DISCUSSION PAPER SERIES

IZA DP No. 3042

# Childhood Economic Conditions and Length of Life: Evidence from the UK Boyd Orr Cohort, 1937-2005

Paul Frijters
Michael A. Shields
Timothy J. Hatton
Richard M. Martin

September 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Childhood Economic Conditions and Length of Life: Evidence from the UK Boyd Orr Cohort, 1937–2005

**Paul Frijters**
*Queensland University of Technology*

**Michael A. Shields**
*University of Melbourne and IZA*

**Timothy J. Hatton**
*University of Essex, Australian National University and IZA*

**Richard M. Martin**
*University of Bristol*

Discussion Paper No. 3042
September 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3042
September 2007

# ABSTRACT

## Childhood Economic Conditions and Length of Life: Evidence from the UK Boyd Orr Cohort, 1937–2005

We study the importance of childhood socioeconomic conditions in explaining differences in life expectancy using data from a sample of around 5,000 children collected in the UK in 1937-39, who have been traced through official death records up to 2005. We estimate a number of duration of life models that control for unobserved household heterogeneity. Our results confirm that childhood conditions such as household income and the quality of the home environment are significant predictors of longevity. Importantly, however, the role of socio-economic status appears to differ across cause of death, with household income only being a significant predictor of death from cancer. Moreover, we find that children born in a location with relatively high infant mortality rates live significantly fewer years, that 1st born children in the family live significantly more years, and that there is a very high correlation in longevity across children from the same family across all causes of death. We estimate that the difference in life expectancy between 'good' and 'bad' households is as large as 11 years.

JEL Classification: I12

Keywords: childhood, socio-economic characteristics, length of life, duration models

Corresponding author:

Timothy J. Hatton
School of Economics
Australian National University
HW Arndt Building 25a
ACT 0200
Australia
E-mail: tim.hatton@anu.edu.au

## 1. Introduction

A great deal of recent attention and debate by economists has focused on the role of early childhood socio-economic conditions and future life outcomes, including educational attainment, occupational success, morbidity and longevity. In a recent paper, Van den Berg et al. (2006) matched information on prevailing macroeconomic conditions when an individual was born, together with some basic parental characteristics, to explore the predictive power of these socio-economic variables in explaining mortality differentials in three provinces of the Netherlands. The advantage of their data was the long time frame, 1812 to 1992, which provided considerable variation in the business cycle together with a number of disease outbreaks including cholera in 1849, smallpox in 1870-71 and influenza in 1918. Their data were obtained from registers of birth, marriage and death certificates. Using a Mixed Proportional Hazard (MPH) model, and conducting a variety of robustness checks, they found a significant relationship between the macroeconomic conditions at birth and longevity, with being born in a recession *leading* to 'a few years less' of life than an individual born in a boom time. They also found a significant relationship between father's social class, father's level of literacy and length of life.

In this paper, we contribute to this literature by exploring the role of childhood socio-economic conditions of children living in England and Scotland in the 1930s in predicting their length of life. The data collected between 1937 and 1939 are from the Carnegie United Kingdom Trust's Study of Family Diet and Health in Pre-War Britain (hereafter the Boyd Orr cohort, after Sir John Boyd Orr, the Director of the Rowett Institute, where the study was undertaken). This was one of the first surveys to focus on the well being of children, and it collected a wide-range of information about living conditions and household characteristics for around 5,000 children. We use this information together with matched mortality data obtained from official death registers, collected up to 2005 by researchers in the Department of Social Medicine at the University of Bristol, to investigate if childhood socio-economic conditions were significant predictors of mortality. In this respect, we build on several papers by medical researchers who have used these data (see Frankel et al., 1999; Dedman et al., 2001).[1]

A review of the wider epidemiology literature on childhood socioeconomic conditions and mortality can be found in Galobardes et al. (2004), who note that there has been 'relatively little investigation of how early life social circumstances influence adult health' (p.8). In an analysis of the

[1] More generally, the Boyd Orr cohort has also been used to examine the relationship between childhood conditions and health during the subsequent life course (see Martin et al., 2005). Some of the studies have focused on the links between anthropometric or medical indicators observed in 1937-39, and the onset of a variety of medical conditions later in life. One example is the link between childhood leg length and cancer risk (Gunnell et al., 1998a; Gunnell et al., 1998b); another is between childhood body mass index, cancer risk and cardiovascular mortality (Gunnell et al., 1998c; Jeffreys et al., 2005). Other studies have focused on the connection between food intake and its dietary constituents and life course outcomes (see Frankel et al., 1998; Ness et al., 2005).

National Child Development Survey (NCDS), a 1958 UK birth cohort, Case et al. (2005) find that individual health and socio-economic background during childhood are strongly related to self-reported health status in adulthood. Exploring the intervening mechanisms they conclude that health outcomes in adulthood are, to a large extent, carried forward from childhood. Birth cohort data such as these do not allow the modelling of unobserved family effects, and more important still, the cohorts are typically too recent to permit the analysis of mortality. However in a recent study Strand and Kunst (2007) analyse the relationship between childhood socio-economic conditions and early adult mortality of Norwegians born in the period 1955-65. Another recent example is Hamil-Luker and O'Rand (2007) who examine the link between childhood socio-economic condition and heart attack risk in adulthood using data from the US Health Retirement Study.

We estimate multivariate single and competing-risks duration models that utilise the fact that we observe multiple children in nearly all of the households. We pay particular attention to the role of household income in predicting mortality and we provide new evidence on the influence of variables such as birth order and housing conditions. We also evaluate neighbourhood effects using matched data on local unemployment and infant mortality rates.

Our data has both strengths and weakness compared to that analysed by Van den Berg et al. (2006). In addition to examining data from a different country in a very different time period, there are three main advantages of our data. Firstly, we have far greater information on child living conditions and parental socio-economic characteristics than is available from birth and marriage registers alone. Secondly, as already noted, we observe multiple children in each household, and we use this information to improve the identification of unobserved heterogeneity in our models. Thirdly, we have information on disease-specific mortality that allows us to test whether childhood living conditions have differential effects of the probability of dying from heart disease, cancer or other causes. The two main disadvantages of our data are that we observe only one cohort of children and so have little variation in macroeconomic conditions (although we do observe large differences across areas), and we do not follow all the children from birth as the children in our sample range between ages 1 and 17 at the time of the household interview.

The paper is setout as follows. In Section 2, we describe the historical context of the period and introduce the data. Section 3 provides detail of our duration framework modelling strategy, while the results are discussed in Section 4. We conclude the paper in Section 5.

## 2. Data and Historical Context

The Boyd Orr survey that we draw on in this paper represents the culmination in the interwar period of a line of social investigation originating in the late nineteenth century that focused on the relationship between poverty and life chances. It is one of the few interwar surveys for which the original records have survived and the only one for which conditions in childhood have been linked to subsequent longevity.[2] For our purposes, it also has two other important advantages. One is that it took place at a time when severe poverty still existed, particularly among children, and it pre-dates the dramatic levelling of incomes and the fall in the incidence of poverty that took place during the Second World War. This, together with the post-war expansion of the welfare state, means that children observed in the 1930s typically faced much better conditions in later life.[3] As a result, we are better able to identify the effects of deprivation in childhood rather than the persistent effects of poverty throughout the life cycle.[4] The second is that the survey took place long enough ago to make possible the study of longevity: by 2005 one third of the children observed in 1937-9 had died.

In his pioneering household survey of York in 1899, Rowntree (1901, p. 111, 117) found that 9.9% of the population were living in 'primary' poverty (having incomes insufficient even to maintain physical efficiency) while a further 17.9% were in what he described as 'secondary' poverty. He also gave a graphic description of the health deficits suffered by children in the poorest households. From the turn of the century to the 1930s, a number of studies reported similar results for different towns and cities.[5] They found over time that low wages coupled with large family size had diminished as a cause of poverty while unemployment became a more important cause, particularly in the 1930s.

In 1936, Rowntree undertook a second survey of York, which afforded comparisons with the earlier survey. He found that primary poverty had diminished from 9.9% to 5.1% of those in working class households, of whom 46.6% were children under 14. However, using a poverty line a little higher than bare subsistence 17.8% were in poverty in 1936 of whom 33.6% were children (Rowntree, 1941, pp. 35-6, 110). Rowntree also identified significant improvements in the health of children since 1899, including a sharp reduction in death rates, particularly due to infectious diseases, and gains of 1-2 inches in average height and 5-6 pounds in average weight. Nevertheless, differences in height, weight and general health between children in the poorest and least poor families remained as marked as ever.[6] This was also revealed in the data collected by Boyd Orr in 1937-9 (see further below). The gap between the poorest and least poor children at age 10-15 was 2.8 inches in height and 13 pounds in weight (Rowett Institute, 1955, p. 144), and there were large differences across income classes in a variety of clinical abnormalities and medical conditions.

---

[2] Other surveys of working class households, for which reasonably complete records have survived are the New Survey of London Life and Labour, taken in 1929-31, and Seebohm Rowntree's second survey of York, taken in 1936.

[3] In their 1950 survey of working class households in York, Rowntree and Lavers (1951) found that poverty had almost disappeared since the 1930s (falling from 17.8% to 2.1% of the population) and that much of this was due to the postwar expansion of the welfare state. Although these findings now appear to have been exaggerated (Hatton and Bailey, 2000) the fall in poverty was nevertheless dramatic.

[4] This is not to say that the link between income and health outcomes evaporated after the War, only that absolute poverty became far less important.

[5] For an excellent survey of the evolution of poverty, its measurement and policy towards it (see Gazeley (2003)).

[6] The gap between the poorest and least poor children in heights, weights and death rates was roughly equivalent to the improvement in the overall averages of these measures between 1899 and 1936 (Rowntree, 1941, p. 296).

Interest in the effects of poverty on health was stimulated by concerns about national efficiency and by improvements in nutritional science. In 1896-1900, 34% of army recruits presented for medical examination were rejected on account of lack of physical fitness. From that time it was increasingly recognised that stature, weight and health were linked to standards of nutrition, which were determined largely by income. While contemporary studies were not able to identify the effects of childhood poverty over the whole life course, there was a growing recognition that conditions early in life were important. This translated into increasing political pressure for better health services and especially for a more comprehensive social security system that would include family allowances (Macnicol, 1980). Yet until the War, the government resisted these budgetary pressures, refusing to acknowledge the extent of deprivation among poor children and suppressing evidence to the contrary (Mayhew, 1988).

Up to the late 1930s, health insurance covered only employees (and not their spouses or children), public services for maternal and child welfare were very limited (especially in poor areas), and only 2% of children received free school meals (Webster, 1985). In 1936, two studies shook the assumptions of the establishment. One by M'Gonigle and Kirby (1936) provided direct evidence on the relationship between food expenditure and children's health in a sample of households in Stockton-on-Tees. The other by Sir John Boyd Orr (1936) calculated that 10% of the national population had diets that were nutritionally deficient (their average weekly food expenditure was just £8.80 per head at the prices of 2005). To provide further direct evidence Boyd Orr embarked on a household survey of diets and health in 1937-9 (the survey that we use here), but the research was overtaken by events and the results were not published until 1955 (Rowett Institute, 1955).

The records from this study survive and have been supplemented with additional data by a team of medical researchers at the University of Bristol (see Martin et al., 2005). The original survey covered a total of 1,343 households, containing 4,999 children and young adults under the age of 19. A selection of households was surveyed in 16 different towns and villages in England and Scotland, which were intended to be representative of both rural and urban locations.[7] It was confined to families with children and was targeted so as to over-represent poor families. The survey recorded details of the demographic structure of each household together with details of income and housing conditions. It also recorded itemised details of expenditure on and consumption of food during the survey week. Finally, the clinical part of the survey collected a variety of indicators of the anthropometric and health status of the children including height, leg length, weight, incidence of medical conditions and dental decay. The Bristol team conducted follow up surveys in 1997/8 and 2002/3 of the survivors of the Boyd Orr cohort who could be traced, collecting details of their health

[7] The 16 locations are, in Scotland: Aberdeen, Kintore, Hopeman, Barthol Chapel, Methlick, Tarves, West Wemyss, Coaltown of Wemys, Dundee and Edinburgh; and in England: Barrow-in-Furness, Liverpool, Yorkshire, Wisbech, Fulham and Bethnal Green. These locations are identified on a map in Martin et al. (2005. p. 743).

and socio-economic status later in life. More important, for our purposes they used the NHS Central Register to identify the date and cause of death, for those who were resident in Britain on 1st January 1948 and for deaths occurring up to 31st July 2005. Deaths before 1948 were identified from other sources and with less detail. Of the 4,999 children, we have reliable follow-up information up to 2005 or date of death for 4,510 (or 90%).[8]

In this paper, we focus on the sample of 4,460 children aged less than 18 years at the time of interview and who have been successfully tracked. We drop 65 cases that are possible combat deaths in the Second World War and a further 50 because household income is missing.[9] These children lived in 1,251 households, with the average number of children observed in each family being 3.6. Table 1 provides means and standard deviations of the key variables for this sample. Importantly, one-third of these children had died by 2005, and the average age at death was 62 years. Cause of death was ascertained from death certificates and classified according to the International Classification of Diseases, 9th and 10th Revisions. In terms of broad causes of death, 8.2% of the sample died from ischaemic heart disease (364 individuals; average age of death 63.6 years), 11.1% from cancer (495 individuals; average age of death 64.8 years) and 14.5% from 'other causes' (645 individuals; average age of death 58.2 years). The group 'other causes' contains 2.6% (116 individuals) of the sample who died from stroke.

Household income per capita is categorised as low (less than 10 shillings per week), medium (10-20 shillings) and high (greater than 20 shillings). To put this in perspective, 10 shillings per week is equivalent to 2.4 US dollars per person per day at 2006 prices—a figure that is close to the World Bank's international poverty line. Boyd Orr (1936, p. 49) identified those living with incomes of less than 10 shillings per week as having a food intake that was deficient in almost every constituent, while those with slightly higher incomes suffered deficiencies mainly in certain minerals and vitamins. Unemployment of the household head, an indicator that was strongly linked to poverty in the interwar period, affected 34% of the children in the sample. As a proxy for poor housing conditions, 32% of them lived in homes with inadequate ventilation as recorded by the interviewer. To these household level variables we have added two measures of conditions in the locality, the unemployment rate and the infant mortality rate.[10] The average unemployment rate across the 16 areas was 12.3%, but there was a great deal of variation across areas with the rate ranging from 6.5% to

[8] The trace rate has increased slightly since earlier research that used the same database as a result of further searches of archived records, contacts with surviving study members and additional notifications from the NHSCR. The dataset used here includes some individuals for whom only partial follow up information is available.

[9] Since we have no cause of death before 1948 we cannot identify those who died in the War. We therefore exclude males born before 15th August 1927 and who died between 3rd September 1939 and 15th August 1945.

[10] The average unemployment rate for 1937-9 was calculated from the Ministry of Labour's local unemployment index; the local infant mortality rate for the same years was obtained from Registrar General Reports for England and Wales and for Scotland.

22.2%. Similarly, while the average area infant mortality rate was 62 deaths per 1,000, this ranges in the data from 51.3 to 79.5.[11]

To get a preliminary feel for the relationship between childhood socio-economic conditions and length of life, we plot the death hazard rate for those living in a low-income household (i.e. less than 10 shillings per capita), compared to those living in the higher income households (Figure 1). We also plot the death hazard by area infant mortality rate, where we simply distinguish between areas with lower and greater than the average rate. The data clearly suggest that children living in a higher income household have a lower probability of death at any age than their low-income equivalents. This is the case at virtually every age, and the differential appears to be greatest for deaths in the age range of approximately 55 and 80 years. Similarly, although the differences are not as clear or large as for the case of household income, individuals who resided in areas with relatively high infant mortality as a child had higher death hazards particularly evident in the age range of approximately 60 to 75 years.

## 3. Duration Models

To further explore the predictive power of childhood socio-economic conditions on length of life, we estimate both Proportional Hazard (PH) and single and competing risks Mixed Proportional Hazard Models (MPH) of mortality. Importantly, with the MPH models we make use of the fact that we observe multiple children in nearly all of the sample families. The basic building block of the MPH models is a hazard function that is multiplicative in its components:

$$\theta(t \mid x, v) = \lambda(t) * \phi(x) * f(v) \quad (1)$$

where $\theta(t \mid x, v)$ is the hazard rate of the event in question, in this case death, and *t* denotes the elapsed duration i.e. age. Here, $\lambda(t)$ is called the baseline hazard and attempts to capture time varying unobservables that are the same for all individuals and move systematically with age. The function $\phi(x)$ is a regressor function dependent on an observable vector of characteristics *x*. This function is usually taken to be $e^{x\beta} > 0$ where *x* can include polynomials or other functions. Here, $f(v)$ is a time-invariant unobserved positive heterogeneity component whose distribution is not necessarily ex ante known. In our application, we take $v$ to be a fixed term for all children in the same family and thus rely on stronger information for the identification of $v$ than is possible under single-risk MPH models. The standard formulation is to take $\theta(t \mid x, v) = \lambda(t) * e^{x\beta + v}$.

[11] This infant mortality rate is roughly the same as current rates in, for example, Cambodia, Myanmar, Nepal and Sudan.

In our empirical specification for the baseline hazard, we take a very flexible specification for $\lambda(t)$ i.e. we take $\lambda(t)$ to be piece-wise constant (=Kaplan-Meyer):

$$
\begin{aligned}
&\lambda(t) = e^{\lambda 1} \text{ for } t_1^\lambda > t \geq t_0^\lambda = 0 \\
&\lambda(t) = e^{\lambda 2} \text{ for } t_2^\lambda > t \geq t_1^\lambda \\
&\lambda(t) = e^{\lambda 3} \text{ for } t_3^\lambda > t \geq t_2^\lambda \\
&\ldots
\end{aligned} \tag{2}
$$

where $t_1^\lambda,...,t_M^\lambda$ are constants and $\lambda_1,...,\lambda_M$ are parameters to be estimated. The log-likelihood of an individual observed dying at age $T_i$ when initially observed at age $T_0$ is then:

$$L_i(T \mid x_i) = \int L_i(T_i \mid x_i, T_0, v_i) \partial G(v) \tag{3}$$

where $G(v)$ is the unobserved heterogeneity distribution that needs to be integrated out. We take a flexible mass-point distribution with:

$$P(v_i = v^0) = p_0,\ P(v_i = v^1) = p_1, ...,\ P(v_i = v^K) = p_K \tag{4}$$

Here, $v^k$ are points to be estimated and $p_k$ are probabilities to be estimated. In our empirical applications we take *K*=2 though we also tried higher numbers for K. In the case that all the points $v^k$ are identical, the MPH model reverts back to the PH model.

The eventual likelihood then becomes:

$$
\begin{aligned}
&L_i(T_i \mid x_i) = \int L_i(T_i \mid x_i, T_0, v_i) G(v_i) \\
&= \sum_{k=1}^{K} p_k * L_i(T_i \mid x_i, T_0, v^k) \\
&= \sum_{k=1}^{K} p_k * [\theta(T_i \mid x_i, v^k)] e^{-\int_{T_0}^{T_i} \lambda(s) e^{x_i\beta + v^k} \partial s}
\end{aligned} \tag{5}
$$

and the likelihood of observing an individual still alive at age $T_i$ equals:

$$L_i(T_i \mid x_i) = \sum_{k=1}^{K} p_k * e^{-\int_{T_0}^{T_i} \lambda(s) e^{x_i\beta + v^k} \partial s}$$
$$= \sum_{k=1}^{K} p_k * e^{-e^{(x_i\beta + v^k)} \int_{T_0}^{T_i} \lambda(s)\partial s} \tag{6}$$

Since we presume $v^k$ to be the same for different individuals within the same family, the likelihood over the *J* families becomes:

$$L_j = \sum_{k=1}^{K} p_k \prod_{i \in S_j} L_i(T_i \mid x_i, T_0, v^k) \tag{7}$$

where $S_j$ is the set of individuals in family *j*. The total likelihood is then $L = \prod_j L_j$ .

We also estimate a model that captures the possibility that different childhood factors might be associated with different disease-specific deaths. In this respect, the data allow us to differentiate between three broad causes of death: (1) heart disease, (2) cancer and (3) other deaths. In order to allow for the possibility that the same variable may be associated with increased hazard rates of death for one cause of death but not another, we will also estimate a competing risks MPH model. Formally, this means that we look at three distinct hazard rates, $\theta^h(t \mid x, v)$, $\theta^c(t \mid x, v)$ and $\theta^o(t \mid x, v)$ associated with heart disease, cancer and other causes of death, respectively. Each of these is modelled as above. The likelihood of observing someone dying from heart disease at age $T_i$ is then for instance:

$$L_i(T_i \mid x_i) = \sum_{k=1}^{K} p_k * \theta^h(T_i \mid x_{i,} v^k) e^{-\int_{T_0}^{T_i} (\theta^h + \theta^c + \theta^o)\partial s} \tag{8}$$

where the term $e^{-\int_{T_0}^{T_i} (\theta^h + \theta^c + \theta^o)\partial s}$ denotes the probability of surviving all three risks from the ages of $T_0$ to $T_i$. Other possible data outcomes (death from one of the other two risks or no death at all) are modelled analogously. Again, we allow for unobservable heterogeneity with *K* points of support.

A final point of model selection is the assumption regarding the moment of mixing i.e. the moment in time that the unobserved heterogeneity term is presumed independent of the other observed characteristics. We presume that the unobserved heterogeneity is household specific in which case it is more natural to assume that the moment of mixing is at the interview. All observed characteristics in our sample, except age, gender and age at interview, are household characteristics; the sample was designed as a household sample rather than a sample of individuals. In addition, within households premature deaths of children in the period of the survey were likely to have led households to respond by producing more children. These factors imply that the unobserved heterogeneity term is best viewed as a household characteristic that is orthogonal to the other characteristics at the moment of the interview.[12]

## 4. Results

We estimate both Proportional Hazard (PH) and Mixed Proportional Hazard (MPH) models of length of life. We also allow for the socio-economic childhood conditions to have a differential effect on length of life by estimating a competing-risks MPH (CRMPH) model that distinguishes between death from ischaemic heart disease, cancer and other causes. The results are shown in Tables 2, 3 and 4, respectively. For each model we provide two specifications in order to illustrate the change in the parameter estimates when we include additional explanatory variables that are likely to be correlated with household income. In Specification 1 (Spec 1), we control for only age categories, sex and household income per capita. In the second specification (Spec 2), we also include controls for other socio-economic conditions expected to be important in predicting lifetime health. These are: whether the head-of-household was unemployed, the number of children in the household, whether the child was first-born in the family and whether the home had adequate ventilation (as assessed by the interviewer). To test whether the external environment to the household plays an additional role in predicting longevity we also control for the area infant mortality rate at the time of interview. We note here the very high correlation between infant mortality and the unemployment rate (0.92) across the 16 areas in question. Since we observe unemployment at the household level, we drop the local unemployment rate and focus on infant mortality as an indicator of the local health environment.

*Single-Risks (All-Cause Mortality)*

The results for all-cause mortality are presented in Table 2 and the first column in Tables 3 and 4. Those reported here are not sensitive to the exclusion of all deaths before 1948, for which our information is less reliable. For Spec 1, the parameter estimates are very similar for the PH and the MPH models. As expected, all the models show a monotonically increasing probability of death with age, and that males live significantly fewer years than females. Importantly, we observe that higher household income per capita is associated with significantly longer life in both the PH and MPH all-cause models, and the differential in life expectancy predicted by income is large. For example, using the income estimates from the MPH model we calculate that at the sample means life expectancy was higher by 0.84 years for those in the middle income group and by 3.13 years for the high income group, compared to those children living in low income households (i.e. less than 10 shillings per capita per week).

[12] The alternative would be to assume that the unobserved heterogeneity is individual-specific and orthogonal at birth to the other observed characteristics. Technically, this would not create a problem as it would simply require us to integrate out the probability of surviving up to the moment of interview.

Moving from Spec 1 to Spec 2 (Table 2), we add into the model other socioeconomic and demographic variables that are likely to be correlated with household income. As a result of correlation with these variables, particularly unemployment, the coefficients on income are substantially reduced although the high-income dummy remains significant for the PH model. The coefficient on unemployment of the household head is negative as expected, and significant at the 10 percent level, which is consistent with the results reported by Frankel et al (1999) using the Boyd Orr cohort[13] and also with the business cycle effects found by Van den Berg et al (2006). This effect may reflect more than the loss of income, with which it is correlated. Interwar observers stressed that prolonged unemployment of the main family breadwinner also had significant psychological effects, which could affect the well-being of the whole household (see Beales and Lambert, 1934; Eisenberg and Lazarsfeld, 1938).

Interestingly, we find no evidence that the number of the children in the household was a significant predictor of mortality, but we do find evidence of greater longevity for first-born children, whose life expectancy is estimated to be greater by 1.88 years than those of higher parities. Similar results were found by Modin (2002) using data from the Uppsala Birth Cohort Study that tracked children born in the Uppsala Academic Hospital in Sweden during the period 1915-29. This finding is also consistent with other evidence that first-born children enjoy family resource advantages over children at higher parities.[14]

Turning to indicators of housing conditions, we initially included a measure of overcrowding (persons per room) which we found to be insignificant in all the models. Since there are missing observations for this variable we have dropped it in order to maximise the sample (dropping these cases has little effect on the other coefficients). However, the models do suggest that housing quality as captured by good ventilation (assessed by the interviewer) has a positive effect on longevity, which is significant at the 95% level of confidence in the PH model (90% level of confidence in the MPH model). This is consistent with Dedman et al. (2005) who found, using the Boyd Orr data, that this was the only indicator of housing conditions that significantly affected longevity. This variable probably captures housing conditions more generally, distinguishing for instance between dark dingy tenements and more modern housing.

Infant mortality may affect health and longevity in different ways. On one hand the selection effect may leave the remaining population healthier but, as an indicator of the disease environment, infant mortality may have a scarring effect on those who survive into childhood (see Bozzoli et al., 2007). Interestingly, we find that children who lived in an area of high infant mortality in the 1930's,

[13] Frankel et al. (1999) analysed the 767 deaths recorded for the Boyd Orr sample up to September 1997 (3,750 of the children had been traced at that point). They did not include household income or the other socio-economic control variables in their models, but rather focused on social class indicators.

[14] See, for example, evidence from British national samples reported by Kaplan et al. (1992) for health outcomes, and by Booth and Kee (2005) for educational outcomes.

experienced significantly (with 95% confidence for the PH model, and 90% confidence for the MPH model) lower life expectancy than those residing in low infant mortality areas. Thus the scarring effect dominates. In fact, we calculate that a one per thousand increase in area infant mortality rate cost these individuals about 0.08 years of expected length of life.[15] Infant mortality varied between 50 and 80 deaths per thousand births over the survey areas, implying that people born in the 'best areas' with an infant mortality rate of only 50 lived about 2 years longer than those born in the 'worst areas'.

Finally, using the estimates from the MPH model, we can look at the importance of unobserved household characteristics in determining longevity. In particular, we can compare the expected length of life for the unobserved 'good types', which consist of some 86% of the households with the 'bad types', which consist of the remaining 14% of the households. The 'bad types' have an average life expectancy that is around 10.9 years lower, suggesting that household unobservable characteristics that we can quantify given the data and the model have a large amount of predictive power. We reiterate that the identification of this finding is helped by the fact that we have more than one child per household for the majority of households. This implies that the results are not driven by the interaction between other variables and duration, which is the source of identification of unobserved heterogeneity in single-spell duration models that lack multiple observations per household.

*Competing-Risks (Cause-Specific Mortality)*

In contrast to many other studies we treat the causes of death as competing risks, explicitly recognising their interdependence. In Tables 3 and 4 we find that males have a higher mortality risk from all causes but that this gender differential is larger for heart disease than for other causes. In Spec 1 (Table 3) household income significantly reduces mortality resulting from cancer and (for high income) from other causes. However, we note that the income coefficients are very similar in magnitude across for each of the causes of death.

When we add other variables (Table 4) the income effects become weaker, consistent with the results for all causes. Although the coefficient on head-of-household unemployment is positive (associated with higher mortality) for all three broad causes of death, it is not statistically significant. Interestingly, having a large number of siblings was associated with a significantly increased risk of death from other causes but not from cancer or ischaemic heart disease. We also find that being 1st born was associated with a significantly reduced probability of death from other causes relative to later-born children. The effect of housing conditions as reflected in good ventilation, that we identified for all-cause mortality also works principally though death from other causes. Finally, while the

[15] Regidor et al. (2002) found that area infant mortality was associated with higher adult mortality using data from deaths in Madrid in 1996 and 1997 matched to 1996 Census area data.

coefficients on area infant mortality are positive for all three broad causes of death, they are not statistically significant.

## 5. Conclusion

In this paper, we have contributed to the growing economics literature that focuses on the role that childhood socio-economic conditions play in determining educational, health and labour market outcomes in later life. We exploit a unique dataset that offers a lifetime perspective on children who grew up in the 1930s, when conditions were very different from those of today. It allows us to observe the effects of serious childhood deprivation in a number of dimensions and to link these with later mortality. We also develop a methodology that allows us to identify household specific effects by exploiting the fact that we observe all the children in a given household. Overall our results support the conclusions of epidemiological studies as well as those by economists such as Case et al. (2005) that identify the effects of conditions during childhood on health outcomes later in life.

Five key findings emerge from this analysis. First, growing up in poverty, as reflected in low income and unemployment of the household head, has negative effects on longevity. Second, housing conditions influence longevity through the quality of housing rather than through overcrowding. Third, there appear to be neighbourhood effects that led to substantial differences for those who grew up in different localities. Fourth, there are large variations in longevity associated with unobserved differences between ‘good’ and ‘bad’ households, which could be due to genetics or to unobserved environmental factors at the household level. Finally, these conditions had differential effects on the causes of death, with low income most strongly associated with death from cancer and other conditions such as housing quality influencing death from other causes.

## References


Beales, H. L. and Lambert, R. S. (1934). *Memoirs of the Unemployed*. London: Gollancz.

Booth, A. L. and Kee H. J. (2005). Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment. *IZA Discussion Paper 1713*: Bonn: IZA.

Boyd Orr, J. (1936). *Food, Health and Income*. London: Macmillan.

Bozzoli, C., Deaton, A. S. and Quintana-Domeque, C. (2007), "Child Mortality, Income and Height," NBER Working Paper 12966: Cambridge: NBER.

Case, A., Fertig, A. and Paxon, C. (2005) "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics,* 24, pp. 365-389.

Dedman, D. J., Gunnell, D., Davey Smith, G. and Frankel, S. (2001). Childhood Housing Conditions and Later Mortality in the Boyd Orr Cohort. *Journal of Epidemiology and Community Health,* 55, pp. 10-15.

Eisenberg, P. and Lazarsfeld, P. F. (1938). The Psychological Effects of Unemployment. *Psychological Bulletin*, 35, pp. 358-391.

Frankel, S., Davey Smith, G. and Gunnell, D. (1999). Childhood Socioeconomic Position and Adult Cardiovascular Mortality: The Boyd Orr Cohort. *American Journal of Epidemiology*, 150, pp. 1081-1086.

Galobardes, B., Lynch, J. W. and Davey Smith, G. (2004). Childhood Socioeconomic Circumstances and Cause-Specific Mortality in Adulthood: Systematic Review and Interpretation. *Epidemiologic Reviews*, 26, pp. 7-21.

Gazeley, I. (2003), *Poverty in Britain, 1900-1965*, Basingstoke: Palgrave-Macmillan.

Gunnell, D., Davey Smith, G., Frankel, S., Nanchahal, K., Braddon, F. E., Pemberton, J. and Peters, T. J. (1998a). Childhood Leg Length and Adult Mortality: Follow up of the Carnegie (Boyd Orr) Survey of Diet and Health in Pre-War Britain. *Journal of Epidemiology and Community Health*, 52, pp. 142-152.

Gunnell, D., Davey Smith, G., Holly, J. M. P. and Frankel, S. (1998b). Leg Length and Risk of Cancer in the Boyd Orr Cohort. *British Medical Journal*, 317, pp. 1350-1.

Gunnell, D. J., Frankel, S., Nanchahal, K., Peters, T., Maynard, M. and Davey Smith G. (1998c). Childhood Obesity and Adult Cardiovascular Mortality. *American Journal of Clinical Nutrition,* 67, pp. 1111-8.

Hamil-Luker, J. and O'Rand, A. (2007). Gender differences in the link between childhood socioeconomics conditions and heart attack risk in adulthood. *Demography*, 44, pp. 137-158.

Hatton, T. J, and Bailey, R. E. (2000). Seebohm Rowntree and the Postwar Poverty Puzzle. *Economic History Review*, 53, pp. 517-543.

Jeffreys, M., Davey Smith, G., Martin, R. M., Frankel, S. and Gunnell, D. Childhood Body Mass Index and Later Cancer Risk: A 50 Year Follow up of the Boyd Orr Study. *International Journal of Cancer*, 112, p-p. 348-351.

Kaplan, B. A., Mascie-Taylor, C. G. N. and Bolsen J. (1992). Birth Order and Health Status in a British National Sample. *Journal of Biosocial Science*, 24, pp. 25-33.

Macnicol, J. (1980). *The Movement for Family Allowances: A Study in Social Policy*. London: Heineman.

Martin, R. M., Gunnell, D., Pemberton, J., Frankel, S. and Davey Smith, G. (2005). Cohort Profile: The Boyd Orr Cohort - An Historical Cohort Study Based on the 65 Year Follow-up of the Carnegie Survey of Diet and Health (1937-39). *International Journal of Epidemiology*, 34, pp. 742-749.

Mayhew, M. (1988). The 1930s Nutrition Controversy. *Journal of Contemporary History*, 23, pp. 445-464.

Modin, B. (2002). Birth Order and Mortality: A Life-Long Follow-up of 14,200 Boys and Girls Born in Early 20th Century Sweden. *Social Science and Medicine*, 54, pp. 1051-64.

Ness, A. R., Maynard, M., Frankel, S., Davey Smith, G., Frobisher, S., Leary, S. D., Emmett, P. M. and Gunnell, D. (2005). Diet in Childhood and Adult Cardiovascular and All Cause Mortality: The Boyd Orr Cohort. *Heart,* 91, pp. 894-898.

Regidor, E., Gutiérrez-Fisac, J. L., Elisa Calle, M., Navarro, P. and Domínguez, V. (2002). Infant Mortality at Time of Birth and Cause-Specific Adult Mortality among Residents of the Region of Madrid born elsewhere in Spain. *International Journal of Epidemiology*, 31, pp. 368-374
Rowett Institute (1955). *Family Diet and Health in Pre-War Britain*. Dunfermline: Carnegie UK Trust.
Rowntree, B. S. (1901), *Poverty: A Study of Town Life*, London: Macmillan.
Rowntree, B. S. (1941). *Poverty and Progress: A Second Social Survey of York,* London: Longmans Green.
Rowntree, B. S. and Lavers, G. R. (1951). *Poverty and the Welfare State*, London: Longmans Green.
Strand, B. H. and Kunst, A. (2007). Childhood Socioeconomic Position and Cause-specific Mortality in Early Adulthood. *American Journal of Epidemiology*, 165, pp. 85-93.
Van den Berg, G. J., Lindeboom, M. and Portrait, F. (2006), "Economic Conditions in Early Life and Infant Mortality," *American Economic Review,* 96, pp. 290-302.
Webster, C. (1985). Health, Welfare and Unemployment during the Depression. *Past and Present,* 109, pp. 204-230.

Figure 1: Death Hazards by High and Low Household Income

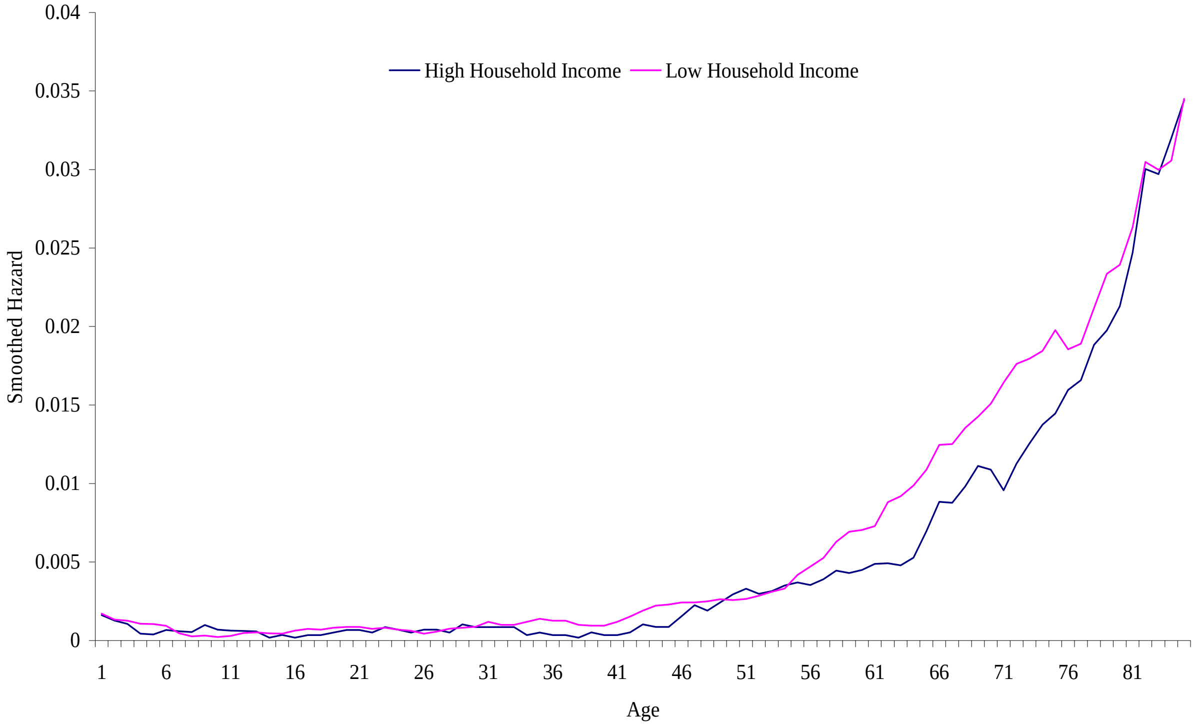


Figure 2: Death Hazards by High and Low Area Infant Mortality

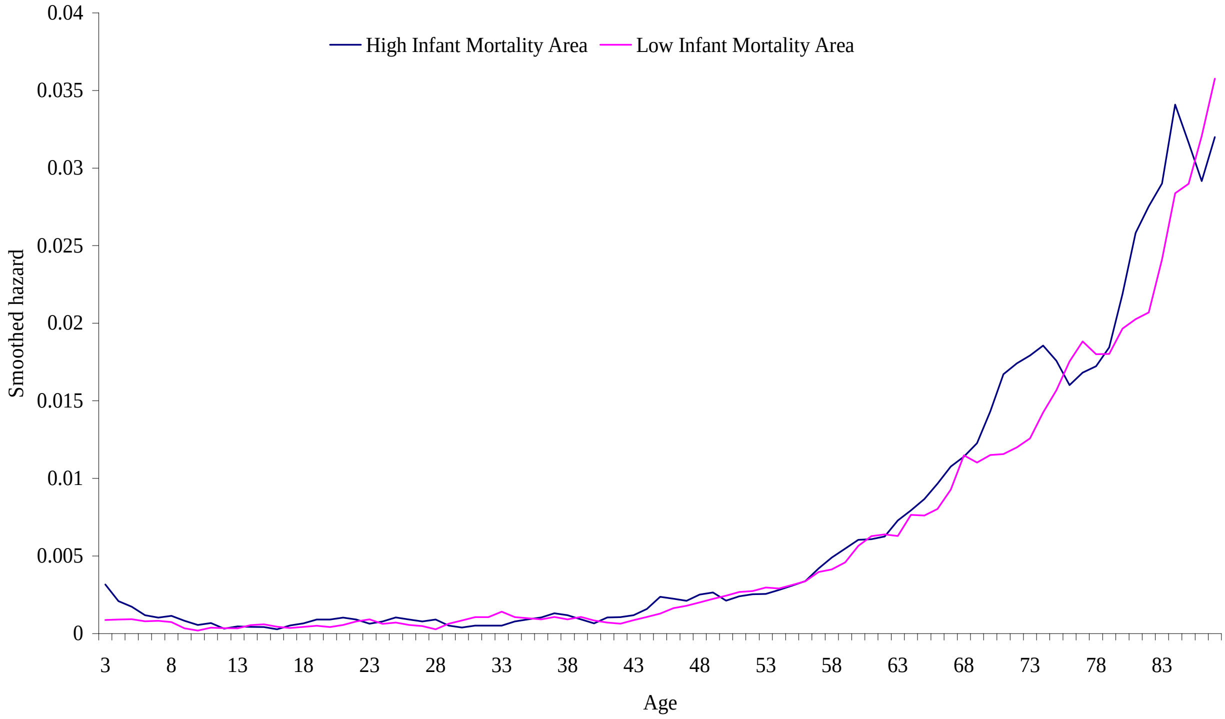

**Table 1:**

**Sample Characteristics**

| | **Average** | **Standard Deviation** |
|---|---|---|
| Male | 0.491 | 0.500 |
| Low household income per capita | 0.733 | 0.443 |
| Middle household income per capita | 0.218 | 0.413 |
| High household income per capita | 0.049 | 0.217 |
| Head of household unemployed* | 0.335 | 0.472 |
| Number of children | 4.804 | 2.084 |
| 1st born child in family | 0.261 | 0.439 |
| Good ventilation in house* | 0.681 | 0.466 |
| Area unemployment (%) | 12.305 | 4.866 |
| Area infant mortality rate per 1000 | 62.009 | 8.116 |
| Total died by 2005 | 0.337 | 0.473 |
| Died from ischaemic heart disease | 0.082 | 0.274 |
| Died from cancer | 0.111 | 0.314 |
| Died from other causes | 0.145 | 0.352 |
| Sample | 4460 | |

*Note*: * means calculated excluding missing cases.

**Table 2:**

**Parameter Estimates from Proportional Hazard (PH) Models of All-Cause Mortality**

| | **PH (Spec 1)** | | **PH (Spec 2)** | |
|---|---|---|---|---|
| | **All deaths** | | **All deaths** | |
| | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** |
| Age<40 | -7.620 | 61.14 | -8.250 | 29.74 |
| Age40-49 | -7.059 | 46.05 | -7.683 | 26.44 |
| Age50-59 | -6.139 | 50.85 | -6.762 | 24.63 |
| Age60-69 | -5.262 | 52.86 | -5.882 | 22.14 |
| Age70-79 | -4.550 | 50.67 | -5.150 | 19.54 |
| Age>79 | -3.741 | 48.82 | -4.298 | 16.78 |
| Male | 0.469 | 8.94 | 0.465 | 8.85 |
| Middle household income | -0.210 | 3.27 | -0.069 | 0.94 |
| High household income | -0.450 | 3.49 | -0.299 | 2.07 |
| HOH unemployed | - | - | 0.114 | 1.84 |
| Number of children | - | - | 0.019 | 1.17 |
| $1^{st}$ born child in family | - | - | -0.164 | 2.68 |
| Good ventilation in house | - | - | -0.141 | 2.44 |
| Area infant mortality /100 | - | - | 0.770 | 2.34 |
| Mean Log Likelihood | -6.583 | | -6.571 | |
| Number of individuals | 4459 | | 4459 | |
| Number of families | 1251 | | 1251 | |
| Number of deaths | 1503 | | 1503 | |

*Notes*: Absolute *t*-statistic shown. The omitted categories are female, low per capita weekly household income, head of household not unemployed, not $1^{st}$ born child in family, house not overcrowded, house assessed by the interviewer as having less than good ventilation. Low household income is per capita weekly household income of less than 10 shilling, Middle household income is 10-20 shillings per capita and High household income is greater than 20 shillings per capita. All models also include a control for ‘age at interview’. In Specification 2, additional controls for missing household unemployment and missing ventilation information are included.

## Table 3:

## Parameter Estimates from Mixed Proportional Hazard (MPH) Models of Mortality

## (Basic Specification)

| | **MPH (Spec 1)** | | **MPH Competing Risks (Spec 1)** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **All deaths** | | **Heart Disease** | | **Cancer** | | **Other Causes** | |
| | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** |
| Age<40 | -7.853 | 51.60 | -12.299 | 18.02 | -10.595 | 27.97 | -7.571 | 46.41 |
| Age40-49 | -7.288 | 40.58 | -9.887 | 21.81 | -9.278 | 24.58 | -7.324 | 34.57 |
| Age50-59 | -6.364 | 41.58 | -8.440 | 24.41 | -7.837 | 30.43 | -6.840 | 36.28 |
| Age60-69 | -5.474 | 39.78 | -7.481 | 24.10 | -6.702 | 32.76 | -6.208 | 34.92 |
| Age70-79 | 4.740 | 36.69 | -6.848 | 22.88 | -5.836 | 30.84 | -5.508 | 36.69 |
| Age>79 | -3.896 | 33.03 | -6.098 | 22.24 | -5.149 | 31.38 | -4.403 | 35.20 |
| Male | 0.4794 | 9.08 | 1.111 | 8.93 | 0.328 | 3.41 | 0.312 | 3.83 |
| Middle household income | -0.224 | 3.16 | -0.132 | 0.92 | -0.332 | 2.87 | -0.164 | 1.71 |
| High household income | -0.463 | 3.49 | -0.403 | 1.49 | -0.467 | 2.22 | -0.477 | 2.41 |
| Heterogeneity | 0.911 | 4.48 | 1.528 | 6.70 | 0.751 | 3.73 | 0.464 | 2.59 |
| Probability | -1.677 | 2.01 | | | -1.194 | 1.97 | | |
| Mean Log Likelihood | -6.583 | | -7.772 | | | | | |
| Number of individuals | 4459 | | 4459 | | | | | |
| Number of families | 1251 | | 1251 | | | | | |
| Number of deaths | 1503 | | 1503 | | | | | |

*Notes*: Absolute *t*-statistic shown. The omitted categories are female, low per capita weekly household income, head of household not unemployed, not 1st born child in family, house not overcrowded, house assessed by the interviewer as having less than good ventilation. Low household income is per capita weekly household income of less than 10 shilling, Middle household income is 10-20 shillings per capita and High household income is greater than 20 shillings per capita. All models also include a control for ‘age at interview’.

**Table 4:**

**Parameter Estimates from Mixed Proportional Hazard (MPH) Models of Mortality**

**(Extended Specification)**

| | **MPH (Spec 2)** | | **MPH Competing Risks (Spec 2)** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **All deaths** | | **Heart Disease** | | **Cancer** | | **Other Causes** | |
| | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** | **Coefficient** | **\|*t*\|** |
| Age<40 | -8.435 | 26.27 | -10.984 | 11.73 | -10.413 | 17.07 | -8.029 | 17.59 |
| Age40-49 | -7.863 | 23.27 | -8.571 | 11.11 | -9.090 | 15.38 | -7.771 | 16.54 |
| Age50-59 | -6.938 | 21.59 | -7.124 | 9.88 | -7.649 | 14.39 | -7.286 | 15.71 |
| Age60-69 | -6.045 | 19.01 | -6.164 | 8.54 | -6.514 | 12.97 | -6.651 | 14.48 |
| Age70-79 | -5.291 | 16.84 | -5.522 | 7.67 | -5.634 | 11.15 | -5.919 | 13.07 |
| Age>79 | -4.406 | 14.18 | -4.756 | 6.61 | -4.917 | 9.78 | -4.752 | 10.66 |
| Male | 0.476 | 8.98 | 1.108 | 8.74 | 0.325 | 3.34 | 0.305 | 3.68 |
| Middle household income | -0.080 | 0.98 | -0.086 | 0.49 | -0.234 | 1.77 | 0.083 | 0.74 |
| High household income | -0.301 | 1.90 | -0.379 | 1.14 | -0.364 | 1.46 | -0.154 | 0.68 |
| HOH unemployed | 0.119 | 1.66 | 0.047 | 0.32 | 0.129 | 1.10 | 0.128 | 1.32 |
| Number of children | 0.019 | 0.96 | -0.017 | 0.41 | 0.006 | 0.17 | 0.054 | 2.08 |
| 1st born child in family | -0.181 | -2.77 | -0.104 | 0.82 | -0.138 | 1.22 | -0.263 | 2.58 |
| Good ventilation in house | -0.147 | -2.31 | -0.174 | 1.26 | -0.039 | -0.36 | -0.236 | 2.70 |
| Area infant mortality /100 | 0.747 | 1.87 | 0.619 | 0.73 | 0.675 | 1.18 | 0.761 | 1.36 |
| Heterogeneity | 0.923 | 3.99 | -1.521 | 6.60 | -0.677 | 3.22 | -0.422 | 2.23 |
| Probability | -1.821 | -2.13 | | | 1.067 | 1.56 | | |
| Mean Log Likelihood | -6.565 | | -7.757 | | | | | |
| Number of individuals | 4459 | | 4459 | | | | | |
| Number of families | 1251 | | 1251 | | | | | |
| Number of deaths | 1503 | | 1503 | | | | | |

*Notes*: Absolute *t*-statistic shown. The omitted categories are female, low per capita weekly household income, head of household not unemployed, not 1st born child in family, house not overcrowded, house assessed by the interviewer as having less than good ventilation. Low household income is per capita weekly household income of less than 10 shilling, Middle household income is 10-20 shillings per capita and High household income is greater than 20 shillings per capita. All models also include a control for ‘age at interview’. Additional controls for missing household unemployment and missing ventilation information are included.