under	982 (9.20%)	164 (10.12%)
21-25	3288 (30.81%)	542 (33.46%)
26-30	3440 (32.24%)	471 (29.07%)
31-35	1881 (17.63%)	289 (17.84%)
36-40	873 (8.18%)	130 (8.02%)
Over 40	207 (1.94%)	24 (1.48%)
Mother's height (inches)	63.43 (2.49)	63.49 (2.56)
Father's age in years		
20 or under	208 (1.95%)	25 (1.54%)
21-25	2111 (19.78%)	354 (21.85%)
26-30	3454 (32.37%)	522 (32.22%)
31-35	2543 (23.83%)	371 (22.90%)
36-40	1310 (12.28%)	192 (11.85%)
41-45	517 (4.84%)	90 (5.56%)
45-50	178 (1.67%)	22 (1.36%)
Over 50	72 (0.67%)	3 (0.19%)
Information missing	278 (2.61%)	41 (2.53%)
Mother's smoking prior to pregnancy		
Non-smoker	6318 (59.21%)	969 (59.81%)
1-4 cigarettes per day	820 (7.68%)	135 (8.33%)
5-9 cigarettes per day	1494 (14.00%)	220 (13.58%)
10-14 cigarettes per day	1459 (13.67%)	213 (13.15%)
15-19 cigarettes per day	306 (2.87%)	39 (2.41%)
20+ cigarettes per day	274 (2.57%)	44 (2.72%)
Weight prior to pregnancy based on body mass index (BMI)		
Normal/healthy [18.5<BMI<25]	7787 (72.97%)	1170 (72.22%)
Underweight [BMI<18.5]	403 (3.78%)	41 (2.53%)
Overweight [25<BMI<30]	2028 (19.00%)	327 (20.19%)
Obese [BMI>30]	453 (4.25%)	82 (5.06%)
Antenatal visits within 7 weeks of pregnancy	432 (4.05%)	69 (4.26%)
Any co-morbidity†	730 (6.84%)	105 (6.48%)
Previous abortion	1269 (11.89%)	191 (11.79%)
Previous premature birth (<5lb,9oz)	855 (8.01%)	148 (9.14%)
Previous large birth (>8lb,14oz))	965 (9.04%)	177 (10.93%)
Previous stillbirth/neonatal death	549 (5.14%)	79 (4.88%)
Complication in previous pregnancy	904 (8.47%)	155 (9.57%)

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.
†Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy

Table A2 (continued): Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

	Not exposed to influenza *in-utero* (86.8%)	**Exposed to influenza *in-utero* *(13.2%)***
	N (%) or Mean (SD)	**N (%) or Mean (SD)**
Persons per room		
<1	7397 (69.32%)	1119 (69.07%)
1-1.5	1851 (17.35%)	275 (16.98%)
1.5-2	969 (9.08%)	148 (9.14%)
>2	454 (4.25%)	78 (4.81%)
Number of older siblings		
No older siblings	4061 (38.06%)	577 (35.62%)
1	3468 (32.50%)	514 (31.73%)
2	1575 (14.76%)	263 (16.23%)
3	791 (7.41%)	116 (7.16%)
4	360 (3.37%)	65 (4.01%)
5	182 (1.71%)	45 (2.78%)
6+	234 (2.19%)	40 (2.47%)
Stillbirth rate (Still births/total births)	0.02 (0.00)	0.02 (0.01)
Proportion of male working population unskilled	0.14 (0.04)	0.14 (0.04)
Proportion of men leaving school aged 16 or older	0.13 (0.05)	0.13 (0.05)
Population density (population in thousands/square km)	2.40 (2.45)	2.27 (2.35)
Pneumonia Notifications Per Thousand Population	12.61 (17.07)	12.00 (15.53)
Region of birth (Base: Northern England)		
North West England	765 (7.17%)	114 (7.04%)
East & West Riding, England	1437 (13.47%)	183 (11.3%)
North Midlands, England	899 (8.42%)	147 (9.07%)
Midlands, England	776 (7.27%)	138 (8.52%)
East England	975 (9.14%)	198 (12.22%)
South East England	901 (8.44%)	146 (9.01%)
South England	1913 (17.93%)	243 (15%)
South West England	587 (5.5%)	96 (5.93%)
Wales	585 (5.48%)	85 (5.25%)
Scotland	541 (5.07%)	100 (6.17%)

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.

Appendix A4: Derivation of the estimation sample for each outcome

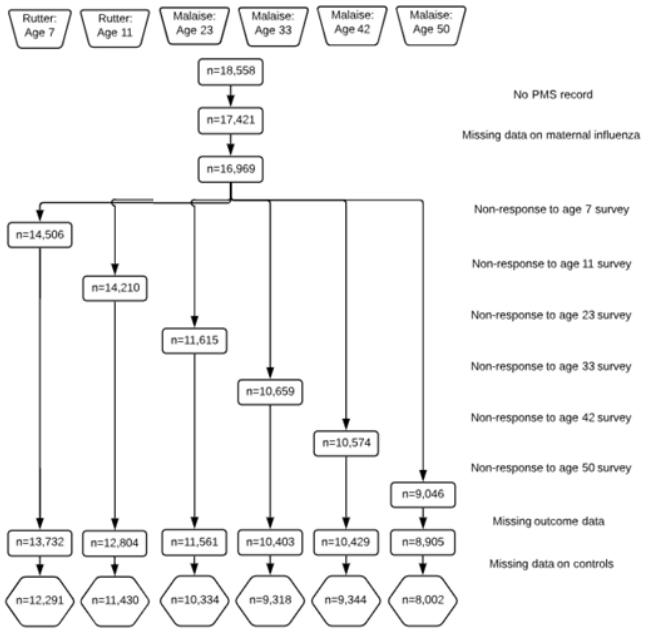


Figure A3: Derivation of the estimation sample for each outcome

Appendix A5: Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

Table A3: Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

-	Age 11		Age 23		Age 33		Age 42		Age 50	
-	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted
Male gender	-0.020	0.006	0.008	-0.001	0.015	-0.003	0.003	0.000	-0.002	-0.005
Ethnic minority	0.001	-0.011	-0.002	-0.012	-0.036	-0.013	-0.013	-0.018	-0.055	-0.009
Mother's occupational social class (Base: V: Unskilled)										
Unemployed	-0.030	0.003	-0.004	-0.003	-0.032	-0.001	-0.039	-0.002	-0.055	-0.006
I or II: Professional, managerial or technical	0.018	-0.008	0.007	-0.013	0.013	-0.004	0.021	-0.007	0.021	-0.008
III: Skilled manual/non-manual	0.000	0.000	0.001	0.005	0.029	-0.001	0.014	0.003	0.018	0.006
IV: Partly skilled	0.034	0.001	-0.005	0.008	-0.002	0.006	0.032	0.003	0.034	0.006
Father's occupational social class (Base: V: Unskilled)										
I: Professional	-0.056	0.010	-0.013	0.004	-0.024	0.012	-0.012	0.018	-0.021	0.010
II: Managerial or technical	-0.108	0.000	-0.095	-0.009	-0.092	-0.010	-0.104	-0.004	-0.112	-0.004
III: Skilled manual/non-manual	0.091	-0.003	0.074	-0.009	0.075	0.001	0.070	-0.006	0.081	-0.007
IV: Partly skilled	0.005	0.000	-0.004	0.008	0.008	-0.001	0.001	0.001	0.020	0.006
Information missing	-0.022	-0.004	-0.004	-0.001	-0.029	-0.008	-0.001	-0.004	-0.014	-0.003
Mother schooled past leaving age	-0.048	0.004	-0.003	0.004	-0.054	-0.002	-0.047	0.008	-0.049	-0.005
Mother's age in years (Base: 20 or under)										
21-25	0.078	-0.001	0.064	0.007	0.040	0.005	0.061	0.004	0.062	-0.003
26-30	-0.077	0.011	-0.082	0.003	-0.103	0.000	-0.067	0.002	-0.067	0.008
31-35	-0.020	-0.010	-0.009	-0.009	0.035	-0.010	0.003	-0.010	-0.003	-0.006
36-40	-0.009	0.008	0.017	0.001	0.014	0.000	-0.027	0.005	-0.025	0.007
Over 40	-0.036	-0.014	-0.005	-0.005	-0.014	0.000	-0.022	0.001	-0.023	-0.003
Mother's height (inches)	0.022	-0.006	0.023	-0.007	-0.002	-0.002	0.016	0.004	0.003	-0.003
Father's age in years (Base: 20 or under)										
21-25	0.071	0.000	0.063	0.000	0.045	-0.001	0.040	0.003	0.034	-0.003
26-30	0.012	-0.002	-0.001	0.003	0.003	0.005	0.015	-0.003	0.028	-0.003
31-35	-0.044	0.001	-0.044	0.003	-0.041	0.003	-0.028	0.005	-0.039	0.000
36-40	-0.030	0.000	-0.016	-0.003	-0.004	0.000	-0.011	-0.004	-0.017	0.002
41-45	0.018	0.003	0.048	-0.001	0.035	0.000	0.018	0.004	0.025	0.004
45-50	-0.030	0.000	-0.019	-0.014	-0.035	-0.014	-0.057	0.001	-0.027	0.000
Over 50	-0.039	-0.005	-0.051	-0.003	-0.060	-0.020	-0.058	-0.019	-0.050	-0.018
Information missing	-0.001	-0.001	-0.011	0.004	0.014	0.000	-0.006	0.007	-0.007	0.013
Mother's smoking prior to pregnancy (Base: Non-smoker)										
1-4 cigarettes per day	0.020	0.002	0.041	0.003	0.027	0.007	0.032	0.005	0.076	0.003
5-9 cigarettes per day	0.000	0.001	-0.011	0.002	0.003	0.005	0.006	-0.005	0.005	-0.002
10-14 cigarettes per day	-0.020	-0.004	-0.023	-0.006	0.003	-0.007	-0.005	-0.005	0.020	-0.004
15-19 cigarettes per day	-0.047	0.011	0.009	0.010	-0.013	0.008	-0.027	0.018	-0.028	0.021
20+ cigarettes per day	0.006	0.004	0.011	-0.001	0.026	-0.009	0.026	-0.013	0.043	-0.010
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])										
Underweight [BMI<18.5]	-0.073	0.011	-0.066	0.004	-0.079	0.000	-0.048	0.002	-0.053	-0.002
Overweight [25<BMI<30]	0.048	0.007	0.063	0.004	0.052	0.008	0.029	0.004	0.030	0.006
Obese [BMI>30]	0.018	-0.004	-0.004	0.001	0.012	0.008	0.010	0.005	-0.016	0.008
Antenatal visits within 7 weeks of pregnancy	0.017	0.005	0.009	0.007	0.024	0.007	0.020	0.004	0.016	0.001
Any comorbidity[a]	-0.033	0.001	-0.034	0.002	-0.020	-0.004	-0.025	-0.010	-0.022	-0.004
Previous abortion	-0.006	0.002	0.010	-0.008	0.022	-0.008	0.024	-0.010	0.018	-0.007
Previous premature birth (<5lb,9oz)	0.048	-0.003	0.051	0.002	0.053	0.006	0.045	-0.003	0.031	0.000
Previous large birth (>8lb,14oz))	0.068	-0.006	0.078	-0.003	0.086	0.001	0.053	0.000	0.083	-0.005
Previous stillbirth/neonatal death	-0.008	-0.002	0.019	0.008	0.026	-0.003	0.024	-0.003	0.019	-0.001
Complication in previous pregnancy	0.024	0.004	0.053	0.004	0.053	0.009	0.053	0.005	0.019	0.010

Formatted: Normal, Line spacing: 1.5 lines

Formatted: Font: 9 pt

Formatted Table

Formatted: Normal, Line spacing: 1.5 lines

Table A3 (continued): Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

-	Age 11		Age 23		Age 33		Age 42		Age 50	
	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted
Persons per room (Base: <1)										
1-1.5	0.003	-0.001	-0.007	-0.004	-0.020	-0.007	-0.002	-0.008	0.004	0.000
1.5-2	0.001	-0.003	0.005	-0.002	0.043	0.000	0.016	-0.005	0.045	-0.006
>2	0.018	0.001	0.051	0.003	0.024	0.007	0.057	0.001	0.079	0.005
Number of older siblings (Base: No siblings)										
1	-0.017	-0.006	-0.045	-0.002	-0.055	-0.009	-0.030	-0.006	-0.065	-0.015
2	0.025	0.012	0.061	0.008	0.053	0.007	0.066	0.005	0.025	0.015
3	0.007	0.007	0.019	0.011	0.007	-0.001	-0.005	0.000	0.016	-0.001
4	0.034	-0.004	0.049	-0.003	0.060	-0.003	0.029	0.002	0.071	0.003
5	0.057	0.002	0.077	-0.004	0.061	0.008	0.078	0.001	0.100	-0.001
6+	0.019	-0.006	0.046	-0.001	0.066	-0.003	0.033	-0.001	0.032	0.005
Stillbirth rate (Still births/total births)	0.058	0.002	0.069	0.003	0.068	0.007	0.069	0.002	0.076	-0.002
Proportion of male working population unskilled	-0.070	-0.002	-0.060	-0.006	-0.030	0.001	-0.041	-0.001	-0.021	0.001
Proportion of men leaving school aged 16 or older	-0.037	0.004	-0.049	-0.001	-0.061	-0.007	-0.046	-0.006	-0.024	-0.002
Population density (population in thousands/square km)	-0.072	-0.001	-0.054	-0.008	-0.041	-0.011	-0.042	-0.007	-0.031	-0.009
Pneumonia Notifications Per Thousand Population	-0.067	0.011	-0.060	0.009	-0.011	0.014	-0.032	0.010	-0.057	0.015
Region of birth (Base: Northern England)										
North West England	-0.062	-0.004	-0.052	-0.002	-0.066	0.002	-0.074	-0.008	-0.047	-0.005
East & West Riding, England	0.017	-0.006	0.040	-0.004	0.023	-0.003	0.042	-0.004	0.016	-0.005
North Midlands, England	0.065	-0.003	0.079	0.000	0.071	0.008	0.066	0.011	0.075	0.008
Midlands, England	0.093	0.002	0.102	0.009	0.086	0.003	0.058	0.007	0.047	0.006
East England	0.015	0.002	-0.029	0.000	0.007	0.001	0.002	0.001	0.014	0.002
South East England	-0.073	-0.001	-0.081	-0.007	-0.091	-0.020	-0.054	-0.015	-0.079	-0.013
South England	0.034	-0.009	0.030	-0.007	0.031	-0.002	0.016	-0.001	0.051	0.000
South West England	-0.005	-0.003	-0.028	-0.004	-0.027	-0.001	-0.043	0.006	-0.033	0.002
Wales	0.051	0.000	0.055	-0.004	0.055	0.003	0.061	0.002	0.054	-0.003
Scotland	-0.076	0.016	-0.082	0.019	-0.047	0.021	-0.066	0.014	-0.059	0.024

Formatted Table

Formatted: Normal, Line spacing: 1.5 lines

Appendix A65: Probit marginal effects from the propensity score model

Table A43: Marginal effects of probit models of *in-utero* influenza exposure.

	Age 7		Age 11		Age 23		Age 33		Age 42		Age 50	
	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err
Male gender	-0.0051	(0.0061)	-0.0053	(0.0063)	0.0013	(0.0066)	0.0031	(0.0069)	0.0001	(0.0069)	-0.0001	(0.0074)
Ethnic minority	-0.0340	(0.0312)	-0.0055	(0.0297)	-0.0097	(0.0339)	-0.0569	(0.0418)	-0.0299	(0.0365)	-0.0832	(0.0430)
Mother's occupational social class (Base: V: Unskilled)												
Unemployed	-0.0153	(0.0203)	-0.0088	(0.0203)	-0.0129	(0.0221)	-0.0105	(0.0237)	-0.0055	(0.0231)	-0.0274	(0.0268)
I or II: Professional, managerial or technical	0.0183	(0.0260)	0.0277	(0.0263)	0.0121	(0.0278)	0.0289	(0.0299)	0.0398	(0.0297)	0.0147	(0.0332)
III: Skilled manual/non-manual	0.0078	(0.0213)	0.0083	(0.0214)	0.0057	(0.0232)	0.0155	(0.0247)	0.0222	(0.0242)	-0.0040	(0.0277)
IV: Partly skilled	0.0118	(0.0226)	0.0164	(0.0228)	0.0001	(0.0245)	0.0032	(0.0260)	0.0244	(0.0257)	-0.0001	(0.0292)
Father's occupational social class (Base: V: Unskilled)												
I: Professional	-0.0195	(0.0185)	-0.0238	(0.0190)	0.0042	(0.0205)	0.0074	(0.0219)	0.0137	(0.0221)	0.0162	(0.0228)
II: Managerial or technical	-0.0246	(0.0137)	-0.0292*	(0.0143)	-0.0199	(0.0150)	-0.0153	(0.0159)	-0.0172	(0.0155)	-0.0085	(0.0165)
III: Skilled manual/non-manual	0.0066	(0.0112)	0.0050	(0.0118)	0.0113	(0.0124)	0.0100	(0.0132)	0.0093	(0.0128)	0.0190	(0.0136)
IV: Partly skilled	-0.0025	(0.0135)	-0.0077	(0.0141)	-0.0015	(0.0148)	0.0000	(0.0158)	-0.0026	(0.0154)	0.0106	(0.0166)
Information missing	0.0006	(0.0244)	-0.0296	(0.0228)	0.0032	(0.0263)	-0.0397	(0.0237)	0.0017	(0.0284)	-0.0052	(0.0292)
Mother schooled past leaving age	0.0022	(0.0077)	0.0004	(0.0079)	0.0131	(0.0083)	-0.0025	(0.0086)	-0.0024	(0.0086)	0.0001	(0.0091)
Mother's age in years (Base: 20 or under)												
21-25	-0.0142	(0.0137)	-0.0145	(0.0141)	-0.0118	(0.0151)	-0.0228	(0.0158)	-0.0164	(0.0159)	-0.0199	(0.0174)
26-30	-0.0385*	(0.0152)	-0.0423**	(0.0157)	-0.0400*	(0.0168)	-0.0481**	(0.0176)	-0.0428*	(0.0177)	-0.0480*	(0.0194)
31-35	-0.0341*	(0.0173)	-0.0408*	(0.0178)	-0.0357	(0.0190)	-0.0287	(0.0203)	-0.0389	(0.0200)	-0.0458*	(0.0220)
36-40	-0.0461*	(0.0199)	-0.0480*	(0.0206)	-0.0385	(0.0223)	-0.0415	(0.0235)	-0.0577*	(0.0226)	-0.0655**	(0.0244)
Over 40	-0.0728**	(0.0252)	-0.0774**	(0.0255)	-0.0548	(0.0299)	-0.0569	(0.0309)	-0.0630*	(0.0305)	-0.0822**	(0.0298)
Mother's height (inches)	0.0019	(0.0013)	0.0019	(0.0013)	0.0015	(0.0014)	0.0007	(0.0015)	0.0015	(0.0015)	0.0006	(0.0016)
Father's age in years (Base: 20 or under)												
21-25	0.0391	(0.0205)	0.0382	(0.0213)	0.0422	(0.0239)	0.0389	(0.0240)	0.0235	(0.0250)	0.0330	(0.0254)
26-30	0.0360	(0.0211)	0.0336	(0.0219)	0.0337	(0.0245)	0.0378	(0.0247)	0.0245	(0.0259)	0.0427	(0.0263)
31-35	0.0304	(0.0221)	0.0232	(0.0229)	0.0212	(0.0255)	0.0205	(0.0257)	0.0151	(0.0269)	0.0303	(0.0273)
36-40	0.0293	(0.0236)	0.0210	(0.0245)	0.0168	(0.0269)	0.0175	(0.0273)	0.0144	(0.0287)	0.0316	(0.0293)
41-45	0.0544	(0.0280)	0.0425	(0.0287)	0.0438	(0.0311)	0.0315	(0.0308)	0.0282	(0.0326)	0.0507	(0.0338)
45-50	0.0157	(0.0321)	0.0107	(0.0332)	0.0024	(0.0349)	-0.0095	(0.0339)	-0.0235	(0.0354)	0.0165	(0.0391)
Over 50	-0.0558	(0.0316)	-0.0181	(0.0394)	-0.0448	(0.0383)	-0.0564	(0.0350)	-0.0586	(0.0384)	-0.0288	(0.0429)
Information missing	0.0221	(0.0324)	0.0488	(0.0375)	0.0209	(0.0368)	0.0817	(0.0457)	0.0112	(0.0392)	0.0345	(0.0432)
Mother's smoking prior to pregnancy (Base: Non-smoker)												
1-4 cigarettes per day	0.0081	(0.0119)	0.0056	(0.0123)	0.0166	(0.0132)	0.0121	(0.0135)	0.0131	(0.0136)	0.0358*	(0.0155)
5-9 cigarettes per day	-0.0040	(0.0090)	0.0001	(0.0095)	-0.0007	(0.0100)	0.0054	(0.0106)	0.0027	(0.0104)	0.0069	(0.0113)
10-14 cigarettes per day	-0.0056	(0.0091)	-0.0062	(0.0094)	-0.0043	(0.0099)	0.0044	(0.0105)	-0.0004	(0.0104)	0.0091	(0.0113)
15-19 cigarettes per day	-0.0152	(0.0180)	-0.0280	(0.0182)	0.0129	(0.0216)	-0.0016	(0.0216)	-0.0134	(0.0206)	-0.0074	(0.0225)
20+ cigarettes per day	0.0091	(0.0200)	0.0059	(0.0204)	0.0114	(0.0221)	0.0267	(0.0236)	0.0202	(0.0227)	0.0361	(0.0255)
Observations	12,291		11,430		10,344		9,318		9,344		8,002	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects.

	Marginal effects	Std. Err
Male gender	-0.0051	(0.0061)
Ethnic minority	-0.0340	(0.0312)
Mother's occupational social class (Base: V: Unskilled)		
Unemployed	-0.0153	(0.0203)
I or II: Professional, managerial or technical	0.0183	(0.0260)
III: Skilled manual/non-manual	0.0078	(0.0213)

Formatted: Left: 0.5", Right: 0.5", Top: 0.5", Bottom: 0.5", Width: 11.69", Height: 8.27"

Formatted Table

IV: Partly skilled	0.0118	(0.0226)
Father's occupational social class (Base: V: Unskilled)		
I: Professional	-0.0195	(0.0185)
II: Managerial or technical	-0.0246	(0.0137)
III: Skilled manual/non-manual	0.0066	(0.0112)
IV: Partly skilled	-0.0025	(0.0135)
Information missing	0.0006	(0.0244)
Mother schooled past leaving age	0.0022	(0.0077)
Mother's age in years (Base: 20 or under)		
21-25	-0.0142	(0.0137)
26-30	-0.0385*	(0.0152)
31-35	-0.0341*	(0.0173)
36-40	-0.0461*	(0.0199)
Over 40	-0.0728**	(0.0252)
Mother's height (inches)	0.0019	(0.0013)
Father's age in years (Base: 20 or under)		
21-25	0.0391	(0.0205)
26-30	0.0360	(0.0211)
31-35	0.0304	(0.0221)
36-40	0.0293	(0.0236)
41-45	0.0544	(0.0280)
45-50	0.0157	(0.0321)
Over 50	-0.0558	(0.0316)
Information missing	0.0221	(0.0324)
Mother's smoking prior to pregnancy (Base: Non-smoker)		
1-4 cigarettes per day	0.0081	(0.0119)
5-9 cigarettes per day	-0.0040	(0.0090)
10-14 cigarettes per day	-0.0056	(0.0091)
15-19 cigarettes per day	-0.0152	(0.0180)
20+ cigarettes per day	0.0091	(0.0200)
Observations	12,291	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects. Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7).

Table A43 (continued): Marginal effects of probit models of *in-utero* influenza exposure

-	Marginal effects	Std. Err
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])		
Underweight [BMI<18.5]	-0.0397**	(0.0141)
Overweight [25<BMI<30]	0.0085	(0.0080)
Obese [BMI>30]	0.0206	(0.0161)
Antenatal visits within 7 weeks of pregnancy	0.0122	(0.0152)
Any comorbidity†	-0.0036	(0.0122)
Previous abortion	0.0015	(0.0096)
Previous premature birth (<5lb,9oz)	0.0117	(0.0119)
Previous large birth (>8lb,14oz))	0.0127	(0.0110)
Previous stillbirth/neonatal death	-0.0129	(0.0145)
Complication in previous pregnancy	0.0125	(0.0110)
Persons per room (Base: <1)		
1-1.5	-0.0166	(0.0087)
1.5-2	-0.0136	(0.0116)
>2	0.0003	(0.0167)
Number of older siblings (Base: No siblings)		
1	0.0222*	(0.0086)
2	0.0462***	(0.0120)
3	0.0315*	(0.0153)
4	0.0651**	(0.0228)
5	0.1173***	(0.0324)
6+	0.0577*	(0.0286)
Stillbirth rate (Still births/total births)	1.6088*	(0.7599)
Proportion of male working population unskilled	-0.0310	(0.1431)
Proportion of men leaving school aged 16 or older	-0.0665	(0.1066)
Population density (population in thousands/square km)	-0.0008	(0.0019)
Pre-pregnancy pneumonia notifications per thousand population	-0.0001	(0.0003)
Region of birth (Base: Northern England)		
North West England	-0.0165	(0.0141)
East & West Riding, England	0.0131	(0.0155)
North Midlands, England	0.0170	(0.0162)
Midlands, England	0.0376*	(0.0156)
East England	0.0178	(0.0170)
South East England	-0.0026	(0.0166)
South England	0.0192	(0.0195)
South West England	0.0050	(0.0183)
Wales	0.0283	(0.0188)
Scotland	-0.0041	(0.0156)
Observations	12,291	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects.
†Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy
Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7).

-	Age 7		Age 11		Age 23		Age 33		Age 42		Age 50	
-	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])												
Underweight [BMI<18.5]	-0.0397**	(0.0141)	-0.0400**	(0.0144)	-0.0367*	(0.0160)	-0.0415*	(0.0162)	-0.0284	(0.0168)	-0.0308	(0.0178)
Overweight [25<BMI<30]	0.0085	(0.0080)	0.0139	(0.0084)	0.0176*	(0.0089)	0.0132	(0.0093)	0.0085	(0.0092)	0.0066	(0.0098)
Obese [BMI>30]	0.0206	(0.0161)	0.0088	(0.0162)	-0.0024	(0.0165)	0.0027	(0.0179)	0.0036	(0.0178)	-0.0117	(0.0184)
Antenatal visits within 7 weeks of pregnancy	0.0122	(0.0152)	0.0152	(0.0154)	0.0119	(0.0166)	0.0199	(0.0170)	0.0181	(0.0170)	0.0164	(0.0183)
Any comorbidity†	-0.0036	(0.0122)	-0.0103	(0.0127)	-0.0126	(0.0135)	-0.0087	(0.0140)	-0.0110	(0.0140)	-0.0103	(0.0151)
Previous abortion	0.0015	(0.0096)	0.0015	(0.0099)	0.0031	(0.0103)	0.0075	(0.0107)	0.0087	(0.0107)	0.0068	(0.0115)
Previous premature birth (<5lb,9oz)	0.0117	(0.0119)	0.0169	(0.0124)	0.0070	(0.0130)	0.0071	(0.0137)	0.0043	(0.0136)	0.0005	(0.0149)
Previous large birth (>8lb,14oz))	0.0127	(0.0110)	0.0168	(0.0115)	0.0111	(0.0122)	0.0164	(0.0126)	0.0059	(0.0128)	0.0191	(0.0135)

Formatted Table

Previous stillbirth/neonatal death	-0.0129	(0.0145)	-0.0073	(0.0153)	0.0049	(0.0157)	0.0058	(0.0162)	0.0074	(0.0165)	0.0070	(0.0178)
Complication in previous pregnancy	0.0125	(0.0110)	0.0059	(0.0115)	0.0154	(0.0120)	0.0178	(0.0128)	0.0175	(0.0126)	0.0066	(0.0140)
Persons per room (Base: <1)												
1-1.5	-0.0166	(0.0087)	-0.0115	(0.0091)	-0.0182	(0.0094)	-0.0217*	(0.0098)	-0.0146	(0.0099)	-0.0103	(0.0107)
1.5-2	-0.0136	(0.0116)	-0.0115	(0.0121)	-0.0136	(0.0129)	-0.0061	(0.0140)	-0.0079	(0.0138)	0.0058	(0.0157)
>2	0.0003	(0.0167)	-0.0012	(0.0171)	0.0172	(0.0201)	-0.0070	(0.0193)	0.0199	(0.0208)	0.0375	(0.0234)
Number of older siblings (Base: No siblings)												
1	0.0222*	(0.0086)	0.0211*	(0.0090)	0.0189*	(0.0093)	0.0170	(0.0097)	0.0250*	(0.0098)	0.0127	(0.0104)
2	0.0462***	(0.0120)	0.0415***	(0.0124)	0.0573***	(0.0133)	0.0498***	(0.0137)	0.0599***	(0.0139)	0.0401**	(0.0145)
3	0.0315*	(0.0153)	0.0380*	(0.0162)	0.0502**	(0.0176)	0.0401*	(0.0177)	0.0423*	(0.0176)	0.0409*	(0.0195)
4	0.0651**	(0.0228)	0.0655**	(0.0238)	0.0852***	(0.0257)	0.0890**	(0.0279)	0.0777**	(0.0276)	0.0959**	(0.0311)
5	0.1173***	(0.0324)	0.1003**	(0.0323)	0.1236***	(0.0348)	0.1133**	(0.0389)	0.1386***	(0.0384)	0.1401***	(0.0422)
6+	0.0577*	(0.0286)	0.0565	(0.0297)	0.0841*	(0.0340)	0.0995**	(0.0375)	0.0894*	(0.0384)	0.0646	(0.0393)
Area-level characteristics												
Stillbirth rate (Still births/total births)	1.6088*	(0.7599)	1.4399	(0.8179)	1.2514	(0.9054)	1.4086	(0.9399)	1.6924	(0.9256)	2.0726	(1.0576)
Proportion of male working population unskilled	-0.0310	(0.1431)	-0.0506	(0.1523)	-0.0662	(0.1573)	0.0436	(0.1665)	-0.0174	(0.1643)	0.1253	(0.1746)
Proportion of men leaving school aged 16 or older	-0.0665	(0.1066)	-0.0675	(0.1120)	-0.0273	(0.1147)	-0.0286	(0.1215)	-0.0580	(0.1182)	0.1214	(0.1255)
Population density (population in thousands/square km)	-0.0008	(0.0019)	-0.0012	(0.0020)	-0.0004	(0.0021)	-0.0011	(0.0022)	-0.0006	(0.0022)	-0.0001	(0.0023)
Pneumonia Notifications Per Thousand Population	-0.0001	(0.0003)	-0.0002	(0.0003)	-0.0003	(0.0003)	0.0001	(0.0003)	-0.0001	(0.0003)	-0.0005	(0.0004)
Region of birth (Base: Northern England)												
North West England	-0.0165	(0.0141)	-0.0143	(0.0147)	-0.0173	(0.0161)	-0.0213	(0.0165)	-0.0340*	(0.0167)	-0.0251	(0.0180)
East & West Riding, England	0.0131	(0.0155)	0.0111	(0.0160)	0.0147	(0.0176)	0.0083	(0.0178)	0.0067	(0.0184)	0.0085	(0.0201)
North Midlands, England	0.0170	(0.0162)	0.0258	(0.0171)	0.0247	(0.0182)	0.0279	(0.0192)	0.0128	(0.0192)	0.0288	(0.0209)
Midlands, England	0.0376*	(0.0156)	0.0358*	(0.0161)	0.0338	(0.0175)	0.0318	(0.0181)	0.0092	(0.0182)	0.0161	(0.0197)
East England	0.0178	(0.0170)	0.0159	(0.0174)	-0.0097	(0.0183)	0.0102	(0.0197)	-0.0005	(0.0198)	0.0061	(0.0211)
South East England	-0.0026	(0.0166)	0.0005	(0.0171)	-0.0149	(0.0185)	-0.0093	(0.0194)	-0.0090	(0.0200)	-0.0191	(0.0207)
South England	0.0192	(0.0195)	0.0277	(0.0205)	0.0139	(0.0216)	0.0250	(0.0227)	0.0070	(0.0230)	0.0219	(0.0248)
South West England	0.0050	(0.0183)	0.0070	(0.0187)	-0.0127	(0.0197)	-0.0062	(0.0205)	-0.0201	(0.0207)	-0.0143	(0.0222)
Wales	0.0283	(0.0188)	0.0275	(0.0190)	0.0232	(0.0205)	0.0262	(0.0213)	0.0196	(0.0217)	0.0152	(0.0230)
Scotland	-0.0041	(0.0156)	-0.0096	(0.0158)	-0.0163	(0.0173)	-0.0149	(0.0179)	-0.0274	(0.0178)	-0.0097	(0.0199)
Observations	12,291		11,430		10,344		9,318		9,344		8,002	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects.

[a]Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy

Appendix A76: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of mortality in childhood and adulthood by gender

Table A54: Gender-specific inverse probability weight estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50.

	Full sample		Males		Females	
	Observations	**ATE**	**Observations**	**ATE**	**Observations**	**ATE**
Rutter: age 7						
Overall	12,291	0.103 (0.095) [1.63%]	6,308	0.183 (0.130) [2.79%]	5.843	0.026 (0.137) [0.43%]
Internalising	12,291	0.076 (0.060) [2.29%]	6,273	0.124 (0.082) [3.83%]	5.983	0.038 (0.088) [1.12%]
Externalising	12,291	0.027 (0.058) [0.90%]	6,301	0.066 (0.083) [2.00%]	5.957	-0.019 (0.079) [-0.72%]
Rutter: age 11						
Overall	11,430	0.194* (0.098) [3.04%]	5,854	0.163 (0.133) [2.44%]	5.976	0.220 (0.141) [3.66%]
Internalising	11,430	0.131* (0.060) [3.68%]	5,806	0.012 (0.085) [0.34%]	5.576	0.275** (0.086) [7.72%]
Externalising	11,430	0.061 (0.058) [2.17%]	5,843	0.143 (0.080) [4.57%]	5.527	-0.050 (0.082) [-2.05%]
Malaise: age 23	10,344	0.054 (0.081) [1.99%]	5,149	0.043 (0.097) [2.14%]	5,195	0.065 (0.131) [1.93%]
Malaise: age 33	9,318	0.082 (0.089) [3.42%]	4,585	0.003 (0.112) [0.15%]	4,733	0.131 (0.132) [4.79%]
Malaise: age 42	9,344	0.046 (0.108) [1.39%]	4,573	0.013 (0.145) [0.42%]	4,771	0.093 (0.160) [2.32%]
Malaise: age 50	8,002	0.093 (0.066) [6.35%]	3,919	0.176* (0.089) [15.45%]	4,083	0.052 (0.101) [2.98%]

Robust standard errors in parenthesis rounded parenthesis; ATEs as a % of mean outcomes for the non-exposed group in squared parenthesis;
*p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

Appendix A~~8~~7: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of mortality in childhood and adulthood

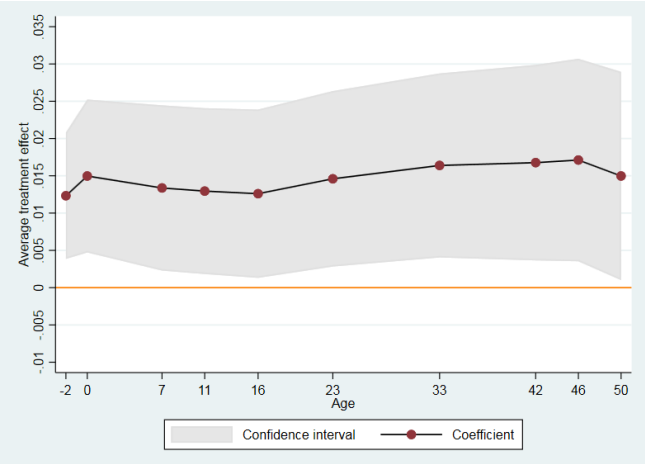


Figure A4: Inverse-probability-weight estimates of the ATE of *in-utero* exposure to influenza on the probability of death at birth, 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50. Ages -2 and 0 represent birth and age 28 days, respectively.

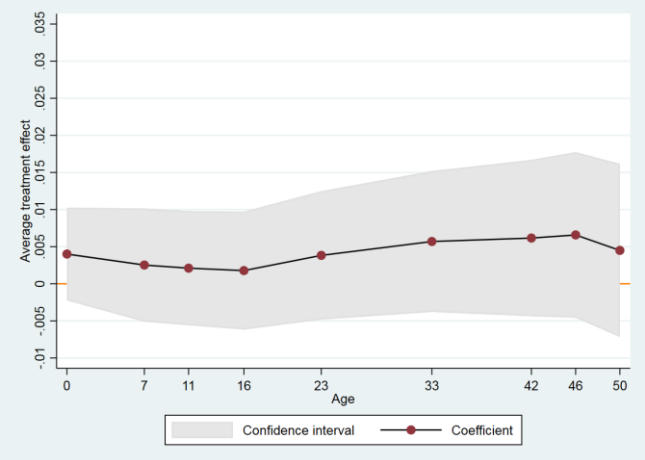


Figure A5: Inverse-probability-weight estimates of the ATE of *in-utero* exposure to influenza on the probability of death at 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50 after conditioning on survival beyond birth. Age 0 represents birth 28 days..

Appendix A98: ATE bounds under mortality-driven non-response

Table A65: ATE bounds under mortality-driven non-response

	Worst case bounds		Bounds replacing missing outcomes with 90th percentile	
	ATE	**ATE as % of non-exposed mean**	**ATE**	**ATE as % of non-exposed mean**
Rutter: age 7				
Overall	[0.103, 0.445]	[1.63%, 6.13%]	[0.103, 0.176]	[1.63%, 2.7%]
Internalising	[0.076, 0.284]	[2.29%, 7.28%]	[0.076, 0.118]	[2.29%, 3.41%]
Externalising	[0.027, 0.182]	[0.9%, 5.35%]	[0.027, 0.080]	[0.9%, 2.56%]
Rutter: age 11				
Overall	[0.194, 0.503]	[3.04%, 6.9%]	[0.194, 0.262]	[3.04%, 3.98%]
Internalising	[0.131, 0.348]	[3.68%, 8.37%]	[0.131, 0.192]	[3.68%, 5.13%]
Externalising	[0.061, 0.232]	[2.17%, 7.08%]	[0.061, 0.120]	[2.17%, 4.05%]
Malaise: age 23	[0.054, 0.384]	[1.99%, 10.01%]	[0.054, 0.128]	[1.99%, 4.33%]
Malaise: age 33	[0.082, 0.563]	[3.42%, 14.64%]	[0.082, 0.166]	[3.42%, 6.28%]
Malaise: age 42	[0.046, 0.563]	[1.29%, 10.6%]	[0.046, 0.180]	[1.29%, 4.48%]
Malaise: age 50	[0.093, 0.285]	[6.35%, 12.06%]	[0.093, 0.148]	[6.35%, 8.42%]

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.

Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50. Upper bound or worst case bounds calculated by replacing missing outcome values due to death with the maximum of the outcome.

Appendix A109: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of non-response in each NCDS follow-up

Table A76: Inverse-probability-weighted estimates of the ATE of *in-utero* exposure to influenza on the probability of non-response at age 7, 11, 23, 33, 42, and age 50 follow-ups.

	Observations	Potential outcome mean under no exposure	ATE
Age 7	14,413	0.853*** (0.003)	-0.003 (0.009)
Age 11	14,398	0.793*** (0.004)	0.001 (0.010)
Age 23	14,305	0.722*** (0.004)	0.008 (0.011)
Age 33	14,233	0.655*** (0.004)	-0.006 (0.012)
Age 42	14,110	0.663*** (0.004)	-0.010 (0.012)
Age 50	13,893	0.578*** (0.004)	-0.019 (0.013)

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Estimation sample reductions due to increases in the number of deaths.

Appendix A10: Unadjusted and IPTW estimates of the ATE of *in-utero* exposure to influenza on mental health and mortality outcomes

Table A7: Unadjusted and inverse-probability-weighted estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50 and the probability of death at birth, 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50.

	Observations	ATE	
		Unadjusted	IPTW
Rutter: age 7			
Overall	12,291	0.103 (0.098)	0.103 (0.095)
Internalising	12,291	0.080 (0.061)	0.076 (0.060)
Externalising	12,291	0.026 (0.060)	0.027 (0.058)
Rutter: age 11			
Overall	11,430	0.189 (0.099)	0.194* (0.098)
Internalising	11,430	0.127* (0.061)	0.131* (0.060)
Externalising	11,430	0.060 (0.060)	0.061 (0.058)
Malaise in adulthood			
Age 23	10,344	0.142 (0.085)	0.054 (0.081)
Age 33	9,318	0.156 (0.091)	0.082 (0.089)
Age 42	9,344	0.113 (0.110)	0.046 (0.108)
Age 50	8,002	0.133* (0.067)	0.093 (0.066)
Mortality			
Birth	15,019	0.010** (0.004)	0.012** (0.004)
28 days	15,019	0.012* (0.005)	0.015** (0.005)
Age 7	15,019	0.011* (0.005)	0.013* (0.006)
Age 11	15,019	0.010* (0.005)	0.013* (0.006)
Age 16	15,019	0.010 (0.005)	0.013* (0.006)
Age 23	15,019	0.012* (0.006)	0.015* (0.006)
Age 33	15,019	0.013* (0.006)	0.016** (0.006)
Age 42	15,019	0.013* (0.006)	0.017* (0.007)
Age 46	15,019	0.014* (0.006)	0.017* (0.007)
Age 50	15,019	0.012 (0.007)	0.015* (0.007)

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

Formatted: Indent: Left: 0", First line: 0"

Highlights

- This is the study to examine the effects of *in-utero* exposure to influenza at the individual level.
- Uses unique mother-reported data on influenza infection during pregnancy and data on mental health and mortality measured throughout the lifecourse.
- Influenza exposure is associated with small reductions in mental health in mid-childhood, and increases in depressive symptoms in mid-life for males.
- Influenza exposure is also associated with substantial increases in mortality, driven primarily driven by a 75% increase in the probability of being stillborn.

Title: The effects *in-utero* exposure to influenza on mental health and mortality risk throughout the life-course.

Authors and affiliations:

Alex J. Turner[a*], Eleonora Fichera[b], Matt Sutton[a]

[a]Health Organisation, Policy and Economics (HOPE) group, Centre for Primary Care & Health Services Research, The University of Manchester, Manchester, United Kingdom, M13 9PL.

[b]Department of Economics, University of Bath, Bath, United Kingdom, BA2 7JP

[*]Corresponding author

Email addresses & phone numbers:

Alex J. Turner: alexander.turner@manchester.ac.uk; +44 (0)161 275 1139

Eleonora Fichera: ef404@bath.ac.uk

Matt Sutton: matt.sutton@manchester.ac.uk

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interest statement: Declarations of interest: none

Original publication statement: This is original unpublished work and has not being submitted for publication elsewhere.

Ethics: Ethical approval was not required for this study.

Abstract:

Studies examining the later-life health consequences of *in-utero* exposure to influenza have typically estimated effects on physical health conditions, with little evidence of effects on mental health outcomes or mortality. Previous studies have also relied primarily on reduced-form estimates of the effects of exposure to influenza pandemics, meaning they are unlikely to recover effects of influenza exposure at an individual-level. This paper uses inverse probability of treatment weighting and "doubly-robust" methods alongside rare mother-reported data on *in-utero* influenza exposure to estimate the individual-level effect of *in-utero* influenza exposure on mental health and mortality risk throughout childhood and adulthood. We find that *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms, and increases in depressive symptoms in mid-life for males. There is also evidence that *in-utero* influenza exposure is associated with substantial increases in mortality, although these effects are primarily driven by a 75% increase in the probability of being stillborn, with limited evidence of additional survival disadvantages at later ages. The potential for mortality selection implies that estimated effects on mental health outcomes are likely to represent a lower bound.

Key words: Influenza; Prenatal exposure; Mental Health, Mortality

1. Introduction

Mental health disorders account for 13 percent of the global disease burden (Collins et al., 2011) and are the second leading cause of years lived under disability worldwide (Ferrari et al., 2017). It is also estimated that mental health problems are associated with a 4% reduction in GDP across 28 European Union countries, driven by increases in direct spending on healthcare, greater spending on social security programmes, and lower rates of employment and in-work productivity (OECD, 2018). Approximately half of mental disorders in adults begin before the mid-teenage years (Kessler et al., 2007; World Health Organization, 2013). Given this, a well-documented upward trend in child and adolescent mental health problems is worrying (Collishaw, 2015; McManus et al., 2019; Sadler et al., 2018). Understanding the determinants of mental health problems and multiple stages of the life-course is therefore vital.

A large body of literature examines how health shocks occurring in the prenatal (*in-utero*) period may impact the development of disease in later life (Almond et al., 2018; Currie and Almond, 2011). However, this literature has focused almost solely on physical health problems, with limited, although growing, evidence on how *in-utero* shocks may impact mental health (Currie, 2020). Studies assessing effects on mental health have typically identified mental health problems using mental health-related hospital admissions (Brown et al., 2000) or drug prescriptions for mental disorders (Persson and Rossin-Slater, 2018), reducing their ability to detect effects on milder mental health symptoms and to capture problems faced by individuals not seeking treatment or not having access to healthcare.

Stemming from Almond (2006)'s seminal work, exposure to influenza has emerged as one of the most widely-examined markers of the *in-utero* environment. Influenza is highly prevalent, infecting 5% - 20% of the population in the seasonal influenza season each year, and has significant effects on both mortality and morbidity (Cassini et al., 2018; Iuliano et al., 2018), particularly for pregnant women (Campion et al., 2014; Jamieson et al., 2006). A large number of studies have established effects of exposure to influenza *in-utero* on the prevalence of diabetes (Almond and Mazumder, 2005; Garthwaite, 2008; Lin and Liu, 2014), cardiovascular diseases (Garthwaite, 2008; Mazumder et al., 2010), metabolic and kidney diseases (Garthwaite, 2008; Lin and Liu, 2014), respiratory problems (Lin and Liu, 2014), and the incidence of strokes (Almond and Mazumder, 2005).[1] However, there is currently no evidence on how influenza exposure may impact the development of mental health problems. This is surprising given several hypothesised mechanisms linking the *in-utero* environment to these outcomes (Schlotz and Phillips, 2009). This includes hypothermia and inflammation which may directly damage the foetal brain during its early stages of development (Rasmussen et al., 2008) and indirect damage to the brain through under-nutrition (Colombo et al., 2004; Gale et al., 2008; Parsons et al., 2008), driven by influenza suppressing appetite, interfering with the absorption of nutrients such as fats and proteins, and through fever-induced increases in the rates of nutrient loss and energy consumption (Tomkins et al., 1994).

Previous studies are also limited by being unable to examine the impacts of *in-utero* exposure to influenza at an individual level. Long-run effects on offspring have primarily been examined by studying the impacts of influenza pandemics, making use of temporal and/or geographical variation in pandemic intensity under the assumption that influenza cases during pandemics are randomly assigned. However, a drawback of these empirical strategies is that *in-utero* exposure to an influenza pandemic, or exposure to a high-virulence area, is not equivalent to *in-utero* exposure

[1] A large number of studies also demonstrate an effect of *in-utero* influenza exposure on labour market and human capital outcomes, such as childhood cognition (Kelly, 2011), educational attainment (Almond, 2006; Garthwaite, 2008; Lin and Liu, 2014; Neelsen and Stratmann, 2012; Nelson, 2010; Percoco, 2016), literacy (Nelson, 2010), income (Almond, 2006; Nelson, 2010) and employment (Nelson, 2010), and weaker effects on family formation (Fletcher, 2018).

to influenza itself. Even in areas where the virulence of a pandemic is high, only a fraction of pregnant women will contract influenza. Also, given that influenza is still a prevalent disease even in non-pandemic periods, some pregnant mothers will contract influenza even in areas where the severity of the pandemic is low or even non-existent. As a result, reduced-form effects of pandemic exposure will not reflect consistent estimates of the average causal effect of *in-utero* exposure to influenza on later-life outcomes.

Schwandt (2017) uses individual-level data in a sibling fixed-effects design to study the effects of seasonal influenza on health at birth and labour market outcomes. However, influenza cases are identified using admissions to hospital for influenza-like illnesses, meaning cases which aren't severe enough to warrant hospital attendance are not recorded, and so mothers with these milder symptoms will be incorrectly identified as not contracting influenza. Estimates may therefore not test the effects of *in-utero* exposure to influenza versus no exposure, but the effect of being exposed to particularly severe cases of influenza compared to not contracting influenza or contracting a form of influenza with milder symptoms.

This study addresses these limitations using data from the 1958 National Child Development Study (NCDS); a rare dataset which collects longitudinal information on mental health symptoms of a cohort of individuals in both childhood and adulthood, and also includes mother-reported information on a range of diseases contracted during pregnancy. As data on influenza exposure is self-reported and mental health outcomes are collected via validated scales rather than recorded diagnoses, milder cases of influenza and mental health problems can be detected. Importantly, the NCDS follows all children in Great Britain born from 2nd to 9th March 1958, meaning that mothers of the NCDS cohort were pregnant during the end of the 1957 Asian Flu epidemic, which reached its peak around June when 95% of the NCDS cohort were between 16 and 25 weeks into gestation.[2] Throughout the 1957 epidemic in the UK, there was evidence that a large proportion of influenza cases were concentrated amongst young adults, and therefore not necessarily concentrated among the old and less-healthy (Clarke et al., 1958; Ministry of Health, 1960; Vynnycky and Edmunds, 2008; Woodall et al., 1958). This has the potential to aid identification of the effects of *in-utero* influenza exposure estimated through a simple comparison of outcomes for exposed and non-exposed groups, as exposure is less likely to be correlated with (unobserved) markers of maternal health. We find evidence of this in our data, with mothers reporting having contracted and not contracting influenza during pregnancy being similar on markers of maternal health and socio-demographic characteristics.

However, to adjust for small remaining differences in characteristics of those exposed and not-exposed to influenza, we estimate the effects of *in-utero* exposure to influenza using inverse propensity of treatment weighting and "doubly-robust" inverse-probability-weighted regression adjustment methods to ensure mothers contracting influenza in pregnancy are identical in observed characteristics to those not contracting influenza.

In a further contribution, we use linked data on NCDS and administrative death records to examine whether *in-utero* exposure to influenza impacts mortality risk at different stages of the lifecourse. This allows an assessment of whether effects on mental health are likely to be affected by mortality selection, but also addresses the paucity of studies examining the later-life mortality effects of *in-utero* influenza. Fletcher (2019) finds that exposure to the 1918 Spanish Flu did not lead to increases in overall or cause-specific mortality, but uses only death information for individuals with data on education and labour market outcomes in adulthood, meaning impacts on mortality risk in early-life and how this risk changes as individuals age could not be examined.

[2] See Kelly (2011) for further details on the 1957 Asian Flu epidemic in Great Britain.

Our results indicate that *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms of mental health, and increases in depressive symptoms in mid-life for males. Findings are robust to the methods used to adjust for observed covariates, and tests indicate they are likely to be unaffected by attrition bias. We also demonstrate that, consistent with small differences in observed child and family background characteristics between exposed and non-exposed groups, estimates are stable to the removal of these characteristics from the covariate set, providing evidence that omitted variable bias is unlikely to be impacting results greatly. However, we find that *in-utero* influenza is associated with a significant reduction in survival rates, primarily driven by effects on the rate of stillbirths, suggesting effects on mental health may be impacted by survival bias. Estimating bounds for treatment effects under survival bias, we find that estimated mental health effects are likely to be conservative.

The remainder of this article is organised as follows. The data is described in Section 2. Section 3 outlines the empirical strategy. The main findings are presented in Section 4. Robustness checks are presented in Section 5. Section 6 is devoted to the discussion and concluding remarks.

2. Data

2.1. National Child Development Study (NCDS)

Our primary data source is the National Child Development Study (NCDS). The NCDS is a longitudinal study which follows a cohort of approximately 17,000 individuals born in England, Scotland and Wales between the 2nd and 9th of March, 1958. The NCDS began with the Perinatal Mortality Survey (PMS), which collected data via questionnaires completed by mothers and midwifes in attendance at delivery, on factors associated with stillbirth and infant death. Population coverage was high, with the PMS accounting for 98% of all live and still births occurring in that week (Butler and Bonham, 1963). There currently exists data from nine subsequent interviews which took place at ages 7 (in 1965), 11 (1969), 16 (1974), 23 (1981), 33 (1991), 42 (1999/2000), 46 (2004/05), 50 (2008/09) and 55 (2013/14), which gathered information on a range of health, education, social and economic outcomes. For interviews during childhood ages, the NCDS also traced immigrants born in reference week and added them to the sample (Power and Elliott, 2006). Data at follow-ups was drawn from a variety of sources including the cohort members themselves, their parents and partners, local authority medical officers, and schools. 17,415 individuals completed the PMS, with the number completing falling as the cohort aged.

We also use information on deaths of NCDS cohort members identified from the National Health Service Central Register (NHSCR), NCDS death cards, NCDS interviewers, and relatives and friends of cohort members (Johnson and Brown, 2015), as well as secure access data on the exact month and year of death (Institute of Education, 2015a).

2.2. *In-utero* influenza exposure

The key feature of the NCDS is that mothers were asked whether they experienced a range of health conditions during pregnancy. In the PMS, NCDS mothers were asked "Were any of the following abnormalities or illnesses, or any other condition, encountered in pregnancy?" of which influenza was one of the conditions listed. We construct

a binary variable, flu, which is equal to one if the mother of an NCDS cohort member reports having contracted influenza during pregnancy, and zero otherwise.

2.3. Outcomes

We measure childhood mental health at ages 7 and 11 using indicators from the Rutter Behaviour Child Scale A. The scale has good psychometric properties, with estimates of inter-rater reliability of r=0.64 and re-test reliability of r=0.74 (Rutter et al., 1970). When cohort members were aged 7 and 11, parents of each cohort member were asked to state the frequency that they observed their child engaging in range of negative behaviours on a scale of "Never", "Sometimes", "Frequently". The complete scale includes 31 behaviours, but only 14 are used in the NCDS[3] (Centre for Longitudinal Studies, 2012). Responses for each behaviour are recorded as 0, 1 and 2 respectively, and are summed to create an overall index ranging from 0 (perfect mental health) to 28 (worst mental health). Approximately 6% and 10% of parents complete only a subsection of the Rutter items at the age 7 and 11 follow-ups, respectively (Table A1, Appendix A1). Where parental responses are missing for four or more of the 14 behaviours (<0.3%), the index is set to missing. To limit sample loss, for individuals with missing responses for less than four behaviours, missing values for these behaviours were imputed with the average response to the non-missing behaviours for that same individual[4].

To examine effects on different aspects of mental health, we also construct separate Rutter indices for externalising and internalising symptoms. Externalising symptoms relate to an inability to regulate behaviour, with symptoms similar to those associated with attention deficit hyperactivity disorder (ADHD). Internalising symptoms relate to problems regulating emotions and mood, with symptoms usually indicative of disorders relating to depression and anxiety (Klein et al., 2009). Rutter items were assigned as externalising/internalising using categorisations derived using principal components analysis in Blanden et al., (2007) and Klein et al., (2009).

Adult mental health is measured using Rutter's Malaise Inventory (Rutter et al., 1970) at ages 23, 33, 42, and 50. The inventory asks cohort members to self-report the presence of a variety of symptoms of anxiety and depression, through questions such as "Do you often feel depressed?', "'Do you feel tired most of the time?', "Are you easily upset or irritated?', and "Do you often have bad headaches?". The NCDS reports the 24-item measure at ages 23, 33 and 42, and the shorter 9-item version at age 50. Yes/No responses to each item are coded as 1/0, and summed to construct a count of depressive symptoms, consistent with evidence that the inventory measures a single underlying factor of distress (Hirst, 1983; Hirst and Bradshaw, 1983; McGee et al., 1986). Due to low levels of item-specific missing data for adults[5], no imputation was conducted for this measure.

The distributions of each of the mental health outcomes in childhood and adulthood are shown in Figure A1 and Figure A2 (Appendix A2).

---

[3] In the NCDS, parents are asked whether their child: "Is squirmy or fidgety", "Destroys own or others' belongings (e.g. tears or breaks)", "Fights with other children", "Worries about many things", "Prefers to do things on his/her own rather than with others", "Is irritable, often flies off the handle", "Is miserable or tearful", "Has twitches or mannerisms of the face, eyes or body", "Sucks thumb or finger during the day", "Bites nails", "Is disobedient at home", "Has difficulty in settling to anything for more than a few moments", "Is upset by new situation, by things happening for first time", "Is bullied by other children".

[4] Being exposed to influenza in-utero was also not a significant predictor of the Rutter index being partially imputed, when all other covariates were controlled for, giving us confidence that the imputation method is unlikely to affect results substantially.

[5] At ages 23, 33, 42 and 50, responses are missing for a subset of Malaise symptoms for only 1.5%, 1.3%, 0.1%, and 0.2% of observations, respectively, and so imputation of the type implemented for the Rutter index would not result in a large increase in estimation samples.

2.4. Covariates

Child and family background characteristics are derived primarily from the PMS. Gender and ethnicity are measured using binary indicators for being male and an ethnicity minority, respectively[6]. Parental socioeconomic status is measured using the social class of the cohort members' father and mother as well as mother's education. Social class is measured in five categories based on parental occupation prior to pregnancy: Class I (professional occupations), Class II (managerial and lower professional occupations), Class III (manual or non-manual skilled occupations), Class IV (semi-skilled occupations), and Class V (unskilled occupations). For mother's social class, classes I and II are not separable and so are combined, and an additional category is included for the 57% of NCDS mothers who are not employed and therefore cannot be assigned to a social class based on their occupation[7]. Maternal education is measured using a binary variable equalling one if the mother was educated past the minimum school leaving age of 15.

Maternal and paternal age at birth is measured using categorical variables with 5-year age intervals to allow for non-linearity in its effects. For mother's age, these range between age 20 and under to age 40+, and for father's age, between age 20 and under and age 50+[8]. Mother's height is measured in inches. Where information on paternal variables are missing due to an absent father-figure, we include a separate category for missing information for all father variables (father's age and father's social class)[9].

Maternal health investment is measured using maternal smoking and weight prior to pregnancy, and engagement with maternity services. Maternal smoking is included as a categorical variable ranging from non-smoker to 20+ cigarettes per day. Mothers' weight is measured using their body mass index (BMI), and were categorised as being a normal/healthy weight, underweight, overweight or obese, using validated cut-offs (Centers for Disease Control and Prevention, 2015)[10]. An indicator for whether the mother visited a neo-natal unit in the first six weeks of pregnancy is used to willingness to engage with health services[11].

To measure maternal health, we construct a binary co-morbidity indicator which equals one if a mother reported suffering from diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy. As a further proxy of a mother's health, as well as the health of their offspring, we measure NCDS mother's experience in previous pregnancies using separate binary indicators for previous abortions, premature births, large births, stillbirth/neo-natal deaths, and birth-related complications. This is under the assumption the negative birth outcomes in pervious pregnancies is indicative of poor genetic health endowments.

---

[6] The low prevalence of ethnic minorities in this historical sample prohibited the use of a more granular measure of ethnicity. This measure was derived from secure access data on the deaths of NCDS cohort members as self-reported ethnicity was not provided until the age 11 follow-up and so would be missing for cohort member dying prior to this follow-up.

[7] This is consistent with the low female labour force participation rate prevalent in the UK throughout the 1950s (Sprahue, 1988).

[8] A missing value was assigned to maternal age for a cohort member whose mother was recorded as age 8.

[9] We do this because we believe dropping observations with missing data relating to these variables could result in a selected sample. The alternative would be dropping these variables as controls. However, we feel this will likely generate omitted variable bias. For example, given that only 40% of NCDS mothers report working, mother's social class is unlikely to be representative of the socioeconomic status of the family, and so the inclusion of father's social class is required to ensure effects of socioeconomic status are adequately captured.

[10] In the PMS, NCDS mothers report weight in categories of <7 stone, 7-8 stone, 8-9 stone, 9-10 stone, 10-11 stone, 11-12 stone, 12-13 stone, 13-14 stone, 14-15 stone, and >15 stone. To construct BMI, lower and upper limits are taken for the <7 stone and >15 categories, respectively, and mid-points are taken for all other categories. These are then converted to pounds, before being divided by the square of mother's height and multiplied by 703.

[11] This was preferred to a measure of total number of visits as this is more likely increased by contacting influenza. This visit is highly correlated with the total number of visits.

Housing density and the number of older siblings are included, due to evidence that they are key determinants of influenza spread (Woodall et al., 1958). Housing density is measured using a categorical variable describing the average number of persons per room in the cohort member's place of residence. The number of older siblings is first derived from self-reported sibling data at age 16. Given this generates missing data for those who left the NCDS or die prior to this age, we replace missing values with a proxy which combines PMS data on parity (the number of previous pregnancies an NCDS mother has experienced prior to the birth of the NCDS cohort member), and the number of previous still births and neo-natal deaths. The number of live older siblings is proxied as the difference between these measures[12].

We supplement these individual-level characteristics with area-level measures of population health which may determine disease spread and proxy maternal health at an individual-level. Area-level variables were matched to NCDS cohort members via identifiers of their local authority of birth[13], which were obtained under special licence (Institute of Education, 2015b). To proxy the underlying prevalence of influenza, we construct a rate of the annual number of pneumonia notifications occurring between the months of September and November averaged across the two years prior to the pregnancy period of the NCDS cohort (1955 and 1956), per 100,000 persons[14]. These were derived from weekly data on pneumonia notifications by local authority, obtained from the Registrar General's Returns for England and Wales, and of Scotland (Registrar General for England and Wales, 1957; Registrar General for Scotland, 1957)[15].

Local authority stillbirth rates were derived from the Registrar General's Returns as a ratio of the total number of stillbirths in 1956 to the total number of births in the same year. Population density, measured as the population in thousands per square kilometre, was constructed using population figures from the 1956 Registrar General's Returns and local authority area data from UKBORDERS. Finally, further controls are derived from the 1951 Census county reports, which provided statistics based on the total population of each local authority. These include the proportion of the male working population in an unskilled occupation and the proportion of men leaving school aged 16 or older. We also include indicators for each of Great Britain's 11 regions.

3. Methods

3.2. Empirical strategy

A primary concern with estimating the impact of *in-utero* exposure to influenza by comparing outcomes of exposed and non-exposed groups is that mothers who contract influenza may differ in characteristics that predict outcomes. A particular concern here is differences in underlying maternal health. Due to evidence of substantial intergenerational transmission of health even in early life (Coneus and Spiess, 2012), maternal and offspring health are likely to be positively correlated, meaning that if mothers in poorer health are at increased risk of contracting influenza, any negative effects of *in-utero* influenza exposure on mental health outcomes could be inflated.

---

[12] Although we cannot correct for the number of miscarriages here, we still believe this to be good quality proxy. This is supported by the 89% agreement rate between the proxy measure and the self-reported measure in the sample of cohort members for which self-reported data is available.

[13] In 1958, local authorities numbered 173 and contained an average of approximately 300,000 inhabitants.

[14] Official notification of influenza cases by physicians was not required by law in Britain during the study period, and Pneumonia is clinically related to influenza, with approximately 1 in 417 cases of influenza resulting in pneumonia (Hunter and Young, 1971).

[15] We thank Dr Elaine Kelly for the provision of this data.

However, due to previous evidence of high prevalence of influenza even in prime-age healthy adults during pandemic periods (Woodall et al., 1958), it is unclear *a priori* whether a maternal health gradient in influenza exposure should be expected in this sample. Summary statistics for child and family background characteristics available in the NCDS provides evidence against this (Table A2, Appendix A3), with the 13% of mothers reporting contracting influenza during preganacy experiencing similar (and marginally lower) rates of co-morbidity and previous stillbirth/neonatal death compared to the non-exposed group. This provides evidence that confounding due to unobserved markers of maternal health is unlikely to subantially bias simple differences in outcomes across groups.

However, to adjust for residual differences in characteristics across exposed and non-exposed groups, we identify the effects of *in-utero* exposure to influenza using inverse probability of treatment weighting (IPTW), which estimates causal effects under the assumption of no unobserved confounding. IPTW estimates treatment effects by re-weighting the sample based on the probability of treatment (here, the probability of influenza exposure) such that characteristics of exposed and non-exposed cohort members are identical, on average, in observed characteristics. Higher weights are placed on non-exposed cohort members which are more similar in characteristics to exposed cohort members.

IPTW was preferred to recently-developed matching methods such as coarsened exact matching, which have the potential to generate greater balance on covariates (Iacus et al., 2011), but can result in a substantial unmatched sample when the majority of covariates are binary or categorical. IPTW also has benefits over regression adjustment, as the latter relies heavily on extrapolation when there is insufficient overlap in distribution of covariates between exposed and non-exposed group (Stuart, 2010), and tends to preform poorly in these situations (Dehejia and Wahba, 2002, 1999). IPTW offers straightforward diagnostics to assess whether adjsutments can be made without extrapolation. A further benefit of IPTW is that that, similar to randomised experiments, the study design can be finalised prior to the analysis of exposure on the outcome (Rubin, 2001).

To implement IPTW, we first estimate the propensity score, $\rho = P(flu = 1|X)$, representing the probability of being exposed to influenza *in-utero* given values of the covariates (Rosenbaum and Rubin, 1983). Here, propensity scores are derived as predicted probabilities from a probit model including all covariates described above[16]. Weights are then derived which equal the inverse of the probability of treatment that the cohort member received, such that they equal $\frac{1}{\rho}$ for the exposed and $\frac{1}{1-\rho}$ for the non-exposed (Rosenbaum, 1987).

The average treatment effect (ATE), which represents the effects of moving the entire population from non-exposed to exposed (Imbens, 2004), is then estimated as:

$$\frac{1}{n}\sum_{i}^{N}\frac{flu_i Y_i}{\rho} - \frac{1}{n}\sum_{i}^{N}\frac{(1-flu_i)Y_i}{1-\rho}$$

where N is the number of observations.

[16] Results are unchanged when propensity scores are derived from a logit model.

IPTW should ensure that the "balancing property" holds, such that the distribution of covariates are equivalent between the sample of exposed and non-exposed cohort members. We examine covariate balance in multiple ways (Austin and Stuart, 2015). First, we examine standardised differences in covariates between exposed and non-exposed groups, both prior to and following re-weighting the data, as well as examining whether re-weighting results in the ratio of variance of the covariates between groups converging to one (representing equal variances across groups). Finally, we formally test balance using a test derived in Imai and Ratkovic (2014) which treats restrictions imposed by balance as overidentifying restrictions.

IPTW also requires the validity of the overlap (or positivity) assumption, which states that each individual must have a positive probability of treatment. In cases where propensity scores are low, weights become large, which can cause IPTW estimates to become erratic due to an insufficient number exposed of observations to estimate treatment effects at the tails of the propensity score distribution (Cole and Hernán, 2008). To test for sufficient overlap, we compare kernal density plots of the propensity score for both exposed and non-exposed cohort members. The overlap assumption is violated when estimated densities have too much mass around 0 or 1 (Busso et al., 2014). We use heteroskedastic-robust standard errors throughout.

We estimate effects for the full sample and also seperately for males and female offspring, given evidence from previous stuides that the effects of *in-utero* shocks may differ by gender[17].

3.2. Derivation of the estimation samples

Due to substantial attrition in the NCDS, separate estimation samples were constructed for each outcome. The process describing the construction of these samples is depicted in Figure A3 (Appendix A4). In total, 18,558 individuals completed an interview for at least one wave of the NCDS. We began by dropping individuals with no PMS record, reducing the sample to 17,421 cohort members. This primarily included the non-British-born individuals added to the NCDS at the childhood follow-ups. We then excluded the 452 cases where data on individual-level exposure to influenza was missing[18]. For the age at which each outcome was recorded, we dropped individuals who failed to return a survey at this age. We then dropped individuals with missing data on the outcomes. We finally dropped observations with missing data on any of the control variables. Implementing this process resulted in final estimation samples of ranging between 12,291 (Rutter index at age 7) and 8,002 (Malaise at age 50).

[17] See for example Arthi (2018), Fletcher (2018), Havari and Peracchi (2017), Lin and Liu (2014), Lindeboom et al., 2010; Neelsen and Stratmann, 2012; Scholte et al., 2015; van den Berg et al., 2015; Yemelyanau et al., 2012).

[18] 426 of these relate to multiple births (212 twin-pairs and 4 sets of triplets), whose influenza data was not made available in the NCDS.

4. Results

4.1. Covariate balance and overlap

Consistent with the summary statistics, probit estimates used to derive propensity scores (Table A4, Appendix A6), suggest few differences in average levels of observed covariates between those exposed and not exposed to influenza *in-utero* in each estimation sample, with average marginal effects indicating that the majority of covariates, including parental socioeconomic status and maternal health, are not associated with the likelihood of influenza exposure *in-utero*[19]. Some patterns are present for maternal age and the number of siblings, where having a large number of siblings and having a younger mother are both associated with a statistically significant increase in the probability of influenza exposure, the former being consistent with children being a primary source of influenza spread. Rates of exposure were also slightly higher for cohort members whose mothers were underweight prior to pregnancy, and lower for cohort members born in local authorities with high stillbirth rates. The difference in covariates is slighltly more pronounced at adult ages, likely driven by the effects of sample attrition, with the largest differences being for the number of older siblings.

Despite small differences in mean levels of covariates prior to weighting, standardised differences and variance ratios for the unweighted and weighted data suggest that IPTW was successful in acheving greater covariate balance (Table 1; Table A3, Appendix A5). Following weighting, for the majority of covariates, standardized differnces in covariates between exposed and non-exposed groups were zero (to two decimal places), and variance ratios were extremely close to one, implying the mean values and the spread of the covariate distributions were approximately identical across groups. This is confirmed in formal balance tests, where balance could not be rejected in any of the outcome-specific estimation samples (Table 2).

Distributions of propensity scores for the exposed and non-exposed groups provide evidence in favour of the validity of the overlap assumption (Figure 2). Propensity scores were marginally larger on average in the group exposed to influenza *in-utero*, but densities of cohort members from exposed and non-exposed groups are similar at most parts of the propensity score distribution; consistent with the small unweighted differences in covariates across groups in Table 1. Numbers of cohort members with proponsity scores close to zero are also small.

[19] Statistically insignificant marginal effects don't seem to be driven by high levels of multicollinearity between covariates. The mean variance inflation factor (VIF) for the propensity score estimated model using ordinary least squares is 2.84, driven primarily by high values on ordered categorical variables where high VIFs are commonplace.

Table 1: Covariate standardised differences and variance ratios before and after inverse probability of treatment weighting.

	Standardized differences		Variance ratio	
	Unweighted	**Weighted**	**Unweighted**	**Weighted**
Male gender	-0.019	0.001	1.001	1.000
Ethnic minority	-0.023	-0.009	0.797	0.917
Mother's occupational social class (Base: V: Unskilled)				
Unemployed	**-0.034**	0.001	1.016	1.000
I or II: Professional, managerial or technical	**0.010**	-0.007	1.044	0.970
III: Skilled manual/non-manual	**0.010**	0.001	1.014	1.001
IV: Partly skilled	**0.031**	0.001	1.087	1.004
Father's occupational social class (Base: V: Unskilled)				
I: Professional	-0.051	0.011	0.778	1.050
II: Managerial or technical	-0.099	-0.007	0.789	0.985
III: Skilled manual/non-manual	**0.077**	-0.002	0.970	1.001
IV: Partly skilled	**0.009**	0.009	1.021	1.021
Information missing	0.000	-0.010	1.000	0.953
Mother schooled past leaving age	-0.039	0.000	0.955	1.000
Mother's age in years (Base: 20 or under)				
21-25	0.057	0.003	1.045	1.003
26-30	-0.069	0.000	0.944	1.000
31-35	0.006	-0.010	1.010	0.984
36-40	**-0.006**	0.010	0.983	1.032
Over 40	**-0.035**	-0.006	0.768	0.961
Mother's height (inches)	**0.026**	-0.011	1.055	1.062
Father's age in years (Base: 20 or under)				
21-25	0.051	0.003	1.077	1.005
26-30	-0.003	0.001	0.998	1.000
31-35	-0.022	-0.001	0.973	0.998
36-40	-0.013	0.003	0.971	1.007
41-45	0.032	0.002	1.139	1.007
45-50	-0.025	-0.006	0.817	0.951
Over 50	-0.075	-0.008	0.276	0.900
Information missing	-0.005	-0.003	0.973	0.980
Mother's smoking prior to pregnancy (Base: Non-smoker)				
1-4 cigarettes per day	0.024	0.001	1.077	1.004
5-9 cigarettes per day	**-0.012**	0.005	0.975	1.011
10-14 cigarettes per day	**-0.015**	0.001	0.968	1.002
15-19 cigarettes per day	**-0.029**	0.004	0.844	1.024
20+ cigarettes per day	0.009	-0.001	1.057	0.995
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])				
Underweight [BMI<18.5]	**-0.071**	0.000	0.679	1.002
Overweight [25<BMI<30]	0.030	0.001	1.047	1.001
Obese [BMI>30]	0.039	-0.001	1.183	0.994
Antenatal visits within 7 weeks of pregnancy	**0.011**	-0.001	1.050	0.995
Any comorbidity†	**-0.014**	-0.008	0.952	0.972
Previous abortion	**-0.003**	0.005	0.993	1.011
Previous premature birth (<5lb,9oz)	0.040	0.007	1.127	1.022
Previous large birth (>8lb,14oz))	0.063	0.003	1.184	1.008
Previous stillbirth/neonatal death	**-0.012**	-0.001	0.951	0.996
Complication in previous pregnancy	0.038	0.000	1.116	1.000
Persons per room (Base: <1)				
1-1.5	-0.010	-0.005	0.984	0.991
1.5-2	0.002	0.006	1.006	1.016
>2	0.027	0.001	1.126	1.003
Number of older siblings (Base: No siblings)				
1	-0.017	-0.005	0.988	0.996
2	0.041	0.010	1.081	1.020
3	-0.010	0.004	0.969	1.013
4	0.034	-0.006	1.182	0.969
5	0.072	-0.001	1.612	0.996
6+	0.018	0.010	1.123	1.066

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.
†Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy
Highlighted in bold are the standardised differences which suggest that, compared to those who were not exposed, individuals exposed to influenza *in-utero* have levels of maternal characteristics that are less predictive of poor outcomes.

Table 1 (continued): Covariate standardised differences and variance ratios before and after inverse probability of treatment weighting.

	Standardized differences		Variance ratio	
	Unweighted	**Weighted**	**Unweighted**	**Weighted**
Stillbirth rate (Still births/total births)	0.065	0.009	1.396	1.046
Proportion of male working population unskilled	-0.051	-0.002	0.948	1.014
Proportion of men leaving school aged 16 or older	-0.049	0.002	0.897	0.961
Population density (population in thousands/square km)	-0.055	-0.009	0.917	0.962
Pneumonia Notifications Per Thousand Population prior to pregnancy	-0.038	0.008	0.828	1.019
Region of birth (Base: Northern England)				
North West England	-0.066	-0.003	0.860	0.994
East & West Riding, England	0.023	-0.005	1.070	0.986
North Midlands, England	0.046	-0.001	1.156	0.996
Midlands, England	0.100	0.004	1.293	1.010
East England	0.020	0.002	1.061	1.006
South East England	-0.079	-0.012	0.867	0.979
South England	0.018	-0.008	1.073	0.970
South West England	-0.010	0.005	0.960	1.021
Wales	0.048	0.006	1.204	1.024
Scotland	-0.051	0.018	0.883	1.041

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.

Table 2: Over-identification tests for covariate balance following inverse probability of treatment weighting for each outcome-specific estimation sample

	Chi-squared statistic	**P-value**
Rutter: age 7	48.181	0.951
Rutter: age 11	42.762	0.988
Malaise: age 23	51.180	0.910
Malaise: age 33	43.675	0.985
Malaise: age 42	39.818	0.996
Malaise: age 50	41.507	0.992

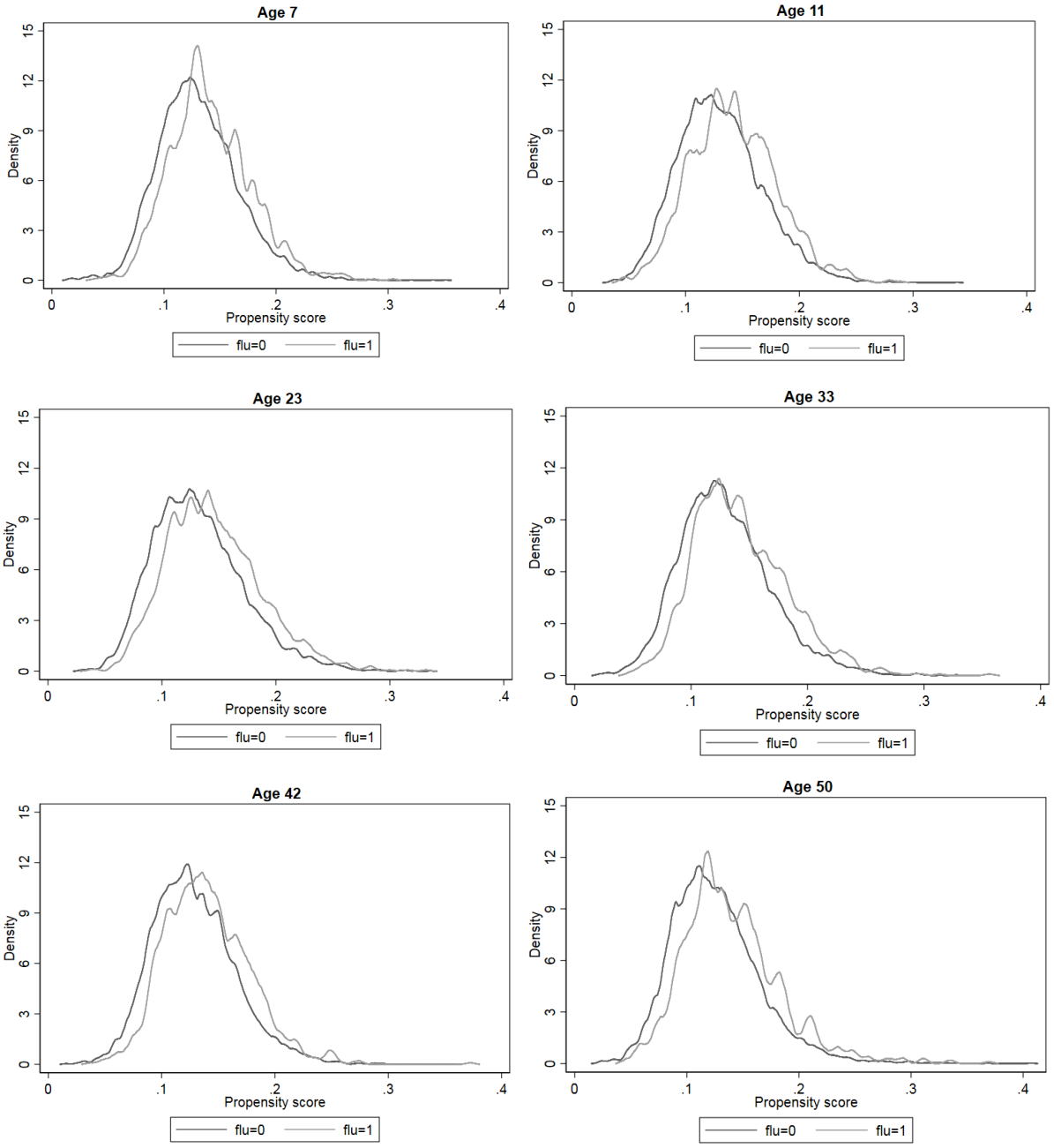


Figure 2: Kernel density plots of estimated propensity scores for both exposed and non-exposed cohort members in each estimation sample. Propensity scores were estimated from a probit regression.

4.2. Average Treatment Effect estimates

IPTW estimates of ATEs suggest that *in-utero* exposure to influenza led to small reductions in mental health in childhood at both ages 7 and 11 (Table 3). Reductions were smaller at age 7, where *in-utero* exposure led to a small but statistically insignificant 0.103 increase in the overall Rutter index; equivalent to 1.63% increase relative to the mean index for the non-exposed group. At age 11, the magnitude of the effect doubles to 0.194 (3.04%) and becomes statistically significant at a 5% level. At both ages, reductions in mental health were primarily driven by an increase in internalising behaviour (3.68% at age 11). Consistent with minimal differences in covariates, these estimates are almost identical to unweighted effects estimated using ordinary least squares.

In-utero influenza exposure also led to small increases in the number of depressive symptoms included in the 24-item Malaise inventory at ages 23, 33, and 42, ranging from an increase of 0.05 symptoms (1.3%) at age 42 to an increase of 0.08 symptoms (3.42%) at age 33, although these were not statistically significant (Table 1). At age 50, the magnitude rises to a 0.09 symptom increase (equivalent to 6.35% given the switch to the 9-item inventory at this age), but again this effect is not statistically significant. The lack of statistical significance despite moderate effect sizes is likely to be driven by reduced sample sizes at later ages. Differences between weighted and unweighted effects are larger in adulthood, consistent with greater differences in covariates prior to weighting.

4.3. Heterogeneity by gender

Effects differ considerably by gender, with this heterogeneity varying by the age at which mental health is measured (Table A5, Appendix A7). At age 7, effects on the overall Rutter index and internalising and externalising subscales are stronger for males, but remain statistically insignificant for both genders. Effects on the Rutter index at age 11 are stronger for females, driven by a statistically significant increase in internalising symptoms of 7.72%. However effects on externalising symptoms remain stronger for males (4.57% vs -2.05%) although effects are again not statistically significant.

Effects of *in-utero* influenza exposure on the Malaise inventory are not statistically significant for both genders at ages 23, 33, and 42, with some evidence of larger effects for females. However, at age 50, effects are much larger for males, with *in-utero* influenza exposure resulting in a statistcially significant 0.176 (15.45%) increase in depressive symptoms.

Table 3: Inverse probability weighted estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50

	Observations	Unweighted ATE	Potential outcome mean for the non-exposed group	Weighted ATE	Weighted ATE as % of non-exposed mean
Rutter: age 7					
Overall	12,291	0.103 (0.098)	6.312*** (0.035)	0.103 (0.095)	1.63%
Internalising	12,291	0.080 (0.061)	3.322*** (0.021)	0.076 (0.060)	2.29%
Externalising	12,291	0.026 (0.060)	2.980*** (0.021)	0.027 (0.058)	0.90%
Rutter: age 11					
Overall	11,430	0.189 (0.099)	6.355*** (0.035)	0.194* (0.098)	3.04%
Internalising	11,430	0.127* (0.061)	3.552*** (0.022)	0.131* (0.060)	3.68%
Externalising	11,430	0.060 (0.060)	2.796*** (0.021)	0.061 (0.058)	2.17%
Malaise: age 23	10,344	0.142 (0.085)	2.686*** (0.031)	0.054 (0.081)	1.99%
Malaise: age 33	9,318	0.156 (0.091)	2.372*** (0.033)	0.082 (0.089)	3.42%
Malaise: age 42	9,344	0.113 (0.110)	3.551*** (0.040)	0.046 (0.108)	1.29%
Malaise: age 50	8,002	0.133* (0.067)	1.449*** (0.023)	0.093 (0.066)	6.35%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

5. Robustness checks

5.1. Alternative methods for adjusting for residual differences in covariates

IPTW estimators model the probability of exposure, with no assumptions made about the functional form of the outcome model. The consistency of IPTW estimators relies solely on specifying the correct functional form of the probability of *in-utero* influenza exposure. To relax this assumption, we employ the inverse-probability-weighted regression adjustment (IPWRA) estimator which combines inverse-probability of treatment weighting with regression adjustment (Wooldridge, 2007). IPWRA is a member of a class of "doubly-robust" (DR) methods which are consistent when the functional form for either the probability of exposure or the outcome is correctly specified. IPWRA estimates are obtained via a four-step process (Wooldridge, 2010):

1. Inverse-probability of treatment weights are estimated as above.
2. A weighted regression of the outcomes on the set of covariates is performed seperately for the exposed and non-exposed groups, with weights equal to the inverse-probabiliy weights from step 1. Consistent with the outcome distributions (Appendix A2), we estimate linear models for the Rutter indices and Poisson models for the Malaise symptom counts.
3. Fitted values are generated for all cohort members from both regressions in step 2. These fitted values represent the predicted outcomes for each cohort member in each treatment state.
4. Cohort member-specific treatment effects are estimated as the difference between their predicted outcomes in each treatment state, and the ATE is estimated as average of these treatment effects across all cohort members.

IPWRA and IPW ATE estimates for the Rutter indices are very similar in magnitude, although the effect of influenza exposure on the overall Rutter index at age 11 becomes statistically insignificant at the 5% level (Table 4). Statistical significance of the effect on the internalising component of the Rutter index is maintained. For adult mental health outcomes, the use of IPWRA increases the magnitude of effects but they remain statistically insignificant at a 5% level.

Table 4: Inverse-probability-weighted regression adjustment estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50

	Observations	Potential outcome mean for the non-exposed group	ATE	ATE as % of non-exposed mean
Rutter: age 7				
Overall	12,291	6.313*** (0.035)	0.109 (0.095)	1.73%
Internalising	12,291	3.322*** (0.021)	0.081 (0.060)	2.44%
Externalising	12,291	2.980*** (0.021)	0.028 (0.059)	0.94%
Rutter: age 11				
Overall	11,430	6.355*** (0.035)	0.188 (0.098)	2.96%
Internalising	11,430	3.552*** (0.022)	0.130* (0.061)	3.66%
Externalising	11,430	2.796*** (0.021)	0.055 (0.058)	1.97%
Malaise: age 23	10,344	2.685*** (0.031)	0.061 (0.082)	2.27%
Malaise: age 33	9,318	2.372*** (0.033)	0.089 (0.090)	3.75%
Malaise: age 42	9,344	3.550*** (0.040)	0.063 (0.108)	1.77%
Malaise: age 50	8,002	1.449*** (0.023)	0.100 (0.067)	6.90%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

5.2. Sources of bias

5.2.1. Non-random sample selection

IPTW estimates could be impacted by two forms of non-random sample selection: mortality selection (survival bias) and survey non-response. Evidence suggests that it is infants at the bottom of the health distribution who die as a result of infection (Bozzoli et al., 2009). Any increase in mortality rates as a result of *in-utero* exposure to influenza therefore indicates that individuals in poor mental health may have been pruned from the exposed group, increasing mental health relative to the unexposed group. This subsequently leads to an under-estimation of the effects of influenza exposure.

Figure 4 depicts how the percentage of NCDS cohort members who have died increases with age, and Table 5 and Figure A4 (Appendix A8) present IPTW estimates of the effect of *in-utero* influenza exposure on the probability of stillbirth/neo-natal death, death within 28 days of birth, and death by ages 7, 11, 16, 23, 33, 42, 46, and 50. For individuals not exposed to influenza *in-utero*, approximately 1.9% of individuals are stillborn, consistent with an infant mortality rate (age <1) of 2.4% in the UK for 1958 (CLOSER, 2020). Rates of death increase to 4.4% by age 7, 5.6% by age 33 and 7.8% by age 50.

A large effect of exposure is found at birth, where *in-utero* exposure to influenza increases the probability a stillbirth/neo-natal death by 1.2 percentage points. This is equivalent to 75% increase compared to the estimated stillbirth rate in the not-exposed group. This effect increases as the cohort ages, although not monotonically, to 1.7 percentage points by age 46. Effects are also stronger relative to unweighted estimated suggesting results are not explained by unobserved confounding. Further analysis suggests that most of the mortality effect occurs before birth, with effects at later ages being positive but losing statistical significance after conditioning on survival beyond birth (Figure A5, Appendix 8). These substantial effects on mortality suggest that survival bias may be causing IPTW estimates to be conservative estimates of the effects of *in-utero* influenza exposure on mental health, with this bias more pronounced at older ages.

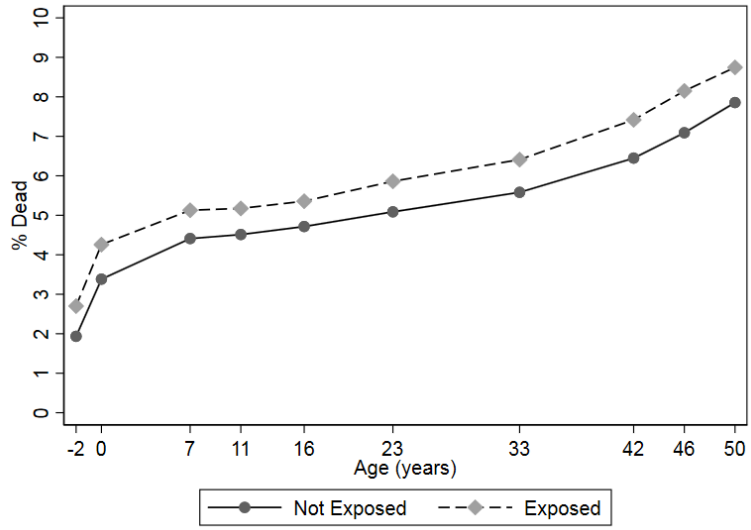


Figure 3: Proportion of NCDS cohort members who are dead at each age at follow-up. Age of -2 refers to birth, and age of 0 refers to 28 day post-birth.

Table 5: Unweighted and inverse probability weighted estimates of the ATE of *in-utero* exposure to influenza on the probability of death at birth, 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50.

	Observations	Unweighted	Potential outcome mean for the non-exposed group	ATE	ATE as % of non-exposed mean
Birth	15,019	0.010** (0.004)	0.016*** (0.001)	0.012** (0.004)	75.00%
28 days	15,019	0.012* (0.005)	0.029*** (0.001)	0.015** (0.005)	51.72%
Age 7	15,019	0.011* (0.005)	0.039*** (0.002)	0.013* (0.006)	33.33%
Age 11	15,019	0.010* (0.005)	0.040*** (0.002)	0.013* (0.006)	32.50%
Age 16	15,019	0.010 (0.005)	0.043*** (0.002)	0.013* (0.006)	30.23%
Age 23	15,019	0.012* (0.006)	0.046*** (0.002)	0.015* (0.006)	32.61%
Age 33	15,019	0.013* (0.006)	0.051*** (0.002)	0.016** (0.006)	31.37%
Age 42	15,019	0.013* (0.006)	0.059*** (0.002)	0.017* (0.007)	28.81%
Age 46	15,019	0.014* (0.006)	0.066*** (0.002)	0.017* (0.007)	25.76%
Age 50	15,019	0.012 (0.007)	0.074*** (0.002)	0.015* (0.007)	20.27%

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.

To examine the potential consequences of differential mortality rates on treatment effects for mental health outcomes, we use a bounding approach similar in spirit to Manski (1990) where "worst case" bounds are estimated by replacing missing values with the maximum or mimumum of the outcomes. Given the postive mortality effects found above, we depart slightly from this approach by assuming monotonicity (similar to assumptions made in the bounding approach in Lee (2009)), such that being exposed to influenza *in-utero* only increases mortality risk and not reduces it. Thus the lower bound is equivalent to the treatment effects from the main specification and the upper bound equivalent to IPTW estimates after replacing missing values due to death with the maximum. In this most severe case, we cannot rule out substantial effects on mental health in childhood and adulthood, with effects on internalising symptoms at age 11 potentially as high as 8.37% relative to the non-exposed mean and effects on depressive symptoms at age 33 as high as 14.6% (Table A6, Appendix A9). As distributtions of the outcomes (Figures A1 and A2, Appendix A2) suggest outcome values at the maximum are extremely rare, we also estimate

bounds instead replacing missing values with the sample 90th percentile. In this case the bounds are much tighter, but still don't rule out effects which are large in magnitude.

Survey non-completion is another possible source of sample selection bias. Even for the sample of individuals who survive, a proportion of NCDS mothers fail to complete the survey at the ages of interest. Consistent with arguments made in published studies using data from the NCDS (Dearden et al., 2002; Jones et al., 2011; Lindeboom et al., 2009), it is possible that the large number of characteristics available in the NCDS means that missing data can be assumed exogenous, conditional on covariates. A lack of correlation between *in-utero* exposure to influenza and non-completion would further reduce the likelihood that non-response is problematic. To test this, we estimate the effects of influenza exposure on the probability of response in each survey wave, controlling for all other covariates[20] and restricting the sample to cohort members who are alive at each wave. The latter allows us to separate the effects of non-response from that of mortality selection. This outcome incorporates both unit non-response, where NCDS mothers did not complete the entire survey wave, and outcome non-response. We find that effects of *in-utero* exposure on the probability of completion are small and not statistically significant at any survey wave (Table A7, Appendix A10), implying that survey non-completion is not a likely source of bias.

5.2.2. Unobserved confounding

As higlighted previously, consistency of IPTW estimates relies on the asumptions of no unobserved confounding, such that there exists no unobserved variables correlated with both influenza exposure and mental health. We therefore implement a series of tests to examine the robustness of results to unobserved confounding under the assumption that the relationship between *in-utero* influenza exposure and the unobservables can be recovered from the relationship between influenza exposure and observables (Altonji et al., 2011). Under this assumption, random assignment with respect to observed confounders implies random assignment with respect to unobserved confounders, and coefficient stability following the inclusion or removal of covariates provides evidence against unobserved confounding (Altonji et al., 2005).

First, we estimate a series of coefficient comparison tests, where we iteratively remove single, pairs, and triplets of all proxies of maternal health and re-estimate ATEs using IPTW. Tight clustering of ATEs around the baseline estimate provides evidence against unobserved confounding, by suggesting that the addition of single unobserved confounders may also not impact results greatly. These proxies include the co-morbidity indicator (which equals one if a mother reported suffering from other diseases during pregnancy) and indicators of mother's experience in previous pregnancies (premature births, large births, stillbirth/neo-natal deaths, and birth-related complications).

Histograms of re-estimated ATEs for child mental health outcomes (Figure 4) show that the removal of proxies result in only small changes in ATEs, with ATEs approximtely normally distributed around the ATEs estimated in the main specification (shown by the vertical red lines). For some outcomes, the removal of covariates leads to a reduction in the ATE, suggesting that ATE from the main specification could represent a lower bound of the true ATE. Similar distributions are found for the ATEs for adult mental health outcomes (Figure 4), although there is some evidence that the removal of some covariates leads to an increase in the ATE. Differential non-response with respect to these covariates could be driving the larger effects of covariate-adjustment for these outcomes.

[20] Missing data on control variables from the PMS are implicitly assumed to be missing at random.

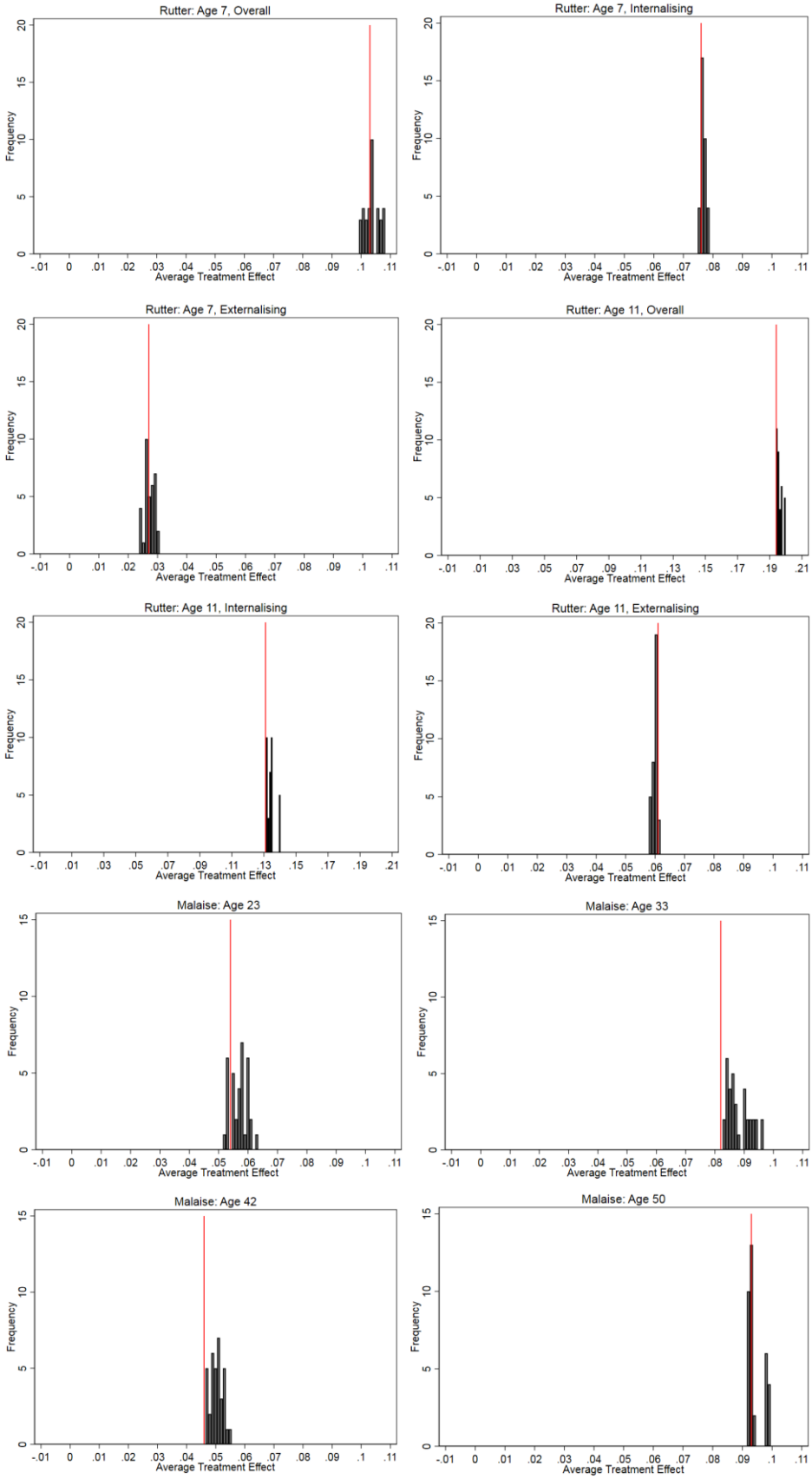


Figure 4: Histograms depicting variation in estimated average treatment effects due to the removal of single, pairs and triplets of proxies for genetic health endowments for mental health outcomes: Red vertical lines represent the average treatment effects with the full specification of covariates.

The overall stability in ATEs is perhaps unsurprising given the small unweighted differences in covariates between those exposed and not-exposed to influenza *in-utero* shown in Table 1.

Second, we implement a series of balancing regressions where the main specification is re-estimated, but with proxies for genetic health endowments used as the outcome. This is preferred to traditional coefficient comparison tests in situations where proxy confounders could be measured with error (Pei et al., 2019). For each proxy, we remove the proxy from the covariate set, re-estimate propensity scores, and use IPTW to estimate the ATE of *in-utero* influenza on the proxy. Null effects on the proxy are suggestive of exogeneity. Results indicate that influenza exposure is not statistically significntly associated with any indicator of genetic health endowments after accounting for other covariates, and in some cases association between influenza exposure and markers of poor maternal health are negative (Table 6).

Results from both tests therefore suggest confounding from unobserved shared health endowments is unlikely to be biasing results considerably.

Table 6: Inverse-probability-weighted estimates of the ATE of *in-utero* exposure to influenza on maternal health characteristics in each estimation sample.

	Comorbidity	**Previous premature birth**	**Previous large birth**	**Previous stillbirth/neo-natal death**	**Previous birth complication**
Age 7 sample	-0.004	0.009	0.008	-0.005	0.008
	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)
Age 11 sample	-0.006	0.009	0.008	-0.003	0.005
	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)
Age 23 sample	-0.006	0.004	0.006	0.004	0.011
	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)
Age 33 sample	-0.006	0.005	0.010	0.002	0.014
	(0.008)	(0.008)	(0.008)	(0.006)	(0.009)
Age 42 sample	-0.008	0.002	0.004	0.002	0.013
	(0.007)	(0.007)	(0.008)	(0.006)	(0.009)
Age 50 sample	-0.006	0.000	0.010	0.002	0.007
	(0.008)	(0.008)	(0.008)	(0.007)	(0.010)

* p<0.05, ** p<0.01, *** p<0.001; robust standard errors in parenthesis

6. Discussion and Conclusion

A large body of research in economics has investigated the relationship between *in-utero* exposure to influenza and later-life health, human capital, and labour market outcomes. However, there is little evidence of effects on mental health outcomes despite numerous potential mechanisms. Evidence on mortality effects are also rare, with no study examining how mortality risk due to influenza exposure develops as an individual ages. In addition, the majority of studies estimate only reduced-form effects of exposure to influenza pandemics, with little evidence on the effects of influenza at an individual-level.

This paper adds to the existing literature by using unique survey data on mother-reported *in-utero* influenza exposure and longitudinal data on the mental health of their offspring to investigate impacts on mental health outcomes recorded at multiple stages of the life-course and early death The data is also drawn from a cohort of individuals who were *in-utero* during the 1957 Asian Flu epidemic, increasing the likelihood of influenza exposure is independent of unobserved confounders. This is supported by the data, which shows that although influenza exposure was correlated with common predictors of disease spread (e.g. the number of siblings), exposed and non-exposed

individuals were similar on markers of maternal health and there was no systematic pattern of socio-demographic selection into maternal influenza. After adjusting for small residual differences between exposed and non-exposed groups using inverse probability weighting and "doubly-robust" inverse-probability-weighted regression adjustment, our findings suggest that *in-utero* exposure to influenza is associated with small reductions in mental health in mid-childhood, driven by increases in internalising symptoms, and small increases in depressive symptoms in mid-life for males. We also find that *in-utero* influenza exposure is associated with a significant reduction in survival rates. This effect is driven primarily by a 75% increase in the relative risk of stillbirth, although we also find small increases in the magnitude of mortality effects in adulthood.

In addition, in a series of balancing regressions and coefficient comparison tests, we demonstrate limited sensitivity of results to removing maternal health and socio-demographic characteristics, particularly for child mental health and mortality outcomes. This, coupled with small unweighted differences in maternal health between exposed and non-exposed groups (with mothers of exposed individuals being in marginally better health on some markers) suggests that unobserved confounding is unlikely to be inflating the magnitudes of our estimates considerably. Nevertheless, the risk of some unobserved confounding remains.

Furthermore, substantial effects of *in-utero* exposure to influenza on early death indicate that estimated effects on mental health are conservative. This may explain the smaller and statistically insignificant effects we find on mental health in adulthood. Estimated bounds under survival bias could not rule out large effects on mental health outcomes. Also, as influenza cases are self-reported there is a potential for measurement error, which is likely to have further attenuated effects towards zero.

Our results are consistent with the limited number of studies finding mixed evidence on the effects of *in-utero* health shocks on mental health outcomes. Persson and Rossin-Slater (2018) find that exposure to maternal stress due to death of a close relative leads to increased use of ADHD medicine in childhood and anti-anxiety and depression medication in adulthood. Brown et al., (2000) find that exposure to the Dutch Winter Famine of 1944-45 led to an increase in the risk of admission for affective disorders. However, Maselko et al., (2015) find that a reduction in prenatal depression caused by a maternal cognitive behavioural therapy intervention did not lead to an increase in socio-emotional development of their offspring.

Our findings of generally stronger health effects of influenza exposure for males are consistent with evidence from other *in-utero* shocks. Havari and Peracchi (2017), Scholte et al. (2015), van den Berg et al. (2015), and Lindeboom et al. (2010) all find that the health effects of *in-utero* exposure to under-nutrition are stronger for males. Similarly, Arthi (2018) find that the effect of *in-utero* exposure to the Dust Bowl on physical disability is statistically significant only for males. These findings are also consistent with previous studies examining the effects of *in-utero* pandemic exposure on human capital and family formation outcomes, with effects on educational attainment and the probability of marriage being stronger for males (Lin and Liu, 2014; Neelsen and Stratmann, 2012).

Our finding of a significant effect on stillbirth rates is consistent with the medical literature that states that influenza can lead to complications in late pregnancy which can subsequently lead to stillbirths (Liu et al., 2012; Pierce et al., 2011). We also find that the magnitude of the effect on the probability of death remained relatively constant up to age 50, indicating a persistence in the survival advantage for those not exposed to influenza *in-utero*, but no large additional survival advantage for those living beyond infancy. This is consistent with the foetal origins hypothesis

which states that the health effects of *in-utero* adversity may remain latent until old age, when the onset of physical disease becomes prevalent (Barker, 1990), and corroborates findings in Fletcher (2019) who find no effects of exposure to the 1918 influenza pandemic on mortality by age 60 for individuals surviving to adulthood. This is also consistent with an evaluation of the long-term health effects of a neonatal program in Sweden, where the mortality advantage for the treated individuals began in infancy and remained constant throughout adulthood, before increasing again at age 70 (Bhalotra et al., 2015). A study examining impacts of influenza exposure on mortality at later ages is warranted once data on future NCDS survey waves become available.

One limitation of our paper is that, if the 1957 Asian Flu was responsible for the majority of the influenza cases in this cohort, our study will only examine the effects of *in-utero* exposure to influenza in a relatively small period of gestation (between 16 and 25 weeks). Therefore, given all NCDS cohort members were born in the same week, we are not able to examine whether there is heterogeneity in effects by trimester of exposure. Furthermore, a lack of statistically significant effects on mental health at some ages may be driven by a lack of power to detect these effects due to limited sample sizes. Examination of this question using longitudinal data on consistent measures of mental health across all ages would be an interesting extension of this study. Finally, external validity must be considered when interpreting results. The NCDS is a historical cohort and our study documents the effects of *in-utero* exposure to influenza in the 1950s. It is unclear whether exposure to influenza would have similar deleterious effects in more recent cohorts. However, the requirement for contemporaneously-measured exposure and outcomes means this limitation is unavoidable in studies such as this.

Despite these limitations, this paper presents findings which are relevant for policy-makers. In the US and UK, pregnant women are advised to have an influenza vaccine to protect against the risk of stillbirths and other pregnancy-related complications, with vaccination also being free-at-the-point-of-use in the UK (Centres for Disease Control and Prevention, 2013; NHS, 2015; Public Health England, 2015). However, recommendations are not ubiquitous elsewhere. Recommendations are still not in place in 5/31 countries in the European Economic Area (ECDPC, 2015), and a World Health Organisation (WHO) report suggests that 37/47 countries were not recommending influenza vaccinations for all pregnant women for the 2008-09 influenza season (World Health Organization, 2014). Furthermore, a 2011 survey found that under 51% of the 37 countries in the Western Pacific Region recommended vaccines for pregnant women (Dwyer et al., 2013).

In addition, even in countries where recommendations are in place, uptake of the influenza vaccines amongst pregnant women is still low. In the 2013-14 influenza season, uptake rates in the US stood at only 52% (Centers for Disease Control and Prevention, 2014), and in the UK, uptake rates fell from 47.2% to 45.2% between the 2017/18 and 2018/19 winter influenza seasons (Public Health England, 2019). A 2014 report found that influenza led to 1 in 11 of maternal deaths, and estimated that over half of influenza deaths could have been prevented with influenza vaccinations (Knight et al., 2014). Our findings, alongside those from previous studies, suggest the negative impacts of influenza during pregnancy extend beyond the mother to their unborn children. Given that current evidence indicates that influenza vaccines are successful in protecting mothers from influenza, and that there are no serious side-effects of these vaccines for either mother or infant (Mak et al., 2008), our results suggest that the roll-out of recommendations for influenza vaccinations for pregnant women across all countries should be explored and that interventions to improve uptake in countries where recommendations are in place are warranted.

7. Bibliography

Almond, D., 2006. Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population. J. Polit. Econ. 114, 672–712.

Almond, D., Currie, J., Duque, V., 2018. Childhood circumstances and adult outcomes: Act II. J. Econ. Lit. 56, 1360–1446. doi:10.1257/jel.20171164

Almond, D., Mazumder, B., 2005. The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data. Am. Econ. Rev. 95, 258–262.

Altonji, J.G., Conley, T.G., Elder, T.E., Taber, C.R., 2011. Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables. Mimeo.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. J. Polit. Econ. 113, 151–184. doi:10.1086/426036

Arthi, V., 2018. "The dust was long in settling": Human capital and the lasting impact of the American Dust Bowl. J. Econ. Hist. 78, 196–230. doi:10.1017/S0022050718000074

Austin, P.C., Stuart, E., 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat. Med. 34, 3661–3679. doi:10.1002/sim.6607

Barker, D., 1990. The fetal and infant origins of adult disease. BMJ Br. Med. J. 301, 1111.

Bhalotra, S., Karlsson, M., Nilsson, T., 2015. Infant health and longevity: evidence from a historical trial in Sweden (No. 2015–08), ISER Working Paper Series.

Blanden, J., Gregg, P., Macmillan, L., 2007. Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. Econ. J. 117, C43–C60. doi:10.1111/j.1468-0297.2007.02034.x

Bozzoli, C., Deaton, A., Quintana-Domeque, C., 2009. Adult height and childhood disease. Demography 46, 647–69. doi:10.1353/dem.0.0079

Brown, A.S., Van Os, J., Driessens, C., Hoek, H.W., Susser, E.S., 2000. Further evidence of relation between prenatal famine and major affective disorder. Am. J. Psychiatry 157, 190–195. doi:10.1176/appi.ajp.157.2.190

Busso, M., DiNardo, J., McCrary, J., 2014. New evidence on the finite sample properties of propensity score reweighting and matching estimators. Rev. Econ. Stat. 96, 885–897. doi:10.1162/REST_a_00431

Butler, N.R., Bonham, D.G., 1963. Perinatal Mortality. E & S Livingstone, Edinburgh.

Campion, E.W., Kourtis, A.P., Read, J.S., Jamieson, D.J., 2014. Pregnancy and Infection. N. Engl. J. Med. 23370, 2211–2218. doi:10.1056/NEJMra1213566

Cassini, A., Colzani, E., Pini, A., Mangen, M.-J.J., Plass, D., McDonald, S.A., Maringhini, G., van Lier, A., Haagsma, J.A., Havelaar, A.H., Kramarz, P., Kretzschmar, M.E., 2018. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. Eurosurveillance 23. doi:10.2807/1560-7917.es.2018.23.16.17-00454

Centers for Disease Control and Prevention, 2015. About Adult BMI [WWW Document]. URL http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/ (accessed 4.20.16).

Centers for Disease Control and Prevention, 2014. Influenza Vaccination Coverage Among Pregnant Women — United States, 2013–14 Influenza Season. Morb. Mortal. Wkly. Rep. 63, 816–821.

Centre for Longitudinal Studies, 2012. Measuring behaviour difficulties of children and teenagers.

Centres for Disease Control and Prevention, 2013. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2013–2014.

Clarke, S.K.R., Heath, R.B., Sutton, R.N.P., Stuart-Harris, C.H., 1958. SEROLOGICAL STUDIES WITH ASIAN STRAIN OF INFLUENZA A. Lancet 271, 814–818. doi:10.1016/S0140-6736(58)91739-2

CLOSER, 2020. Infant mortality rate (deaths under 1 year per 1,000 live births), 1930 - 2019 (United Kingdom) [WWW

Document]. URL https://www.closer.ac.uk/data/infant-mortality/ (accessed 4.15.21).

Cole, S.R., Hernán, M.A., 2008. Constructing inverse probability weights for marginal structural models. Am. J. Epidemiol. 168, 656–664. doi:10.1093/aje/kwn164

Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S., Bordin, I.A., Costello, E.J., Durkin, M., Fairburn, C., Glass, R.I., Hall, W., Huang, Y., Hyman, S.E., Jamison, K., Kaaya, S., Kapur, S., Kleinman, A., Ogunniyi, A., Otero-Ojeda, A., Poo, M.-M., Ravindranath, V., Sahakian, B.J., Saxena, S., Singer, P.A., Stein, D.J., Anderson, W., Dhansay, M.A., Ewart, W., Phillips, A., Shurin, S., Walport, M., Timmermans, H., 2011. Grand challenges in global mental health. Nature 475, 37–30. doi:10.1038/475027a.Grand

Collishaw, S., 2015. Annual research review: Secular trends in child and adolescent mental health. J. Child Psychol. Psychiatry Allied Discip. 56, 370–393. doi:10.1111/jcpp.12372

Colombo, J., Kannass, K.N., Shaddy, D.J., Kundurthi, S., Maikranz, J.M., Anderson, C.J., Blaga, O.M., Carlson, S.E., 2004. Maternal DHA and the development of attention in infancy and toddlerhood. Child Dev.

Coneus, K., Spiess, C.K., 2012. The intergenerational transmission of health in early childhood - Evidence from the German Socio-Economic Panel Study. Econ. Hum. Biol. 10, 89–97. doi:10.1016/j.ehb.2011.03.002

Currie, J., 2020. Child health as human capital. Health Econ. Online onl. doi:10.1002/hec.3995

Currie, J., Almond, D., 2011. Human capital development before age five, in: Card, D., Ashenfelter, T. (Eds.), Handbook of Labour Economics. North-Holland, Amsterdam.

Dearden, L., Ferri, J., Meghir, C., 2002. The effect of school quality on educational attainment and wages. Rev. Econ. Stat. 84, 1–20.

Dehejia, R., Wahba, S., 2002. Propensity Score Matching Methods for Non-experimental Causal Studies. Rev. Econ. Stud. 84, 151–161. doi:10.2139/ssrn.138259

Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. J. Am. Stat. Assoc. 94, 1053–1062.

Dwyer, D., Barr, I., Hurt, A., Kelso, A., Reading, P., Sullivan, S., Buchy, P., Yu, H., Zheng, J., Shu, Y., Wang, D., Lam, Aguon, A., Oliva, R.Q., Odagiri, T., Tashiro, M., Verasahib, K., Yusof, M.A., Nymadawa, P., Alexander, B., Gourinat, A.C., Grangeon, J.P., Jennings, L., Huang, S., Horwood, P., Lucero, M., Roque Jr., V., Lee Suy, L., Cardon, P., Tandoc 3rd, A., Olveda, R.M., Kang, C., Young-Joon, P., Cutter, J., Lin, R., Low, C., Mai le, T.Q., Balish, A., Kile, J., Mei, S., McFarland, J., Moen, A., Olsen, S., Samaan, G., Xiyan, X., Chea, N., Diorditsa, S., Feldon, K., Fox, K., Jamsran, M., Konings, F., Lewis, H.C., McPherson, M., Nilles, E., Olowokure, B., Partridge, J., 2013. Seasonal influenza vaccine policies, recommendations and use in the World Health Organization's Western Pacific Region, Western Pacific Surveillance and Response Journal. doi:10.5365/WPSAR.2013.4.1.009

ECDPC, 2015. Seasonal influenza vaccination in Europe: Overview of vaccination recommendations and coverage rates in the EU Member States for the 2012–13 influenza season. Stockholm.

Ferrari, A., Charlson, F., Norman, R., Patten, S., Freedman, G., Murray, C., Vos, T., Whiteford, H., 2017. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. PLoS Med. 10, 319–327. doi:10.1371/journal.pmed.1001547

Fletcher, J.M., 2019. Examining the long-term mortality effects of early health shocks. Appl. Econ. Lett. 26, 902–908. doi:10.1080/13504851.2018.1520960

Fletcher, J.M., 2018. The effects of in utero exposure to the 1918 influenza pandemic on family formation. Econ. Hum. Biol. 30, 59–68. doi:10.1016/j.ehb.2018.06.004

Gale, C.R., Robinson, S.M., Godfrey, K.M., Law, C.M., Schlotz, W., O'Callaghan, F.J., 2008. Oily fish intake during pregnancy - Association with lower hyperactivity but not with higher full-scale IQ in offspring. J. Child Psychol. Psychiatry Allied Discip. 49, 1061–1068.

Garthwaite, C., 2008. The effect of in-utero conditions on long term health: evidence from the 1918 Spanish flu pandemic.

Havari, E., Peracchi, F., 2017. Growing up in wartime: Evidence from the era of two world wars. Econ. Hum. Biol. 25, 9–32. doi:10.1016/j.ehb.2016.09.002

Hirst, M.A., 1983. Evaluating the Malaise Inventory - An item analysis. Soc. Psychiatry 18, 181–184. doi:10.1007/BF00583528

Hirst, M.A., Bradshaw, J.R., 1983. Evaluating the Malaise inventory: A comparison of measures of stress. J. Psychosom. Res. 27, 193–199. doi:10.1016/0022-3999(83)90022-3

Hunter, J., Young, J., 1971. Diffusion of influenza in England and Wales. Ann. Assoc. Am. Geogr. 61.

Iacus, S.M., King, G., Porro, G., 2011. Causal Inference without Balance Checking: Coarsened Exact Matching. Polit. Anal. 20, 1–24. doi:10.1093/pan/mpr013

Imai, K., Ratkovic, M., 2014. Covariate balancing propensity score. J. R. Stat. Soc. Ser. B Stat. Methodol. 76, 243–263. doi:10.1111/rssb.12027

Imbens, G., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. Rev. Econ. Stat. 86, 4–29.

Institute of Education, 2015a. National Child Development Study Deaths Dataset, 1958-2014: Special Licence Access (SN7717). Accessed via the UK Data Service.

Institute of Education, 2015b. National Child Development Study: Local Authority Data, 1958-1974: Special Licence Access (SN5744). Accessed via the UK Data Service.

Iuliano, A.D., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Gran, J.M., Schanzer, D., Cowling, B.J., Wu, P., Kyncl, J., Ang, L.W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., van Asten, L., Pereira da Silva, S., Aungkulanon, S., Buchholz, U., Widdowson, M.A., Bresee, J.S., Azziz-Baumgartner, E., Cheng, P.Y., Dawood, F., Foppa, I., Olsen, S., Haber, M., Jeffers, C., MacIntyre, C.R., Newall, A.T., Wood, J.G., Kundi, M., Popow-Kraupp, T., Ahmed, M., Rahman, M., Marinho, F., Sotomayor Proschle, C.V., Vergara Mallegas, N., Luzhao, F., Sa, L., Barbosa-Ramírez, J., Sanchez, D.M., Gomez, L.A., Vargas, X.B., Acosta Herrera, a. B., Llanés, M.J., Fischer, T.K., Krause, T.G., Mølbak, K., Nielsen, J., Trebbien, R., Bruno, A., Ojeda, J., Ramos, H., an der Heiden, M., del Carmen Castillo Signor, L., Serrano, C.E., Bhardwaj, R., Chadha, M., Narayan, V., Kosen, S., Bromberg, M., Glatman-Freedman, A., Kaufman, Z., Arima, Y., Oishi, K., Chaves, S., Nyawanda, B., Al-Jarallah, R.A., Kuri-Morales, P.A., Matus, C.R., Corona, M.E.J., Burmaa, A., Darmaa, O., Obtel, M., Cherkaoui, I., van den Wijngaard, C.C., van der Hoek, W., Baker, M., Bandaranayake, D., Bissielo, A., Huang, S., Lopez, L., Newbern, C., Flem, E., Grøneng, G.M., Hauge, S., de Cosío, F.G., de Moltó, Y., Castillo, L.M., Cabello, M.A., von Horoch, M., Medina Osis, J., Machado, A., Nunes, B., Rodrigues, A.P., Rodrigues, E., Calomfirescu, C., Lupulescu, E., Popescu, R., Popovici, O., Bogdanovic, D., Kostic, M., Lazarevic, K., Milosevic, Z., Tiodorovic, B., Chen, M., Cutter, J., Lee, V., Lin, R., Ma, S., Cohen, A.L., Treurnicht, F., Kim, W.J., Delgado-Sanz, C., de mateo Ontañón, S., Larrauri, A., León, I.L., Vallejo, F., Born, R., Junker, C., Koch, D., Chuang, J.H., Huang, W.T., Kuo, H.W., Tsai, Y.C., Bundhamcharoen, K., Chittaganpitch, M., Green, H.K., Pebody, R., Goñi, N., Chiparelli, H., Brammer, L., Mustaquim, D., 2018. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet 391, 1285–1300. doi:10.1016/S0140-6736(17)33293-2

Jamieson, D.J., Theiler, R.N., Rasmussen, S.A., 2006. Emerging Infections and Pregnancy. Emerg. Infect. Dis. 12, 1638–1643.

Johnson, J., Brown, M., 2015. National Child Development Study: User Guide to the Response and Deaths Datasets.

Jones, A.M., Rice, N., Rosa Dias, P., 2011. Long-Term Effects of School Quality on Health and Lifestyle: Evidence from Comprehensive Schooling Reforms in England. J. Hum. Cap. 5, 342–376. doi:10.1086/662441

Kelly, E., 2011. The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children. J. Hum. Resour.

Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Ustün, T.B., 2007. Age of onset of mental disorders: a review of recent literature. Curr. Opin. Psychiatry 20, 359–364.

Klein, J.M., Gonçalves, A., Silva, C.F., 2009. The Rutter Children Behaviour Questionnaire for teachers: from psychometrics to norms, estimating caseness. Psico-USF (Impresso) 14, 157–165. doi:10.1590/S1413-82712009000200004

Knight, M., Kenyon, S., Brocklehurst, P., Neilson, J., Shakespeare, J., Kurinczuk, J., 2014. Saving Lives, Improving Mothers' Care - Lessons learned to inform future maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-12. Oxford. doi:10.1021/cn400230x

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76, 1071–1102. doi:10.1111/j.1467-937X.2009.00536.x

Lin, M.J., Liu, E.M., 2014. Does in utero exposure to illness matter? The 1918 influenza epidemic in taiwan as a natural experiment. J. Health Econ. 37, 152–163.

Lindeboom, M., Llena-Nozal, A., van der Klaauw, B., 2009. Parental education and child health: evidence from a schooling reform. J. Health Econ. 28, 109–31. doi:10.1016/j.jhealeco.2008.08.003

Lindeboom, M., Portrait, F., van den Berg, G.J., 2010. Long-run effects on longevity of a nutritional shock early in life: The Dutch Potato famine of 1846-1847. J. Health Econ. 29, 617–629. doi:10.1016/j.jhealeco.2010.06.001

Liu, S., Wang, J., Yang, X., Chen, J., Huang, R., Ruan, B., He, H., Wang, C., Zhang, H., Sun, Z., Xie, L., Zhuang, H., 2012. Pandemic influenza A(H1N1) 2009 virus in pregnancy. Rev. Med. Virol. 1957. doi:10.1002/rmv.1712

Mak, T.K., Mangtani, P., Leese, J., Watson, J.M., Pfeifer, D., 2008. Influenza vaccination in pregnancy: current evidence and selected national policies. Lancet Infect. Dis. 8, 44–52. doi:10.1016/S1473-3099(07)70311-0

Manski, C.F., 1990. Nonparametric Bounds on Treatment Effects. Am. Econ. Rev. 80, 829–823. doi:10.2307/2006592

Maselko, J., Sikander, S., Bhalotra, S., Bangash, O., Ganga, N., Mukherjee, S., Egger, H., Franz, L., Bibi, A., Liaqat, R., Kanwal, M., Abbasi, T., Noor, M., Ameen, N., Rahman, A., 2015. Effect of an early perinatal depression intervention on long-term child development outcomes: follow-up of the Thinking Healthy Programme randomised controlled trial. The Lancet Psychiatry 0366, 1–9. doi:10.1016/S2215-0366(15)00109-1

Mazumder, B., Almond, D., Park, K., Crimmins, E.M., Finch, C.E., 2010. Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. J. Dev. Orig. Health Dis.

McGee, R., Williams, S., Silva, P.A., 1986. An evaluation of the Malaise inventory. J. Psychosom. Res. 30, 147–152. doi:10.1016/0022-3999(86)90044-9

McManus, S., Gunnell, D., Cooper, C., Bebbington, P.E., Howard, L.M., Brugha, T., Jenkins, R., Hassiotis, A., Weich, S., Appleby, L., 2019. Prevalence of non-suicidal self-harm and service contact in England, 2000–14: repeated cross-sectional surveys of the general population. The Lancet Psychiatry 6, 573–581. doi:10.1016/S2215-0366(19)30188-9

Ministry of Health, 1960. Reports on Public Health and Medical Subjects. The influenza pandemic in England and Wales 1957–58. London.

Neelsen, S., Stratmann, T., 2012. Long-run effects of fetal influenza exposure: Evidence from Switzerland. Soc. Sci. Med. 74, 58–66.

Nelson, R.E., 2010. Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic. Health Econ. 19, 1181–1192.

NHS, 2015. Flu, your pregnancy an you [WWW Document]. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448958/PHE_8879_Flu_Pregnancy_2014_DL__A5_leaflet_05_.pdf

OECD, 2018. Health at a Glance: Europe 2018. doi:10.1787/health_glance_eur-2018-en

Parsons, A.G., Zhou, S.J., Spurrier, N.J., Makrides, M., 2008. Effect of iron supplementation during pregnancy on the behaviour of children at early school age: long-term follow-up of a randomised controlled trial. Br J Nutr 99, 1133–1139. doi:S0007114507853359 [pii]\n10.1017/S0007114507853359 [doi]

Pei, Z., Pischke, J.S., Schwandt, H., 2019. Poorly Measured Confounders are More Useful on the Left than on the Right. J. Bus. Econ. Stat. 37, 205–216. doi:10.1080/07350015.2018.1462710

Percoco, M., 2016. Health Shocks and Human Capital Accumulation: The Case of Spanish Flu in Italian Regions. Reg. Stud. 50, 1496–1508. doi:10.1080/00343404.2015.1039975

Persson, P., Rossin-Slater, M., 2018. Family ruptures, stress, and the mental health of the next generation. Am. Econ. Rev. 108, 1214–1252. doi:10.1257/aer.20161124

Pierce, M., Kurinczuk, J.J., Spark, P., Brocklehurst, P., Knight, M., 2011. Perinatal outcomes after maternal 2009/H1N1 infection: national cohort study. BMJ 342, d3214. doi:10.1136/bmj.d3214

Power, C., Elliott, J., 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). Int. J. Epidemiol. 35, 34–41. doi:10.1093/ije/dyi183

Public Health England, 2019. Surveillance of influenza and other respiratory viruses in the UK Winter 2018 to 2019. London.

Public Health England, 2015. The national flu immunisation programme 2015/16.

Rasmussen, S.A., Jamieson, D.J., Bresee, J.S., 2008. Pandemic influenza and pregnant women. Emerg. Infect. Dis.

Registrar General for England and Wales, 1957. The Registrar General's Weekly Return for England and Wales. Births and Deaths, Infectious Diseases, and Weather. London.

Registrar General for Scotland, 1957. Weekly Return of the Registrar General Scotland. Births, Deaths, and Marriages. Infectious Diseases. Weather report. Edinburgh.

Rosenbaum, P., 1987. Model-based direct adjustment. J. Am. Stat. Assoc. 82, 387–394. doi:10.1080/01621459.1987.10478441

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika

70, 41–55.

Rubin, D.B., 2001. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Heal. Serv. Outcomes Res. Methodol. 2, 169–188. doi:10.1017/cbo9780511810725.030

Rutter, M., Tizard, J., Whitmore, K., 1970. Education, health and behaviour. Longman, London.

Sadler, K., Vizard, T., Ford, T., Goodman, A., Goodman, R., McManus, S., 2018. The Mental Health of Children and Young People in England 2017: Trends and characteristics. London.

Schlotz, W., Phillips, D.I.W., 2009. Fetal origins of mental health: Evidence and mechanisms. Brain. Behav. Immun. 23, 905–916.

Scholte, R.S., van den Berg, G.J., Lindeboom, M., 2015. Long-run effects of gestation during the dutch hunger winter famine on labor market and hospitalization outcomes. J. Health Econ. 39, 17–30. doi:10.1016/j.jhealeco.2014.10.002

Schwandt, H., 2017. The Lasting Legacy of Seasonal Influenza: In-utero Exposure and Labor Market Outcomes (No. 5/2017), COHERE discussion paper. doi:10.1017/CBO9781107415324.004

Sprahue, A., 1988. Post-war fertility and female labour force participation rates. Econ. J. 98, 682–700.

Stuart, E., 2010. Matching methods for causal inference: A review and a look forward. Stat. Sci. 25, 1–21. doi:10.1214/09-STS313.Matching

Tomkins, A., Murray, S., Rondo, P., Filteau, S., 1994. Impact of maternal infection on foetal growth and nutrition. SCN News 11, 18–20.

van den Berg, G.J., Pinger, P.R., Schoch, J., 2015. Instrumental Variable Estimation of the Causal Effect of Hunger Early in Life on Health Later in Life. Econ. J. n/a-n/a. doi:10.1111/ecoj.12250

Vynnycky, E., Edmunds, W.J., 2008. Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol. Infect. 136, 166–179. doi:10.1017/S0950268807008369

Woodall, J., Rowson, K., McDonald, J., 1958. Age and Asian influenza, 1957. Br. Med. J. 2, 1316–1318.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd ed. MIT Press, Cambridge, MA. doi:10.1016/B978-0-323-05712-7.00031-3

Wooldridge, J.M., 2007. Inverse probability weighted estimation for general missing data problems. J. Econom. 141, 1281–1301. doi:10.1016/j.jeconom.2007.02.002

World Health Organization, 2014. Evaluation of seasonal influenza vaccination policies and coverage in the WHO European Region: Results from the 2008/2009 and 2009/2010 influenza seasons. Copenhagen.

World Health Organization, 2013. Mental Health Action Plan 2013-2020 [WWW Document]. WHO Libr. Cat. DataLibrary Cat. Data. URL http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf

Supplementary appendices

Appendix A1: Item-level missing data for the Rutter index at ages 7 and 11

Table A1: Number of completed Rutter items at age 7 and age 11, conditional on mother completing at least one item

Age 7			Age 11		
Items completed	Frequency	Percentage	Items completed	Frequency	Percentage
1	1	0.01	1	0	<0.01
2	3	0.02	2	1	0.01
3	5	0.03	3	3	0.02
4	3	0.02	4	1	0.01
5	4	0.03	5	2	0.01
6	2	0.01	6	0	<0.01
7	3	0.02	7	2	0.01
8	5	0.03	8	4	0.03
9	2	0.01	9	6	0.04
10	13	0.09	10	10	0.07
11	23	0.16	11	45	0.33
12	111	0.76	12	184	1.33
13	782	5.35	13	1,139	8.25
14	13,651	93.45	14	12,408	89.88

All figures are to 2 decimal places

Appendix A2: Distributions of the mental health outcomes in childhood and adulthood

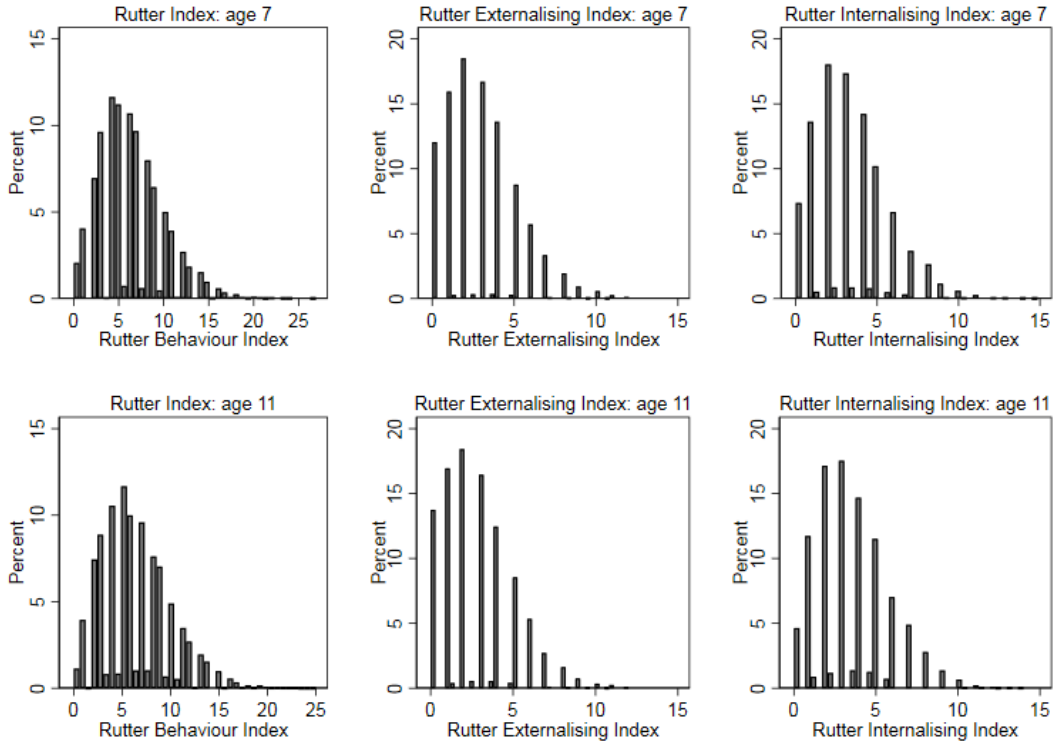


Figure A1: Distributions of the Rutter Behaviour Indices at ages 7 and 11

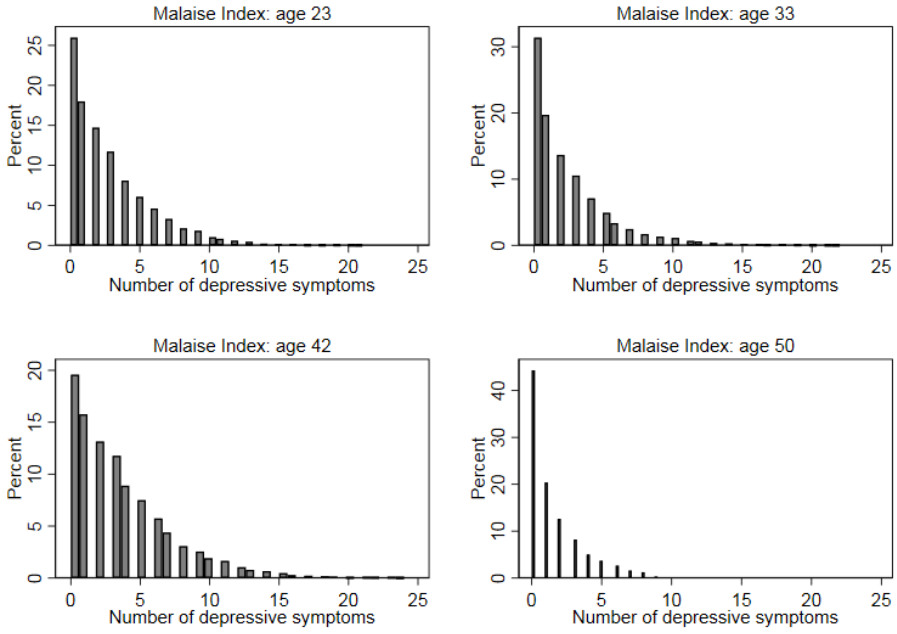


Figure A2: Distributions of the Malaise Index at ages 23, 33, 42 and 50. Derived from 24-item Malaise Inventory at ages 23, 33 and 42, and from the 9-item inventory at age 50.

Appendix A3: Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

Table A2: Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

	Not exposed to influenza *in-utero* (86.8%)	**Exposed to influenza *in-utero* *(13.2%)***
	N (%) or Mean (SD)	**N (%) or Mean (SD)**
Male gender	5490 (51.45%)	818 (50.49%)
Ethnic minority	116 (1.09%)	14 (0.86%)
Mother's occupational social class		
Unemployed	6592 (61.77%)	974 (60.12%)
I or II: Professional, managerial or technical	485 (4.55%)	77 (4.75%)
III: Skilled manual/non-manual	2314 (21.68%)	358 (22.10%)
IV: Partly skilled	1014 (9.50%)	169 (10.43%)
V: Unskilled	266 (2.49%)	42 (2.59%)
Father's occupational social class		
I: Professional	445 (4.17%)	52 (3.21%)
II: Managerial or technical	1420 (13.31%)	164 (10.12%)
III: Skilled manual/non-manual	6206 (58.16%)	1003 (61.91%)
IV: Partly skilled	1235 (11.57%)	192 (11.85%)
V: Unskilled	943 (8.84%)	145 (8.95%)
Information missing	422 (3.95%)	64 (3.95%)
Mother schooled past leaving age	2723 (25.52%)	386 (23.83%)
Mother's age in years		
20 or under	982 (9.20%)	164 (10.12%)
21-25	3288 (30.81%)	542 (33.46%)
26-30	3440 (32.24%)	471 (29.07%)
31-35	1881 (17.63%)	289 (17.84%)
36-40	873 (8.18%)	130 (8.02%)
Over 40	207 (1.94%)	24 (1.48%)
Mother's height (inches)	63.43 (2.49)	63.49 (2.56)
Father's age in years		
20 or under	208 (1.95%)	25 (1.54%)
21-25	2111 (19.78%)	354 (21.85%)
26-30	3454 (32.37%)	522 (32.22%)
31-35	2543 (23.83%)	371 (22.90%)
36-40	1310 (12.28%)	192 (11.85%)
41-45	517 (4.84%)	90 (5.56%)
45-50	178 (1.67%)	22 (1.36%)
Over 50	72 (0.67%)	3 (0.19%)
Information missing	278 (2.61%)	41 (2.53%)
Mother's smoking prior to pregnancy		
Non-smoker	6318 (59.21%)	969 (59.81%)
1-4 cigarettes per day	820 (7.68%)	135 (8.33%)
5-9 cigarettes per day	1494 (14.00%)	220 (13.58%)
10-14 cigarettes per day	1459 (13.67%)	213 (13.15%)
15-19 cigarettes per day	306 (2.87%)	39 (2.41%)
20+ cigarettes per day	274 (2.57%)	44 (2.72%)
Weight prior to pregnancy based on body mass index (BMI)		
Normal/healthy [18.5<BMI<25]	7787 (72.97%)	1170 (72.22%)
Underweight [BMI<18.5]	403 (3.78%)	41 (2.53%)
Overweight [25<BMI<30]	2028 (19.00%)	327 (20.19%)
Obese [BMI>30]	453 (4.25%)	82 (5.06%)
Antenatal visits within 7 weeks of pregnancy	432 (4.05%)	69 (4.26%)
Any co-morbidity†	730 (6.84%)	105 (6.48%)
Previous abortion	1269 (11.89%)	191 (11.79%)
Previous premature birth (<5lb,9oz)	855 (8.01%)	148 (9.14%)
Previous large birth (>8lb,14oz))	965 (9.04%)	177 (10.93%)
Previous stillbirth/neonatal death	549 (5.14%)	79 (4.88%)
Complication in previous pregnancy	904 (8.47%)	155 (9.57%)

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.
†Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy

Table A2 (continued): Unweighted summary statistics for individuals exposed and not exposed to influenza *in-utero*

	Not exposed to influenza *in-utero* (86.8%)	**Exposed to influenza *in-utero* *(13.2%)***
	N (%) or Mean (SD)	**N (%) or Mean (SD)**
Persons per room		
<1	7397 (69.32%)	1119 (69.07%)
1-1.5	1851 (17.35%)	275 (16.98%)
1.5-2	969 (9.08%)	148 (9.14%)
>2	454 (4.25%)	78 (4.81%)
Number of older siblings		
No older siblings	4061 (38.06%)	577 (35.62%)
1	3468 (32.50%)	514 (31.73%)
2	1575 (14.76%)	263 (16.23%)
3	791 (7.41%)	116 (7.16%)
4	360 (3.37%)	65 (4.01%)
5	182 (1.71%)	45 (2.78%)
6+	234 (2.19%)	40 (2.47%)
Stillbirth rate (Still births/total births)	0.02 (0.00)	0.02 (0.01)
Proportion of male working population unskilled	0.14 (0.04)	0.14 (0.04)
Proportion of men leaving school aged 16 or older	0.13 (0.05)	0.13 (0.05)
Population density (population in thousands/square km)	2.40 (2.45)	2.27 (2.35)
Pneumonia Notifications Per Thousand Population	12.61 (17.07)	12.00 (15.53)
Region of birth (Base: Northern England)		
North West England	765 (7.17%)	114 (7.04%)
East & West Riding, England	1437 (13.47%)	183 (11.3%)
North Midlands, England	899 (8.42%)	147 (9.07%)
Midlands, England	776 (7.27%)	138 (8.52%)
East England	975 (9.14%)	198 (12.22%)
South East England	901 (8.44%)	146 (9.01%)
South England	1913 (17.93%)	243 (15%)
South West England	587 (5.5%)	96 (5.93%)
Wales	585 (5.48%)	85 (5.25%)
Scotland	541 (5.07%)	100 (6.17%)

Estimated on the largest outcome-specific estimation sample (Rutter indices at age 7). Sample size=12,291.

Appendix A4: Derivation of the estimation sample for each outcome

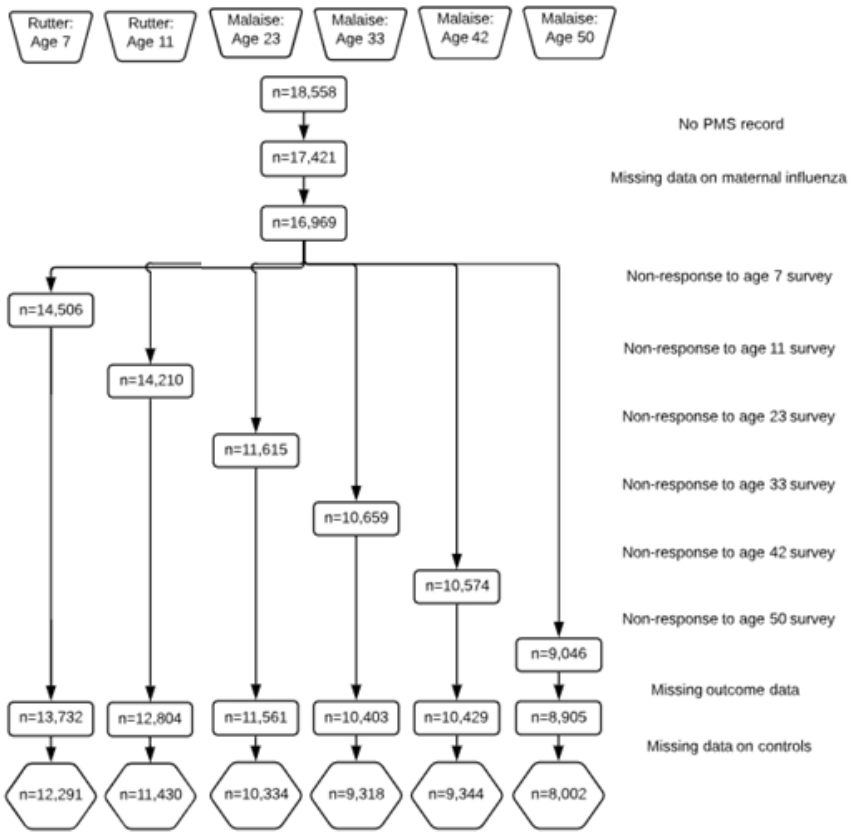


Figure A3: Derivation of the estimation sample for each outcome

Appendix A5: Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

Table A3: Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

	Age 11		Age 23		Age 33		Age 42		Age 50	
	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted
Male gender	-0.020	0.006	0.008	-0.001	0.015	-0.003	0.003	0.000	-0.002	-0.005
Ethnic minority	0.001	-0.011	-0.002	-0.012	-0.036	-0.013	-0.013	-0.018	-0.055	-0.009
Mother's occupational social class (Base: V: Unskilled)										
Unemployed	-0.030	0.003	-0.004	-0.003	-0.032	-0.001	-0.039	-0.002	-0.055	-0.006
I or II: Professional, managerial or technical	0.018	-0.008	0.007	-0.013	0.013	-0.004	0.021	-0.007	0.021	-0.008
III: Skilled manual/non-manual	0.000	0.000	0.001	0.005	0.029	-0.001	0.014	0.003	0.018	0.006
IV: Partly skilled	0.034	0.001	-0.005	0.008	-0.002	0.006	0.032	0.003	0.034	0.006
Father's occupational social class (Base: V: Unskilled)										
I: Professional	-0.056	0.010	-0.013	0.004	-0.024	0.012	-0.012	0.018	-0.021	0.010
II: Managerial or technical	-0.108	0.000	-0.095	-0.009	-0.092	-0.010	-0.104	-0.004	-0.112	-0.004
III: Skilled manual/non-manual	0.091	-0.003	0.074	-0.009	0.075	0.001	0.070	-0.006	0.081	-0.007
IV: Partly skilled	0.005	0.000	-0.004	0.008	0.008	-0.001	0.001	0.001	0.020	0.006
Information missing	-0.022	-0.004	-0.004	-0.001	-0.029	-0.008	-0.001	-0.004	-0.014	-0.003
Mother schooled past leaving age	-0.048	0.004	-0.003	0.004	-0.054	-0.002	-0.047	0.008	-0.049	-0.005
Mother's age in years (Base: 20 or under)										
21-25	0.078	-0.001	0.064	0.007	0.040	0.005	0.061	0.004	0.062	-0.003
26-30	-0.077	0.011	-0.082	0.003	-0.103	0.000	-0.067	0.002	-0.067	0.008
31-35	-0.020	-0.010	-0.009	-0.009	0.035	-0.010	0.003	-0.010	-0.003	-0.006
36-40	-0.009	0.008	0.017	0.001	0.014	0.000	-0.027	0.005	-0.025	0.007
Over 40	-0.036	-0.014	-0.005	-0.005	-0.014	0.000	-0.022	0.001	-0.023	-0.003
Mother's height (inches)	0.022	-0.006	0.023	-0.007	-0.002	-0.002	0.016	0.004	0.003	-0.003
Father's age in years (Base: 20 or under)										
21-25	0.071	0.000	0.063	0.000	0.045	-0.001	0.040	0.003	0.034	-0.003
26-30	0.012	-0.002	-0.001	0.003	0.003	0.005	0.015	-0.003	0.028	-0.003
31-35	-0.044	0.001	-0.044	0.003	-0.041	0.003	-0.028	0.005	-0.039	0.000
36-40	-0.030	0.000	-0.016	-0.003	-0.004	0.000	-0.011	-0.004	-0.017	0.002
41-45	0.018	0.003	0.048	-0.001	0.035	0.000	0.018	0.004	0.025	0.004
45-50	-0.030	0.000	-0.019	-0.014	-0.035	-0.014	-0.057	0.001	-0.027	0.000
Over 50	-0.039	-0.005	-0.051	-0.003	-0.060	-0.020	-0.058	-0.019	-0.050	-0.018
Information missing	-0.001	-0.001	-0.011	0.004	0.014	0.000	-0.006	0.007	-0.007	0.013
Mother's smoking prior to pregnancy (Base: Non-smoker)										
1-4 cigarettes per day	0.020	0.002	0.041	0.003	0.027	0.007	0.032	0.005	0.076	0.003
5-9 cigarettes per day	0.000	0.001	-0.011	0.002	0.003	0.005	0.006	-0.005	0.005	-0.002
10-14 cigarettes per day	-0.020	-0.004	-0.023	-0.006	0.003	-0.007	-0.005	-0.005	0.020	-0.004
15-19 cigarettes per day	-0.047	0.011	0.009	0.010	-0.013	0.008	-0.027	0.018	-0.028	0.021
20+ cigarettes per day	0.006	0.004	0.011	-0.001	0.026	-0.009	0.026	-0.013	0.043	-0.010
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])										
Underweight [BMI<18.5]	-0.073	0.011	-0.066	0.004	-0.079	0.000	-0.048	0.002	-0.053	-0.002
Overweight [25<BMI<30]	0.048	0.007	0.063	0.004	0.052	0.008	0.029	0.004	0.030	0.006
Obese [BMI>30]	0.018	-0.004	-0.004	0.001	0.012	0.008	0.010	0.005	-0.016	0.008
Antenatal visits within 7 weeks of pregnancy	0.017	0.005	0.009	0.007	0.024	0.007	0.020	0.004	0.016	0.001
Any comorbidity[a]	-0.033	0.001	-0.034	0.002	-0.020	-0.004	-0.025	-0.010	-0.022	-0.004
Previous abortion	-0.006	0.002	0.010	-0.008	0.022	-0.008	0.024	-0.010	0.018	-0.007
Previous premature birth (<5lb,9oz)	0.048	-0.003	0.051	0.002	0.053	0.006	0.045	-0.003	0.031	0.000
Previous large birth (>8lb,14oz))	0.068	-0.006	0.078	-0.003	0.086	0.001	0.053	0.000	0.083	-0.005
Previous stillbirth/neonatal death	-0.008	-0.002	0.019	0.008	0.026	-0.003	0.024	-0.003	0.019	-0.001
Complication in previous pregnancy	0.024	0.004	0.053	0.004	0.053	0.009	0.053	0.005	0.019	0.010

Table A3 (continued): Standardised differences in the covariate before and after weighting for estimation samples at ages 11, 23, 33, 42, and 50

	Age 11		Age 23		Age 33		Age 42		Age 50	
	Raw	**Weighted**	**Raw**	**Weighted**	**Raw**	**Weighted**	**Raw**	**Weighted**	**Raw**	**Weighted**
Persons per room (Base: <1)										
1-1.5	0.003	-0.001	-0.007	-0.004	-0.020	-0.007	-0.002	-0.008	0.004	0.000
1.5-2	0.001	-0.003	0.005	-0.002	0.043	0.000	0.016	-0.005	0.045	-0.006
>2	0.018	0.001	0.051	0.003	0.024	0.007	0.057	0.001	0.079	0.005
Number of older siblings (Base: No siblings)										
1	-0.017	-0.006	-0.045	-0.002	-0.055	-0.009	-0.030	-0.006	-0.065	-0.015
2	0.025	0.012	0.061	0.008	0.053	0.007	0.066	0.005	0.025	0.015
3	0.007	0.007	0.019	0.011	0.007	-0.001	-0.005	0.000	0.016	-0.001
4	0.034	-0.004	0.049	-0.003	0.060	-0.003	0.029	0.002	0.071	0.003
5	0.057	0.002	0.077	-0.004	0.061	0.008	0.078	0.001	0.100	-0.001
6+	0.019	-0.006	0.046	-0.001	0.066	-0.003	0.033	-0.001	0.032	0.005
Stillbirth rate (Still births/total births)	0.058	0.002	0.069	0.003	0.068	0.007	0.069	0.002	0.076	-0.002
Proportion of male working population unskilled	-0.070	-0.002	-0.060	-0.006	-0.030	0.001	-0.041	-0.001	-0.021	0.001
Proportion of men leaving school aged 16 or older	-0.037	0.004	-0.049	-0.001	-0.061	-0.007	-0.046	-0.006	-0.024	-0.002
Population density (population in thousands/square km)	-0.072	-0.001	-0.054	-0.008	-0.041	-0.011	-0.042	-0.007	-0.031	-0.009
Pneumonia Notifications Per Thousand Population	-0.067	0.011	-0.060	0.009	-0.011	0.014	-0.032	0.010	-0.057	0.015
Region of birth (Base: Northern England)										
North West England	-0.062	-0.004	-0.052	-0.002	-0.066	0.002	-0.074	-0.008	-0.047	-0.005
East & West Riding, England	0.017	-0.006	0.040	-0.004	0.023	-0.003	0.042	-0.004	0.016	-0.005
North Midlands, England	0.065	-0.003	0.079	0.000	0.071	0.008	0.066	0.011	0.075	0.008
Midlands, England	0.093	0.002	0.102	0.009	0.086	0.003	0.058	0.007	0.047	0.006
East England	0.015	0.002	-0.029	0.000	0.007	0.001	0.002	0.001	0.014	0.002
South East England	-0.073	-0.001	-0.081	-0.007	-0.091	-0.020	-0.054	-0.015	-0.079	-0.013
South England	0.034	-0.009	0.030	-0.007	0.031	-0.002	0.016	-0.001	0.051	0.000
South West England	-0.005	-0.003	-0.028	-0.004	-0.027	-0.001	-0.043	0.006	-0.033	0.002
Wales	0.051	0.000	0.055	-0.004	0.055	0.003	0.061	0.002	0.054	-0.003
Scotland	-0.076	0.016	-0.082	0.019	-0.047	0.021	-0.066	0.014	-0.059	0.024

Appendix A6: Probit marginal effects from the propensity score model

Table A4: Marginal effects of probit models of *in-utero* influenza exposure.

	Age 7		Age 11		Age 23		Age 33		Age 42		Age 50	
	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err
Male gender	-0.0051	(0.0061)	-0.0053	(0.0063)	0.0013	(0.0066)	0.0031	(0.0069)	0.0001	(0.0069)	-0.0001	(0.0074)
Ethnic minority	-0.0340	(0.0312)	-0.0055	(0.0297)	-0.0097	(0.0339)	-0.0569	(0.0418)	-0.0299	(0.0365)	-0.0832	(0.0430)
Mother's occupational social class (Base: V: Unskilled)												
Unemployed	-0.0153	(0.0203)	-0.0088	(0.0203)	-0.0129	(0.0221)	-0.0105	(0.0237)	-0.0055	(0.0231)	-0.0274	(0.0268)
I or II: Professional, managerial or technical	0.0183	(0.0260)	0.0277	(0.0263)	0.0121	(0.0278)	0.0289	(0.0299)	0.0398	(0.0297)	0.0147	(0.0332)
III: Skilled manual/non-manual	0.0078	(0.0213)	0.0083	(0.0214)	0.0057	(0.0232)	0.0155	(0.0247)	0.0222	(0.0242)	-0.0040	(0.0277)
IV: Partly skilled	0.0118	(0.0226)	0.0164	(0.0228)	0.0001	(0.0245)	0.0032	(0.0260)	0.0244	(0.0257)	-0.0001	(0.0292)
Father's occupational social class (Base: V: Unskilled)												
I: Professional	-0.0195	(0.0185)	-0.0238	(0.0190)	0.0042	(0.0205)	0.0074	(0.0219)	0.0137	(0.0221)	0.0162	(0.0228)
II: Managerial or technical	-0.0246	(0.0137)	-0.0292*	(0.0143)	-0.0199	(0.0150)	-0.0153	(0.0159)	-0.0172	(0.0155)	-0.0085	(0.0165)
III: Skilled manual/non-manual	0.0066	(0.0112)	0.0050	(0.0118)	0.0113	(0.0124)	0.0100	(0.0132)	0.0093	(0.0128)	0.0190	(0.0136)
IV: Partly skilled	-0.0025	(0.0135)	-0.0077	(0.0141)	-0.0015	(0.0148)	0.0000	(0.0158)	-0.0026	(0.0154)	0.0106	(0.0166)
Information missing	0.0006	(0.0244)	-0.0296	(0.0228)	0.0032	(0.0263)	-0.0397	(0.0237)	0.0017	(0.0284)	-0.0052	(0.0292)
Mother schooled past leaving age	0.0022	(0.0077)	0.0004	(0.0079)	0.0131	(0.0083)	-0.0025	(0.0086)	-0.0024	(0.0086)	0.0001	(0.0091)
Mother's age in years (Base: 20 or under)												
21-25	-0.0142	(0.0137)	-0.0145	(0.0141)	-0.0118	(0.0151)	-0.0228	(0.0158)	-0.0164	(0.0159)	-0.0199	(0.0174)
26-30	-0.0385*	(0.0152)	-0.0423**	(0.0157)	-0.0400*	(0.0168)	-0.0481**	(0.0176)	-0.0428*	(0.0177)	-0.0480*	(0.0194)
31-35	-0.0341*	(0.0173)	-0.0408*	(0.0178)	-0.0357	(0.0190)	-0.0287	(0.0203)	-0.0389	(0.0200)	-0.0458*	(0.0220)
36-40	-0.0461*	(0.0199)	-0.0480*	(0.0206)	-0.0385	(0.0223)	-0.0415	(0.0235)	-0.0577*	(0.0226)	-0.0655**	(0.0244)
Over 40	-0.0728**	(0.0252)	-0.0774**	(0.0255)	-0.0548	(0.0299)	-0.0569	(0.0309)	-0.0630*	(0.0305)	-0.0822**	(0.0298)
Mother's height (inches)	0.0019	(0.0013)	0.0019	(0.0013)	0.0015	(0.0014)	0.0007	(0.0015)	0.0015	(0.0015)	0.0006	(0.0016)
Father's age in years (Base: 20 or under)												
21-25	0.0391	(0.0205)	0.0382	(0.0213)	0.0422	(0.0239)	0.0389	(0.0240)	0.0235	(0.0250)	0.0330	(0.0254)
26-30	0.0360	(0.0211)	0.0336	(0.0219)	0.0337	(0.0245)	0.0378	(0.0247)	0.0245	(0.0259)	0.0427	(0.0263)
31-35	0.0304	(0.0221)	0.0232	(0.0229)	0.0212	(0.0255)	0.0205	(0.0257)	0.0151	(0.0269)	0.0303	(0.0273)
36-40	0.0293	(0.0236)	0.0210	(0.0245)	0.0168	(0.0269)	0.0175	(0.0273)	0.0144	(0.0287)	0.0316	(0.0293)
41-45	0.0544	(0.0280)	0.0425	(0.0287)	0.0438	(0.0311)	0.0315	(0.0308)	0.0282	(0.0326)	0.0507	(0.0338)
45-50	0.0157	(0.0321)	0.0107	(0.0332)	0.0024	(0.0349)	-0.0095	(0.0339)	-0.0235	(0.0354)	0.0165	(0.0391)
Over 50	-0.0558	(0.0316)	-0.0181	(0.0394)	-0.0448	(0.0383)	-0.0564	(0.0350)	-0.0586	(0.0384)	-0.0288	(0.0429)
Information missing	0.0221	(0.0324)	0.0488	(0.0375)	0.0209	(0.0368)	0.0817	(0.0457)	0.0112	(0.0392)	0.0345	(0.0432)
Mother's smoking prior to pregnancy (Base: Non-smoker)												
1-4 cigarettes per day	0.0081	(0.0119)	0.0056	(0.0123)	0.0166	(0.0132)	0.0121	(0.0135)	0.0131	(0.0136)	0.0358*	(0.0155)
5-9 cigarettes per day	-0.0040	(0.0090)	0.0001	(0.0095)	-0.0007	(0.0100)	0.0054	(0.0106)	0.0027	(0.0104)	0.0069	(0.0113)
10-14 cigarettes per day	-0.0056	(0.0091)	-0.0062	(0.0094)	-0.0043	(0.0099)	0.0044	(0.0105)	-0.0004	(0.0104)	0.0091	(0.0113)
15-19 cigarettes per day	-0.0152	(0.0180)	-0.0280	(0.0182)	0.0129	(0.0216)	-0.0016	(0.0216)	-0.0134	(0.0206)	-0.0074	(0.0225)
20+ cigarettes per day	0.0091	(0.0200)	0.0059	(0.0204)	0.0114	(0.0221)	0.0267	(0.0236)	0.0202	(0.0227)	0.0361	(0.0255)
Observations	12,291		11,430		10,344		9,318		9,344		8,002	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects.

Table A4 (continued): Marginal effects of probit models of *in-utero* influenza exposure

	Age 7		Age 11		Age 23		Age 33		Age 42		Age 50	
	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err	Margins	Std. Err
Weight prior to pregnancy based on body mass index (BMI) (Base: Normal/healthy [18.5<BMI<25])												
Underweight [BMI<18.5]	-0.0397**	(0.0141)	-0.0400**	(0.0144)	-0.0367*	(0.0160)	-0.0415*	(0.0162)	-0.0284	(0.0168)	-0.0308	(0.0178)
Overweight [25<BMI<30]	0.0085	(0.0080)	0.0139	(0.0084)	0.0176*	(0.0089)	0.0132	(0.0093)	0.0085	(0.0092)	0.0066	(0.0098)
Obese [BMI>30]	0.0206	(0.0161)	0.0088	(0.0162)	-0.0024	(0.0165)	0.0027	(0.0179)	0.0036	(0.0178)	-0.0117	(0.0184)
Antenatal visits within 7 weeks of pregnancy	0.0122	(0.0152)	0.0152	(0.0154)	0.0119	(0.0166)	0.0199	(0.0170)	0.0181	(0.0170)	0.0164	(0.0183)
Any comorbidity[a]	-0.0036	(0.0122)	-0.0103	(0.0127)	-0.0126	(0.0135)	-0.0087	(0.0140)	-0.0110	(0.0140)	-0.0103	(0.0151)
Previous abortion	0.0015	(0.0096)	0.0015	(0.0099)	0.0031	(0.0103)	0.0075	(0.0107)	0.0087	(0.0107)	0.0068	(0.0115)
Previous premature birth (<5lb,9oz)	0.0117	(0.0119)	0.0169	(0.0124)	0.0070	(0.0130)	0.0071	(0.0137)	0.0043	(0.0136)	0.0005	(0.0149)
Previous large birth (>8lb,14oz))	0.0127	(0.0110)	0.0168	(0.0115)	0.0111	(0.0122)	0.0164	(0.0126)	0.0059	(0.0128)	0.0191	(0.0135)
Previous stillbirth/neonatal death	-0.0129	(0.0145)	-0.0073	(0.0153)	0.0049	(0.0157)	0.0058	(0.0162)	0.0074	(0.0165)	0.0070	(0.0178)
Complication in previous pregnancy	0.0125	(0.0110)	0.0059	(0.0115)	0.0154	(0.0120)	0.0178	(0.0128)	0.0175	(0.0126)	0.0066	(0.0140)
Persons per room (Base: <1)												
1-1.5	-0.0166	(0.0087)	-0.0115	(0.0091)	-0.0182	(0.0094)	-0.0217*	(0.0098)	-0.0146	(0.0099)	-0.0103	(0.0107)
1.5-2	-0.0136	(0.0116)	-0.0115	(0.0121)	-0.0136	(0.0129)	-0.0061	(0.0140)	-0.0079	(0.0138)	0.0058	(0.0157)
>2	0.0003	(0.0167)	-0.0012	(0.0171)	0.0172	(0.0201)	-0.0070	(0.0193)	0.0199	(0.0208)	0.0375	(0.0234)
Number of older siblings (Base: No siblings)												
1	0.0222*	(0.0086)	0.0211*	(0.0090)	0.0189*	(0.0093)	0.0170	(0.0097)	0.0250*	(0.0098)	0.0127	(0.0104)
2	0.0462***	(0.0120)	0.0415***	(0.0124)	0.0573***	(0.0133)	0.0498***	(0.0137)	0.0599***	(0.0139)	0.0401**	(0.0145)
3	0.0315*	(0.0153)	0.0380*	(0.0162)	0.0502**	(0.0176)	0.0401*	(0.0177)	0.0423*	(0.0176)	0.0409*	(0.0195)
4	0.0651**	(0.0228)	0.0655**	(0.0238)	0.0852***	(0.0257)	0.0890**	(0.0279)	0.0777**	(0.0276)	0.0959**	(0.0311)
5	0.1173***	(0.0324)	0.1003**	(0.0323)	0.1236***	(0.0348)	0.1133**	(0.0389)	0.1386***	(0.0384)	0.1401***	(0.0422)
6+	0.0577*	(0.0286)	0.0565	(0.0297)	0.0841*	(0.0340)	0.0995**	(0.0375)	0.0894*	(0.0384)	0.0646	(0.0393)
Area-level characteristics												
Stillbirth rate (Still births/total births)	1.6088*	(0.7599)	1.4399	(0.8179)	1.2514	(0.9054)	1.4086	(0.9399)	1.6924	(0.9256)	2.0726	(1.0576)
Proportion of male working population unskilled	-0.0310	(0.1431)	-0.0506	(0.1523)	-0.0662	(0.1573)	0.0436	(0.1665)	-0.0174	(0.1643)	0.1253	(0.1746)
Proportion of men leaving school aged 16 or older	-0.0665	(0.1066)	-0.0675	(0.1120)	-0.0273	(0.1147)	-0.0286	(0.1215)	-0.0580	(0.1182)	0.1214	(0.1255)
Population density (population in thousands/square km)	-0.0008	(0.0019)	-0.0012	(0.0020)	-0.0004	(0.0021)	-0.0011	(0.0022)	-0.0006	(0.0022)	-0.0001	(0.0023)
Pneumonia Notifications Per Thousand Population	-0.0001	(0.0003)	-0.0002	(0.0003)	-0.0003	(0.0003)	0.0001	(0.0003)	-0.0001	(0.0003)	-0.0005	(0.0004)
Region of birth (Base: Northern England)												
North West England	-0.0165	(0.0141)	-0.0143	(0.0147)	-0.0173	(0.0161)	-0.0213	(0.0165)	-0.0340*	(0.0167)	-0.0251	(0.0180)
East & West Riding, England	0.0131	(0.0155)	0.0111	(0.0160)	0.0147	(0.0176)	0.0083	(0.0178)	0.0067	(0.0184)	0.0085	(0.0201)
North Midlands, England	0.0170	(0.0162)	0.0258	(0.0171)	0.0247	(0.0182)	0.0279	(0.0192)	0.0128	(0.0192)	0.0288	(0.0209)
Midlands, England	0.0376*	(0.0156)	0.0358*	(0.0161)	0.0338	(0.0175)	0.0318	(0.0181)	0.0092	(0.0182)	0.0161	(0.0197)
East England	0.0178	(0.0170)	0.0159	(0.0174)	-0.0097	(0.0183)	0.0102	(0.0197)	-0.0005	(0.0198)	0.0061	(0.0211)
South East England	-0.0026	(0.0166)	0.0005	(0.0171)	-0.0149	(0.0185)	-0.0093	(0.0194)	-0.0090	(0.0200)	-0.0191	(0.0207)
South England	0.0192	(0.0195)	0.0277	(0.0205)	0.0139	(0.0216)	0.0250	(0.0227)	0.0070	(0.0230)	0.0219	(0.0248)
South West England	0.0050	(0.0183)	0.0070	(0.0187)	-0.0127	(0.0197)	-0.0062	(0.0205)	-0.0201	(0.0207)	-0.0143	(0.0222)
Wales	0.0283	(0.0188)	0.0275	(0.0190)	0.0232	(0.0205)	0.0262	(0.0213)	0.0196	(0.0217)	0.0152	(0.0230)
Scotland	-0.0041	(0.0156)	-0.0096	(0.0158)	-0.0163	(0.0173)	-0.0149	(0.0179)	-0.0274	(0.0178)	-0.0097	(0.0199)
Observations	12,291		11,430		10,344		9,318		9,344		8,002	

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001; coefficients represent average partial effects.

[a]Co-morbidities include diabetes, heart disease, tuberculosis, German measles, or psychiatric disorders during pregnancy

Appendix A7: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of mortality in childhood and adulthood by gender

Table A5: Gender-specific inverse probability weight estimates of the ATE of *in-utero* exposure to influenza on mental health at ages 7, 11, 23, 33, 42, and 50.

	Full sample		Males		Females	
	Observations	**ATE**	**Observations**	**ATE**	**Observations**	**ATE**
Rutter: age 7						
Overall	12,291	0.103 (0.095) [1.63%]	6,308	0.183 (0.130) [2.79%]	5.843	0.026 (0.137) [0.43%]
Internalising	12,291	0.076 (0.060) [2.29%]	6,273	0.124 (0.082) [3.83%]	5.983	0.038 (0.088) [1.12%]
Externalising	12,291	0.027 (0.058) [0.90%]	6,301	0.066 (0.083) [2.00%]	5.957	-0.019 (0.079) [-0.72%]
Rutter: age 11						
Overall	11,430	0.194* (0.098) [3.04%]	5,854	0.163 (0.133) [2.44%]	5.976	0.220 (0.141) [3.66%]
Internalising	11,430	0.131* (0.060) [3.68%]	5,806	0.012 (0.085) [0.34%]	5.576	0.275** (0.086) [7.72%]
Externalising	11,430	0.061 (0.058) [2.17%]	5,843	0.143 (0.080) [4.57%]	5.527	-0.050 (0.082) [-2.05%]
Malaise: age 23	10,344	0.054 (0.081) [1.99%]	5,149	0.043 (0.097) [2.14%]	5,195	0.065 (0.131) [1.93%]
Malaise: age 33	9,318	0.082 (0.089) [3.42%]	4,585	0.003 (0.112) [0.15%]	4,733	0.131 (0.132) [4.79%]
Malaise: age 42	9,344	0.046 (0.108) [1.39%]	4,573	0.013 (0.145) [0.42%]	4,771	0.093 (0.160) [2.32%]
Malaise: age 50	8,002	0.093 (0.066) [6.35%]	3,919	0.176* (0.089) [15.45%]	4,083	0.052 (0.101) [2.98%]

Robust standard errors in parenthesis rounded parenthesis; ATEs as a % of mean outcomes for the non-exposed group in squared parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50.

Appendix A8: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of mortality in childhood and adulthood

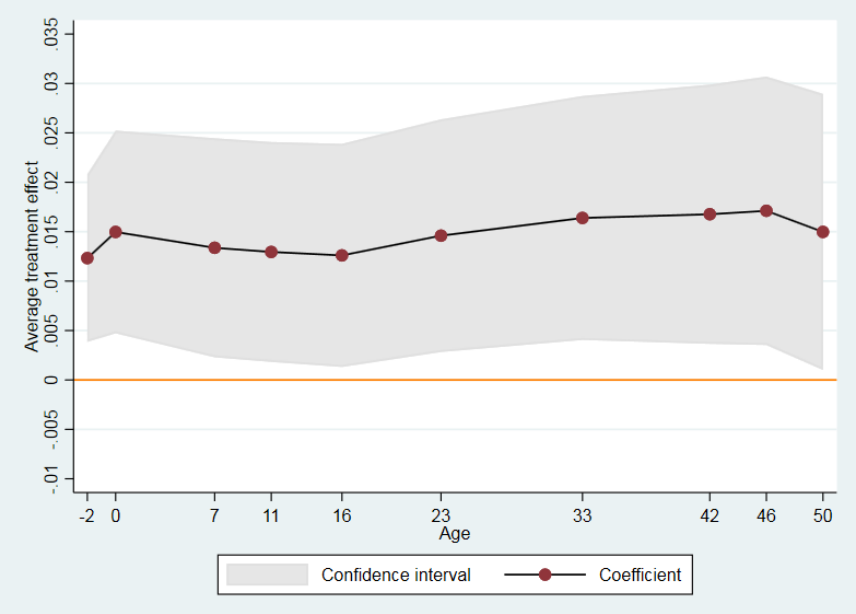


Figure A4: Inverse-probability-weight estimates of the ATE of *in-utero* exposure to influenza on the probability of death at birth, 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50. Ages -2 and 0 represent birth and age 28 days, respectively.

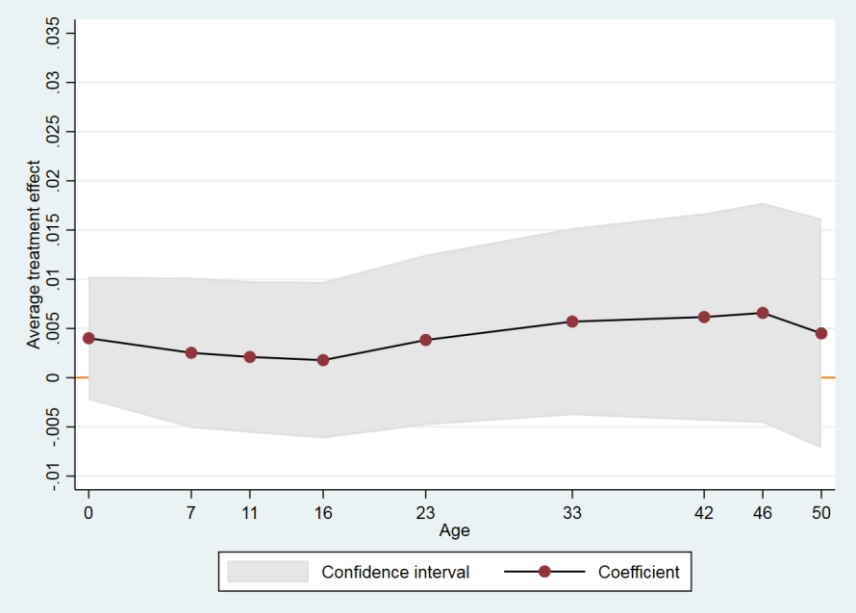


Figure A5: Inverse-probability-weight estimates of the ATE of *in-utero* exposure to influenza on the probability of death at 28 days, and by ages 7, 11, 16, 23, 33, 42, 46 and 50 after conditioning on survival beyond birth. Age 0 represents birth 28 days..

Appendix A9: ATE bounds under mortality-driven non-response

Table A6: ATE bounds under mortality-driven non-response

	Worst case bounds		Bounds replacing missing outcomes with 90th percentile	
	ATE	**ATE as % of non-exposed mean**	**ATE**	**ATE as % of non-exposed mean**
Rutter: age 7				
Overall	[0.103, 0.445]	[1.63%, 6.13%]	[0.103, 0.176]	[1.63%, 2.7%]
Internalising	[0.076, 0.284]	[2.29%, 7.28%]	[0.076, 0.118]	[2.29%, 3.41%]
Externalising	[0.027, 0.182]	[0.9%, 5.35%]	[0.027, 0.080]	[0.9%, 2.56%]
Rutter: age 11				
Overall	[0.194, 0.503]	[3.04%, 6.9%]	[0.194, 0.262]	[3.04%, 3.98%]
Internalising	[0.131, 0.348]	[3.68%, 8.37%]	[0.131, 0.192]	[3.68%, 5.13%]
Externalising	[0.061, 0.232]	[2.17%, 7.08%]	[0.061, 0.120]	[2.17%, 4.05%]
Malaise: age 23	[0.054, 0.384]	[1.99%, 10.01%]	[0.054, 0.128]	[1.99%, 4.33%]
Malaise: age 33	[0.082, 0.563]	[3.42%, 14.64%]	[0.082, 0.166]	[3.42%, 6.28%]
Malaise: age 42	[0.046, 0.563]	[1.29%, 10.6%]	[0.046, 0.180]	[1.29%, 4.48%]
Malaise: age 50	[0.093, 0.285]	[6.35%, 12.06%]	[0.093, 0.148]	[6.35%, 8.42%]

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Malaise symptom counts range from 0-24 at ages 23, 33, and 42, and from 0-9 at age 50. Upper bound or worst case bounds calculated by replacing missing outcome values due to death with the maximum of the outcome.

Appendix A10: IPTW estimates of the ATE of *in-utero* exposure to influenza on the probability of non-response in each NCDS follow-up

Table A7: Inverse-probability-weighted estimates of the ATE of *in-utero* exposure to influenza on the probability of non-response at age 7, 11, 23, 33, 42, and age 50 follow-ups.

	Observations	Potential outcome mean under no exposure	ATE
Age 7	14,413	0.853*** (0.003)	-0.003 (0.009)
Age 11	14,398	0.793*** (0.004)	0.001 (0.010)
Age 23	14,305	0.722*** (0.004)	0.008 (0.011)
Age 33	14,233	0.655*** (0.004)	-0.006 (0.012)
Age 42	14,110	0.663*** (0.004)	-0.010 (0.012)
Age 50	13,893	0.578*** (0.004)	-0.019 (0.013)

Robust standard errors in parenthesis; *p<0.05, **p<0.01, ***p<0.001.
Estimation sample reductions due to increases in the number of deaths.

Author statement

Alex Turner: Conceptualization, Data Curation, Methodology, Formal analysis, Writing- Original Draft, Writing- Reviewing and Editing; **Eleonora Fichera**: Supervision, Methodology, Writing- Reviewing and Editing; **Matt Sutton**: Supervision, Methodology, Writing- Reviewing and Editing.