Public health issues in India and UK: Child Mortality and Survival Expectations

Seetha Menon

A thesis submitted for the degree of Doctor of Philosophy in Economics

Institute for Social and Economic Research

University of Essex

July, 2016

Contents

List of Figures

List of Tables

Declarations

No part of this thesis has been submitted for another degree.

Chapter 1 is a joint work between my supervisor Professor Adeline Delavande at The University of Essex and Dr. Jinkook Lee at The University of Southern California.

The other Chapters in this thesis are exclusively mine.

An early draft of Chapter 3 has been previously published as a wroking paper as *Unfinished Lives: The effect of domestic violence on neonatal & infant mortality in India.*, ISER Working Paper Series No. 2014-27, August, 2014

Acknowledgements

This thesis was supported by a studentship in Labour Economics from the Institute for Social and Economic Research (ISER), University of Essex.

I am very grateful to my supervisor, Adeline Delavande, for being extremely supportive and providing precious guidance. I am also grateful to Sonia Bhalotra whose creative suggestions greatly improved this work. I am thankful to João Santos Silva for teaching me everything I know about Econometrics and ISER as an institution for giving me several opportunities to collaborate and work with other academic staff which has considerably broadened my perspective.

I thank my friends at ISER who made this journey enjoyable and for all the fun we had in the last 3 years. I especially thank Vilma, for sticking around during the bad times and the good.

Finally, thank you to my dear parents for raising me to be who I am and for inculcating in me the discipline to undertake this venture.

Summary

This thesis consists of three self-contained research articles that empirically explore survival expectations of older populations and mortality amongst children. Chapter 1 examines methodological considerations in eliciting survival probabilities in India and tests their associations with known socioeconomic characteristics. Chapter 2 investigates the effect of a survey induced negative health information shock on survival expectations of older Britons. Chapter 3 estimates the causal effect of domestic violence on child mortality in India. The analysis uses the pilot study of the Longitudinal Ageing Study of India (LASI) in Chapter 1, the English Longitudinal Ageing Study (ELSA) in Chapter 2 and the Demographic Health Survey (DHS) in Chapter 3. The general timeframe of the analysis is 1991 to 2010. The findings in Chapter 1 suggests that it is feasible to elicit subjective survival expectations in a developing country context and that these expectations correlate in meaningful ways with previously known social and economic predictors. Chapter 2 finds that individuals update subjective survival expectations in response to new information in meaningful ways and in Chapter 3 I find evidence that suggest a causal and significantly positive relationship between domestic violence and child mortality.

Introduction

The world is facing unprecedented demographic changes in the form of an increasing ageing population and a declining youth. This presents serious challenges for both industrialized and developing countries. For industrialised countries, this translates to rising costs in areas of pension and health. Most developing nations do not have comprehensive social care infrastructure in place, which is likely to magnify the impact of an ageing population.

A country's population composition is a result of its birth and death rates. Accordingly, this thesis revolves around two themes; survival expectations of the older population and child mortality. From a policy perspective, survival expectations have been shown to have effects on economic actions such as healthcare expenditure, retirement planning, savings and bequests. Such probabilistic expectations have been shown to measure private information and are increasingly used in empirical work. They have been shown to be highly correlated with mortality predictions which are mere projections of historical data. Chapter 1 and Chapter 2 examine survival expectations in a developing country context and in an industrialized country respectively. On the other hand, as child mortality continues to be a significant global health concern, in Chapter 3, I examine the effect of domestic violence on child mortality.

As the fastest growing population in the world, India is of particular importance in this deliberation. India's population of ages 60 and above is set to dramatically increase over the next four decades according to the United Nations Population Division (United Nations Population Division (UN), 2014). In Chapter 1, we examine several methodological considerations when eliciting probabilistic expectations in a developing country context using the Longitudinal Ageing Survey of India (LASI). We conclude that although on average individuals are able to understand the concept of probability, responses are sensitive to framing effects with regards to the wording of the questionnaire in a survival versus mortality format. Responses are also sensitive to own versus hypothetical person effects where respondents were asked to report their own survival expectations versus the survival expectations of other similar individuals. We also find that overall people are pessimistic about their survival probabilities as compared to state-specific life tables and that socio economic status does influence beliefs about own survival expectations as found in previous literature in other countries. Higher levels of education and income have a positive association with survival expectations and these associations persist even when conditioning on self-reported health. The results remain robust to several alternative specifications. We then compare the subjective survival measures to biomarkers which are measurable indicators of the presence or severity of diseases and are considered objective measures of health. We find that activity of daily life, height and low haemoglobin levels co-vary with subjective expectations in expected directions.

While existing research has established the use of survival probabilities in empirical work we know very little about the formation of these probabilities. The United Kingdom is also considering policy responses to an increasingly ageing population. In 2012, the number of over 65's in the UK surpassed 10 million for the first time (ONS, 2014). In Chapter 2, I use the longitudinal aspect of the English longitudinal Survey of Ageing (ELSA) to examine the effect of a negative health information shock in the form of an out-of-range biomarker result on subsequent survival expectations. I find that individuals who have received a negative health information shock report lower survival probabilities of approximately 8% on average as compared to individuals who did not receive the negative health information shock. I also find that individuals who have been previously diagnosed by a doctor with a cardiovascular or chronic condition report 3.5% and 5.6% lower probabilities of survival, respectively.

India also accounts for a quarter of global child mortality. For instance in the year 2015, the Under-5 mortality rate (U5MR) in India is 48 per 1000 live births while the corresponding rate in the UK is 4 per 1000 live births (United Nations Population Division

(UN), 2015). The current literature has succeeded in establishing an association between domestic violence and child mortality, but has yet to present evidence of a causal relationship. In Chapter 3, I use an instrumental variable approach to analyse the causal impact of domestic violence against the mother on child mortality in the Indian context. Domestic violence is instrumented with the real price of gold at the month of marriage of the mother. I find that domestic violence has a positive and significant effect on child mortality, which suggests that domestic violence continues to be a serious global health concern which impacts national mortality rates.

Chapter 1

Subjective survival expectations of older Indians

1.1 Introduction

India, with 1.27 billion inhabitants, has a growing elderly population. Currently 60 million people are aged 65 and over. By 2050 this figure is projected to climb to 222 million (DESA, 2010; Factbook, 2010). An Indian born in 1950 could expect to live for 36 years (Roser, 2016) while an Indian born today can expect to live for 69 years (The World Bank, 2015). This is a dramatic increase in the elderly dependency ratio and presents serious impending economic and health challenges which are of particular concern given the level of unemployment and poverty in the country and the lack of an effective healthcare or pension system. This increase in life expectancy also has potential important ramifications for the many intertemporal decisions (such as retirement, bequest, investment, saving, migration and healthcare) that elderly individuals have to make. In this paper, we present unique new evidence on the survival expectations of older Indians.

Asking respondents about verbal expectations (e.g. is this event very likely or very unlikely?) is commonly done in surveys, but those yield only ordinal measures of beliefs. Moreover, responses may not be interpersonally comparable. These concerns lead to the elicitation of probabilistic expectations, where respondents are asked a question that can be interpreted as a probability. Manski (2004) and Hurd (2009) review the literature on the elicitation of probabilistic expectations in developed countries, while Delavande et al. (2011) and Delavande (2014) review the parallel literature in developing countries. Both strands of literature emphasize that survey respondents are able and willing to provide their expectations in probabilistic format, that a majority respects basic properties of probabilities, that there is substantial heterogeneity in beliefs, that expectations tend to vary with observable characteristics in the same way as actual outcomes, and that the expectations are useful predictors of future behaviour. However, the existing literature in developing countries has typically focused on younger respondents, and little is known about whether these findings apply to older individuals.

In developed country surveys, the standard method of eliciting subjective probabilities relies on a percent chance format (e.g., *"What is the percent chance that you will live to be 75 or more?"* as in the U.S Health and Retirement Study - HRS) but this method may be challenging in low numeracy contexts. The most common approach in developing countries has been to use visual aids (such as stones, beans or marbles) to help respondents to express probabilities. For example, Delavande and Kohler (2009), in the Malawi Longitudinal Study of Families and Health, asks respondents to choose 10 beans to express the likelihood of an event happening. We adopt a similar approach in India, and asked respondent to express the likelihood of being alive in 1 year, 5 years and 10 years using 10 beans.

Because little is known about the best way to elicit survival expectations from the elderly in developing countries, our design explicitly addresses several methodological considerations to provide information on the best way to collect these expectations, and to assess their validity and usefulness. First, we asked respondents both about their own survival and about the survival of people like themselves, as respondents may be reluctant to think about their own demise. Second, we randomize the wording in terms of mortality or survival to assess any potential framing effect. In order to assess validity, we investigate how the elicited expectations relate to socioeconomic characteristics and to health biomarkers collected as part of the survey. Finally, we evaluate the relationship between the survival expectations and some intertemporal economic decisions.

Our findings show great promise to elicit subjective expectations from elderly in a context like India. First, response rates are high (e.g., about 87% for ones own probability of survival). Second, violation of the monotonicity property of probabilities is similar among older Indians and older Americans. Third, as one would expect, average survival expectations decrease as the time horizon considered increases. Fourth, survival expectations vary with observable characteristics as expected: younger respondents, those with more education, those from higher caste, those with better self-reported health and those with fewer difficulties in their activity of daily living report higher survival expectations on average. Fifth, shorter respondents (an indicator of poor childhood nutrition (Steckel, 1979) and men with decreased haemoglobin concentration (an indicator of anaemia) report a lower survival expectations on average. Finally, respondents who have a higher one-year survival expectation are more likely to have an outstanding loan, consistent with the idea that they are making an investment for the future. However, we also find that respondents are much more pessimistic about their survival than warranted by existing life table estimates, a pattern seen in other contexts (e.g., Malawi (Delavande and Kohler, 2016) or the US (Hurd, 2009)). Women also appear more pessimistic than men, which is also a pattern that has been seen in other contexts (Malawi (Delavande and Kohler, 2009), US (Hurd, 2009), Europe (Peracchi and Perotti, 2009)).

From a methodological point of view, our findings offer some recommendations on the best way to elicit subjective expectations from older respondents in a context like India. First, response rates are not significantly improved by asking about the survival of people like you instead of own survival. However, there can be large differences in answers, driven by perception of own health. Researchers interested to learn about respondents' own survival should therefore ask about it directly. Second, framing the question in terms of survival or mortality influences respondents' answers. The mean and median subjective survival probability when asked the survival format is higher than when asked the mortality format for both own survival and hypothetical person survival. For the longer time horizons, the difference is quite substantial (for example, 11 percentage points for the 10-year own survival). Once we control for other covariates, this framing effect is observed for the 10-year time horizon only, when uncertainty is likely to be larger, suggesting that responses to expectations questions are reasonable.

This paper complements the existing literature investigating individuals survival expectations. Very few studies have investigated subjective survival expectations in developing countries. Delavande and Kohler (2009) look at survival expectations in Malawi. Like in India, the reported subjective expectations about mortality correspond in broad terms with the actual variation in mortality (e.g., respondents living in regions with higher mortality risks have higher mortality expectations) but they are widely over-estimated. Aguila et al. (2014) report the results of various cognitive interviews to assess the best way to elicit survival expectations from older Mexican, and emphasize the usefulness of visual aids. In the US context, a number of in-depth studies have been conducted using the subjective expectations from the HRS. They appear well-calibrated on average, vary systematically with known risk factors and evolve in panel in response to information relevant to survival, such as parental death or onset of disease. For instance, Hurd and McGarry (1995) show that survival expectations are internally consistent and are good approximations to population probabilities. Schoenbaum (1997) compares the subjective survival expectations of smokers to smoking-specific life tables from nationally representative data on the United States and find that survival expectations were close to actuarial predictions. Subjective survival expectations are also found to be predictive of actual survival (Bloom et al., 2006; Delavande and Rohwedder, 2011; Elder, 2007; Hurd and McGarry, 2002; Perozek, 2008). Similar findings have been reported based on subjective probabilities of survival elicited in the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Ageing and Retirement in Europe (SHARE) (e.g.,(Balia, 2014; Delavande and Rohwedder, 2011; Hurd et al., 2004; Menon, 2015; Winter, 2008)).

The remainder of the paper is organised as follows: Section 1.2 presents the data and the expectations module. Section 1.3 examines the methodological considerations when eliciting subjective expectations from older individuals in a context like India. Section 1.4 examines the relationships between health measures and subjective expectations in India, while Section 1.5 assesses whether subjective survival expectations are predictive of investment in the future. Section 1.6 concludes.

1.2 Data Description

1.2.1 Longitudinal Ageing Survey of India (LASI)

We use data collected in the Longitudinal Ageing Survey of India (LASI) pilot survey, which was fielded between October and December 2010. The LASI pilot was funded by the National Institute on Ageing (NIA) and is a collaboration between three main institutions, the International Institution for Population Sciences (IIPS), RAND Corporation and Harvard School of Public Health (HSPS). LASI collected data on health, retirement, economic and social well-being of India's elderly population. The LASI instrument was developed to be internationally comparable to the HRS of the United States and is harmonized to other surveys such as the China Health and Retirement Longitudinal Study. Using the 2001 Census, a geographically diverse sample was drawn from 4 states in India: Karnataka, Kerala, Rajasthan and Punjab.[1] LASI consists of a household survey, collected once per household and an individual survey for each age-eligible respondent who is at least 45 years of age and their spouse. The LASI pilot achieved an individual response rate of 90.9%. The total individual sample size is 1683 respondents within 950 households out of whom 1486 are aged 45 years or older. The survey was fielded in Hindi, Kannada, Malayalam and Punjabi, the local native language of each of these states. Computer-assisted personal interviews (CAPI) which lasted for approximately 2 hours were conducted.

1.2.2 Expectations Module

LASI implemented an expectations module to a randomly selected 33% of the total number of respondents. This module included questions about subjective probabilities of survival to specific ages. Respondents were given preliminary training questions to introduce them to the concept of probability. Out of the 1486 age-eligible respondents, 531 were asked the expectations module. Out of these, 467 respondents are equal to or above the age of 45 years, which is our analytical sample. The expectations module took an average of 5 minutes to complete. The module uses an interactive elicitation technique based on asking respondents to allocate up to 10 beans on a plate to express the likelihood that an event will be realised (Delavande and Kohler, 2009). Prior to eliciting subjective survival probabilities, the respondents were given an explanation of basic probability concepts and given the following introduction:

I will ask you several questions about the chance or likelihood that certain events are going to happen. There are 10 beans in the cup. I would like you to choose some beans out of these 10 beans and put them in the plate to help me understand what you think the likelihood or chance is of a specific event happening. If you do not put any beans in the plate, it means you are sure that the event will NOT happen. If you add beans, this means that you think the likelihood that the event happens will increase.

[1] An overview of LASI can be found at International Institute for Population Sciences (2010) and Harvard School of Public Health (2010). A detailed external validity check has been done in Arokiasamy et al. (2012)

For example, if you put one or two beans, it means you think the event is not likely to happen but it is still possible. If you pick 5 beans, it means that it is just as likely it happens as it does not happen (fifty-fifty). If you pick 6 beans, it means the event is slightly more likely to happen than not to happen. If you put 10 beans in the plate, it means you are sure the event will happen. One bean represents one chance out of 10. There is not a right or wrong answer; I just want to know what you think.

Our analysis focuses on survival expectations. Respondents were asked about their survival in 1 year, 5 years and 10 years. There were two important features of the design that is relevant from a methodological point of view. First, the wording of the questions in terms of survival (alive) or mortality (not alive) was randomized. Second, all respondents were asked both their own survival expectations and survival expectations of a hypothetical individual like themselves. Out of the 467 age-eligible respondents that answered the expectations module, 239 were asked mortality questions while 228 were asked survival questions. The questions were organized and worded as follows:

1. Mortality wording

   I would like to ask you to consider the likelihood that you and other people may not be alive as time goes by. Think about 10 people like you (same age, gender, income, etc). Pick the number of beans that reflects how many

   (a) *Will die within a one-year period beginning today.*

   (b) *Will die within a 5-year period beginning today.*

   (c) *Will die within a 10-year period beginning today.*

   Now, I would like to ask you to consider the likelihood that you may not be alive as time goes by. We hope that nothing bad will happen to you, but nevertheless, something unfortunate may occur over the next years despite all precautions that you may take. If you dont want to, you do not need to answer this question. Pick the number of beans that reflects how likely you think it is that

   (a) *You will die within a one-year period beginning today.*

   (b) *You will die within a 5-year period beginning today.*

(c) *You will die within a 10-year period beginning today.*

2. Survival wording

I would like to ask you to consider the likelihood that you and other people may be alive as time goes by. Think about 10 people like you (same age, gender, income, etc). Pick the number of beans that reflects how many

   (a) *Will be alive in one year.*

   (b) *Will be alive in 5 years.*

   (c) *Will be alive in 10 years.*

   Now, I would like to ask you to consider the likelihood that you may be alive as time goes by. We hope that nothing bad will happen to you, but nevertheless, something unfortunate may occur over the next years despite all precautions that you may take. If you dont want to, you do not need to answer this question. Pick the number of beans that reflects how likely you think it is that

   (a) *You will be alive in one year.*

   (b) *You will be alive in 5 years.*

   (c) *You will be alive in 10 years.*

1.2.3 Demographics characteristics of the analytical sample

Out of the 467 age eligible respondent who was asked the expectations module, 391 respondents have full information on all demographic variables of interest. Our analysis using demographic controls is thus restricted to these 391 respondents to ensure results are not driven by differing sample compositions. Table 1.1 presents the demographic composition of this analytical sample: respondents who were selected to answer the expectation module. Male respondents and female respondents are almost equally represented in the sample. 46% of our sample are between 45 years and 54 years of age and 8% of the sample are above 75 years of age with the oldest respondent being 96 years of age. 37% of the analytical sample belongs to the high/other caste community with the rest being divided into each of the three lower caste communities. 46%

of the sample has no schooling. There are gender differences in educational attainment with males having an overall greater educational attainment than females, which is consistent with the gender differences in the national representation of educational attainment. The income variable used is self-rated by the respondent in answer to the question: *"Compared to other households in this (geographic) community, how do you consider your household?"* The responses were recorded in income quintiles. The top two quintiles of well off and very well off have been collapsed due to small numbers in these categories. There is an almost equal representation from each of the 4 states that were surveyed. For 7% of the respondents, both parents are alive at the time of the survey. Respondents were also asked about their general health with the question framed as *"Overall, how is your health in general? Would you say it is very good, good, fair, poor, or very poor?"* 62% of the respondents report their health as being good or very good.

1.3 Can we ask survival expectations of older respondents in low-income countries? Methodological considerations

In this section we review the methodological considerations to be taken into account when eliciting survival probabilities from an older population in a developing country. We use the age-eligible sample of 467 respondents for this section to take advantage of the larger sample size and to enable reporting of response rates. Note that respondents are willing to report their beliefs in probabilistic formats: response rates are high for own probability of survival, about 87%.

1.3.1 Do older respondents understand the concept of probabilities?

After reading the introduction, the interviewers checked if the respondents understood the concepts of probability with some practice questions.

Respondents were asked to pick the number of beans that reflects the probability of going to the market within 2 days and within 2 weeks to assess whether they would respect the monotonicity property of nested events. 447 out of the 467 respondents answered this question which translated to a 4% non-response rate. Figure 1.1 presents the difference in the probability of going to the market within 2 weeks and the probability of going to the market within 2 days for 447 of the respondents for whom we have complete data. A negative statistic is a violation of the monotonicity criterion of nested events which occurs in 21% of the sample. This is consistent with previous studies eliciting subjective expectations among the elderly in developed countries (approximately 23% in the survival expectations questions in the HRS in the US). Out of this 54% of the sample who violated monotonicity has no schooling. This is more than what has been reported in other developing countries with a younger sample. For example Delavande and Kohler (2009) find that 1.41% of their sample in Malawi violates monotonicity when asked the probability of going to the market in the first instance.

The respondents, who violated the criterion, were subsequently given the following information:

"Remember, as time goes by, you may find more time to go to the market. Therefore, there is a higher chance that you go to the market within 2 weeks than within 2 days. So you should put more beans for the likelihood of going to the market within 2 weeks than within 2 days. Let me ask you again."

These respondents were then asked the question regarding the probability of going to the market again. Only 20 of the 447 respondents continue to violate the criterion.

Respondents were also asked a question to assess whether they understand that complementary events have a probability summing up to one. In particular, in the context of a game of Ludo, the questions were:

"Pick the number of beans that reflects how likely you think it is that

- *You will win the game*
- *You will lose the game"*

423 out of the 467 respondents answered this question. 35% of this sample correctly as-

signed probabilities to each outcome so that the sum of both would equal to one. Since respondents are given 10 beans where each bean represents 10% likelihood, it is plausible that respondents may be rounding their actual probabilities (Manski and Molinari, 2010). The sum of the proportion of the sample that report probabilities between +1 and -1 is 52%. Note that we are not aware of other surveys asking similar questions about complementary events, so we do not have a benchmark. Overall, respondents seem more familiar with the idea of monotonicity than complementarity.

1.3.2 Survival expectations by time horizon

In order to compare the various formats used to elicit expectations, we recode the mortality responses into survival and express all responses in survival terms on a scale from 0 to 1. Figure 1.2 shows the distribution of all 3 survival periods. It shows that respondents are aware that survival probability decreases as the time horizon increases. For example, the percentage of respondents who report a survival probability of 1 in the 1-year period is 24% as compared to 14% of respondents who report the same for the 10-year survival period. When looking at monotonicity violations in the survival expectations responses at the individual level, we find that on average 25% of the respondents violate among all three time periods. Specifically when comparing the 1-year survival period to the 5-year survival period, 26% of respondents violate monotonicity. When comparing the 5-year to the 10-year survival period and the 1-year to the 10-year survival period, monotonicity violations are at 22% and 27% respectively.

Heaping at 0.5 is also a common feature of subjective expectations. Previous studies have shown that expectations of 0.5 may be indicative of epistemic uncertainty (e.g., de Bruin et al., 2000). This is consistent in the LASI data with uncertainty increasing as the time horizon increases. Respondents are more likely to report 0.5 in the 5-year and 10-year survival periods of 19.27% and 20.25% respectively as compared to 16.22% in the 1-year survival period.

1.3.3 You versus Other people like you

In addition to asking respondents to report their own survival expectations, they are asked to think of 10 people like themselves and to report the survival expectations for these hypothetical individuals. This has been done in previous studies for example in McKenzie et al. (2008) and Aguila et al. (2014). The potential advantage of this wording of the question is to improve response rates, as people may be less reluctant to think about the mortality of others. However, it is conceptually a different expectation than ones own expectations: answers may vary between own and hypothetical person survival as respondents may make unobservable assumptions about the characteristics of the hypothetical individuals. Researchers interested in explaining how mortality expectations influence *individual* decision-making want to elicit respondents' own expectations. There may be trade-off between better response rates and precise survival estimates (see discussion in Delavande (2014)). Table 1.2 presents the summary statistics of own survival and hypothetical survival. Response rates are only slightly higher for hypothetical persons survival probability as opposed to own survival probability, and not statistically significantly different. Regarding the average levels of expectations, beliefs about own survival relative to a hypothetical persons survival are similar. The unpaired t-test for equality in means between own survival and the survival of a hypothetical person is not significant in all three time frames.

We further investigate the difference of expectations at the individual levels. Table 1.3 presents the summary statistics of the difference in the responses between own survival and hypothetical survival for those respondents who had different answers. About 55% of the respondents report a different answer. The differences in the responses are quite small on average and vary between -0.01 and 0.02 in the three survival periods but the percentiles show that they can be large for some individuals. For example, the 25th and 75th percentiles correspond to a very large difference of 20 percentage points.

In Table 1.4, we seek to evaluate whether individual characteristics and self-reported health are predictive of the difference between own survival and that of the hypothetical individual. For this analysis we restrict the sample to respondents whose responses differed between the own survival and hypothetical survival wording of the questionnaire. Table 1.4 presents the OLS coefficients using the difference in beliefs as dependent variables. Demographic and socio-economic characteristics have essentially no predictive power for this difference. However, as one would expect, respondents with relatively poor self-reported health status are also likely to report lower survival probabilities for themselves as compared to a hypothetical individual.

Overall, in the context of this study, response rates are not significantly improved by asking about the survival of people like you instead of own survival. However, there can be large differences in answers, driven by the perception of own health in that respondents who view their health as poor in general tend to report lower survival probabilities for themselves as opposed to a hypothetical person. Researchers interested to learn about respondents' own survival should ask about it directly.

1.3.4 Mortality versus Survival

Previous studies have shown that framing can have an effect on survey responses (e.g.,Tversky and Kahneman, 1981). Studies examining the framing effect specifically on survival and mortality format of questionnaires have shown mixed results. Some studies fail to find a significant effect (e.g., Miller and Fagley, 1991) while some studies find a significant effect that is small in magnitude (Levin et al. (1998) provides an overview).

Respondents in the expectations module of LASI were randomised between the survival format of the question and the mortality format of the question (see Section 2.2). Table 1.5 presents the summary statistics for the mortality versus survival format of the questions. To enable comparison, responses to the mortality format of the questionnaire have been re-coded in survival terms. For our age-eligible sample of 467 respondents, 239 were asked the mortality format of the question while 228 were asked the survival format of the question. The question format does not seem to systematically influence response rate as the difference in response rate is not statistically different across the two formats. However, respondents reported feeling a bit uncomfortable talking about their own mortality to interviewers.

The mean and median subjective survival probability when asked the survival format is higher than when asked the mortality format for both own survival and hypothetical person survival. For the larger time horizons, the difference is quite substantial (for example, 11 percentage points for the 10-year own survival). The t-test for equality of means between mortality and survival format of the questionnaire is significant in the 5 year and 10 year survival period for own survival and in the 10 year survival period for the survival of a hypothetical individual. There is therefore a framing effect for longer time horizons, with respondents allocated in the mortality format being more pessimistic about survival than those allocated in the survival format. Once we control for other covariates, the framing effect is observed for the 10-year time horizon only, when uncertainty is likely to be larger (see discussion in Section 1.4.1).

1.3.5 Do subjective probabilities of survival vary by socioeconomic characteristics?

We now investigate whether the subjective probability of survival vary with socio-economic characteristics similarly as actual survival is known to vary with those. Table 1.6 presents the mean subjective probability of own survival and hypothetical person survival by characteristics. Means are weighted by the pooled individual weight to provide survey design adjusted standard errors across the four states. There are a few important remarks based on this table. First, as already shown in Section 3.2, survival expectations decrease as the time horizon considered increases: for example, the difference in survival subjective probability within 1 year and within 10 years is 10 percentage points. Second, survival expectations decrease as age increases, in almost all cases. For example, respondents aged 45 to 54 expect a 63% chance of being alive in the next 5 years on average while the 75+ expect a 51% chance. Third, there is a clear caste and education gradient in the responses. High caste respondents report higher expectations of own survival in all 3 time horizons. Respondents with at least a high school education report higher survival expectations in all three time horizons for both own survival and the survival of a hypothetical person. Fourth: women and men have similar levels of expectations, while women have larger life expectancy (male life expectancy at birth is 63 years while female life expectancy at birth is 66 years, World

Health Organization, 2011).

The survival responses according to the income category of the respondent are mixed. Respondents from a household with income well below average report lower survival probabilities, as do the really well off respondents. A possible explanation is that very high income respondents may be more health literate and so may adjust their survival expectations accordingly (Bloom, 2005). There is considerable heterogeneity between states with Karnataka reporting lower survival responses than the other states in all three time periods and for own and hypothetical person survival.

1.3.6 Are survival expectations accurate?

To further assess the validity of respondents survival expectations, we compare subjective survival expectations to life table estimates based on the Sample Registration System (SRS) and published by the Government of India (Sample Registration System, 2012). The SRS is a large scale demographic survey based on a dual recording system which provides reliable mortality estimates at state and national levels. Abridged life tables are created using the mortality package MORTPACK 4, the UN's software package for mortality measurements. For the purpose of our analysis, we use the revised life table reports for the period 2006-2010. The comparison is therefore not completely ideal to assess accuracy as we are comparing 2010 life tables with prospective survival, but is still a useful exercise.

The top panel in Table 1.7 presents the state life table estimates for the 5 year and 10 year survival periods for the overall sample and the state specific life table estimates. The lower panel presents the overall and state wise subjective survival estimates.

There are two important things to note from this table. First, people are much more pessimistic about their survival probabilities than is warranted by existing life table estimates. Overall, respondents report a 61% chance of being alive in the next 5 years while equivalent life table statistic is 84%. On average, the survival expectations of men are closer to the life table estimates as compared to women (not shown). If anything, one would expect life expectancy not to deteriorate in the coming years, so this difference is unlikely to be driven by respondents being forward-looking and predicting a reduction in life expectancy. A similar phenomenon has been observed in Malawi (Delavande and Kohler, 2009). Second, individuals seem unaware of the protective effect of residing in certain states. In the state life table estimates, there is a clear ordering of the survival forecasts i.e. Karnataka, Kerala, Punjab and Rajasthan in order of decreasing survival forecast. For both the 5-year and 10-year survival periods, Karnataka has the highest survival probabilities as reported by the state life table estimates while Rajasthan has the lowest survival probabilities. With respect to subjective survival probabilities there does not exist such a clear ranking. Kerala has the highest survival expectation in the 5-year period while Rajasthan has the highest in the 10-year period. Karnataka has the lowest survival expectation in the 5-year period and in the 10-year period.

1.4 Health measures and subjective survival expectations

In this section we evaluate how various measures of health are correlated with the elicited survival expectations. We focus on self-reported health, activities of daily living and objective biomarkers.

1.4.1 Self-reported health

Self-reported health has been shown to be good predictors for mortality (Burström and Fredlund, 2001; Idler and Benyamini, 1997). In the context of India, self-reported health measures have been shown to be reliable measures of health when estimates are conditioned on region (Chen and Mahal, 2010).

Table 1.8 presents an OLS regression investigating the predictive power of self-reported health, after conditioning on demographic characteristics. As a comparison, the first 3 columns shows results when we only control for demographic characteristics.

As already seen in Section 3, some indicators of socio-economic status and state of residence are correlated with beliefs. Also, having one or both parents dead is associated with lower probability of survival. As seen in columns 4 to 6, self-reported health status has a negative relationship with survival probabilities, which is statistically significant in all 3 time horizons. The magnitude of the effect is very large: for example, those who rate their health as very poor have a subjective probability of survival, which is 0.36 point lower compared to those who rate their health as very good. Survival expectations are therefore in line with self-reported health, even after conditioning for other characteristics.[2]

Interestingly, we also find a framing effect of the questionnaire format (survival vs. mortality) in the 10-year survival period of 0.11. This suggests that respondents are more influenced by the framing of the questionnaire when the survival time period in question is longer, and therefore when there is presumably more uncertainty.

1.4.2 Activity of Daily Life

LASI also collected self-reported disability rates measured by difficulty with at least one activity of daily life (ADL). Self-reported measures have previously been shown to be reliable measures of health in India (Subramanian et al., 2009). Table 1.9 presents the proportion of respondents within the analytical sample in each measure of the ADL who reported having a difficulty. We coded a factor score of ADL using the above measures through a principle component analysis. A high score on the ADL thus means the respondent does not have a difficulty in any of the 6 ADLs while a low score indicates that the respondent has difficulties in one or more of the 6 ADLs. The top panel in Table 1.10 presents the estimates of the association between subjective survival probability and self-reported measures of ADL. These are based on OLS regressions similar to those in the first 3 columns of Table 1.8. Each cell in Table 8 reports the results of separate estimations with all the control variables used in the main specification. Subjective survival probabilities in the 1-year time horizon and 5-year time horizon is positively correlated with ADL measures with a coefficient of and 0.03 and 0.02 respectively.

[2]Recoding responses of 0 to 0.1 and responses of 1 to 0.9 does not alter our results significantly

1.4.3 Biomarkers

LASI included a biomarker content, which includes anthropometric measures, blood pressure readings, vision and physical functioning test, and a collection of dried blood samples (Bloom et al., 2014). This data allows us to compare subjective survival expectations in India with objective measures of health collected through the direct assessment of biomarkers. Among the 1683 individuals interviewed for LASI, 1311 completed the biomarker module which translates to a 77.9% completion rate.

The second panel in Table 1.10 shows the association between high blood pressure and survival expectations. High blood pressure is a binary indicator with 1 identifyiong respondents who have blood pressure 140/90mmHg or higher. We find a negative but insignificant relationship in all three time periods.

Several studies have established an association between height, early life nutritional status, morbidity and mortality (Bhalotra and Rawlings, 2011; Monden and Smits, 2009). The average height of men in our analytical sample is 165.5cm and for women it is 153.1cm. The third panel in Table 1.10 presents the association between height and subjective survival probability of the respondents. There exists a positive relationship between height and survival probability all three time horizons with a magnitude of 0.005.

Decreased haemoglobin concentrations are an indicator for anaemia which is highly prevalent in developing countries. Lower levels of haemoglobin have been shown to predict mortality and morbidity (Guralnik et al., 2004; Tolentino and Friedman, 2007). In LASI, haemoglobin levels were measured using an ELISA (Enzyme-Linked Immunosorbent Assay) protocol based on the O'Broin and Gunter method (D O'Broin and Gunter, 1999). Blood haemoglobin levels are measured in grams (g) per decilitre (dl). Mean haemoglobin level for our analytical sample is 14.3 g/dl. We create a binary indicator for low haemoglobin levels based on standard clinical cut points of 12.0g/dl for women and 13.0g/dl for men (World Health Organization, 2001). 19% (66 respondents) of our analytical sample have low haemoglobin levels out of whom 73% (48 respondents) are women.

We find a strong negative association between low haemoglobin concentrations and subjective survival expectations in the 1 year and 5 year period for men with magnitudes of 0.14 and 0.17 respectively as shown in Table 1.11. We find no significant effects for women.

1.5 Survival expectations and expenditure

In this section we explore the association between survival expectations and some economic decisions of the respondents to evaluate whether survival expectations are correlated with forward looking decisions where how long one expect to leave should matter. We use two dependent variables in our analysis; savings and loans. We expect people with higher survival expectations to be more likely to have a loan (i.e. they are making investments) and to have higher savings. Respondents were asked to provide an approximate value of savings accounts, postal accounts and certificates of deposits.[3] The summary statistics for these variables are provided in the last panel of Table 1 in Section 2.3. We drop the top 1% of the data (N=1) to reduce the effect of outliers. The average value of savings reports was INR 35,521 with a standard deviation of INR 66,923. Bank loan is a binary variable with 1 indicating that the respondent has an outstanding loan from a bank. 13% (50 respondents) of the analytical sample reported having an outstanding bank loan. Table 1.12 presents the results of the association between survival expectations, outstanding loans and savings. An increase in the 1-year survival expectation is positively associated with a 0.12 percentage point increase of having an outstanding bank loan. We created quintiles of savings and use this as the dependent variables in the last three columns in Table 1.12. We find a negative association between savings and survival expectations in the 1-year time period and a positive association in the 5-year and 10-year time period but with no statistical significance. These results are not sensitive to coding savings in quartiles or deciles.

[3] Savings is a sub-component of an answer to the question: Do you or members of your household possess any of the following financial assets? 1. Current accounts 2. Savings accounts 3. Stocks or mutual funds 4. Bonds 5. Outstanding balances in kitty parties, chit funds, bishi etc. 6. Other 7. None of the above. Thus our sample size is lower as it only includes the savings sub-component of this response with 0 indicating respondents who do not report any savings.

1.6 Conclusion

This paper is a thorough investigation of older individuals' subjective survival expectations in India. We inspect several methodological contemplations with regards to eliciting subjective survival expectations in the developing country context. We conclude that although on average individuals are able to understand the concept of probability, responses are sensitive to framing effects and own versus hypothetical person effects. This evidence suggests that researchers interested in respondents own survival should ask about it directly. We also find that overall people are pessimistic about their survival probabilities as compared to state specific life tables. This suggests that respondents are not aware of the past increases in longevity and so are not accounting for this in their survival expectations.

Next, we examine socio-economic gradients in the Indian context for three time periods of survival; 1-year, 5-year and 10-year survival. We find that socio economic status does influence beliefs about own survival expectations as found in previous literature in several other countries. Higher levels of education and income have a positive association with survival expectations and these associations persist even when conditioning on self-reported health. There are significant state level differences in survival expectations. The results remain robust to several alternative specifications.

We then compare the survival measure to objective measures of health. The distinct advantage of anthropometric and biomarker data is that they are objective markers of health and free from respondent reporting errors. We find that activity of daily life, height and low haemoglobin levels co-vary with subjective expectations in expected directions. We also find that survival expectations are predictive of investments for the future. Overall, our findings suggest that researchers can ask subjective expectations of older survey respondents in a context like India.

These findings suggest that it is worthwhile to continue to elicit survival expectations in LASI and invest in a longitudinal perspective of these expectations in India. The implications of this for research are also promising in that future analysis will be able to test for the associations between survival expectations and actual survival as has been done in the US. Given the current agenda in India to put an affordable healthcare system in place, survival expectations could also prove to be a useful measure of future mortality statistics as an alternative to life tables.

1.7 Figures and Tables

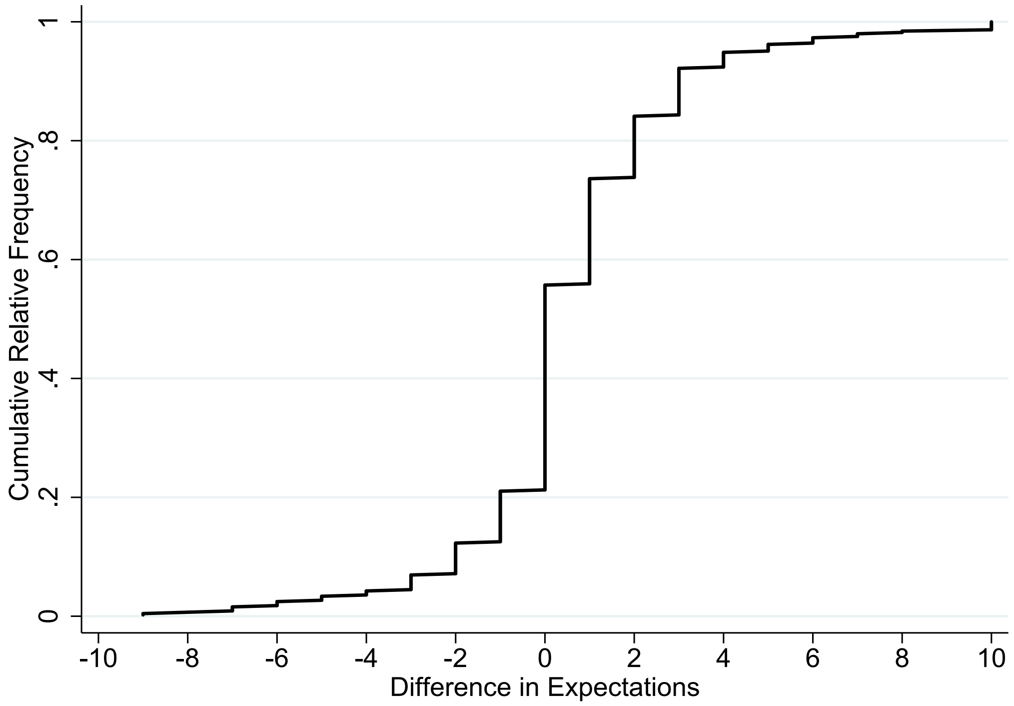

Figure 1.1: Difference in expectations of going to the market

Notes: This figure plots the cumulative relative frequencies of the difference in expectations of going to the market within 2 weeks and 2 days for a respondent. Expectations are recorded on a scale of 0- 10 and so the difference is recorded on a scale of -10 to +10. A negative statistic is a violation of the monotonicity criterion of nested events which occurs in 21% of the sample.

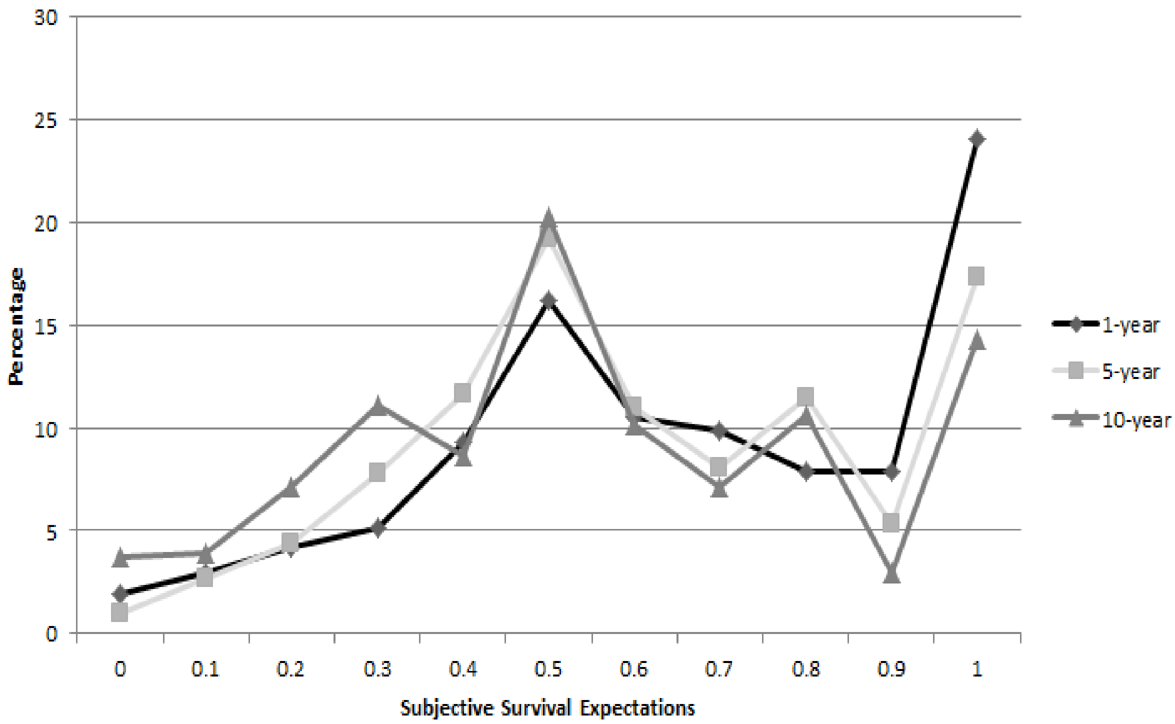

Figure 1.2: Distribution of subjective probability of Survival (weighted)

Notes: This figure plots the subjective survival expectations weighted by pooled individual weights for the three time horizons of 1-year survival, 5-year survival and 10-year survival.

Table 1.1: Summary statistics

Variable	Categories	N	Mean
Gender	Male	391	0.50
	Female	391	0.50
Age	45 - 54	391	0.46
	55 - 64	391	0.29
	65 - 74	391	0.17
	75+	391	0.08
Caste	scheduled caste	391	0.14
	scheduled tribe	391	0.12
	other backward class	391	0.37
	Other Caste	391	0.37
Education	no schooling	391	0.46
	primary/middle school	391	0.36
	high school or more	391	0.18
Income	well below average	391	0.16
	below average	391	0.29
	about average	391	0.47
	well off	391	0.08
State	Punjab	391	0.26
	Rajasthan	391	0.24
	Kerala	391	0.25
	Karnataka	391	0.25
Survival Format		391	0.50
Either parent is dead		391	0.93
Self-reported Health	Very Good	391	0.02
	Good	391	0.6
	Fair	391	0.31
	Poor	391	0.06
	Very Poor	391	0.02
Objective Measures of Health	Activity of Daily Life	319	0.00
	Height (cm)	370	159.08
	High Blood Pressure	309	0.19
	Undiagnosed High BP	308	1.81
	Anaemia	343	0.19
Financial Variables	Savings (INR)	129	35520.54
	Outstanding bank loan	391	0.13

Survival format is a binary indicator where 1 is survival framing of the survival expectations questions and 0 is the mortality framing of the survival expectations questions
Activity of Daily Life is coded as a principle component index of 6 factors which are detailed in Table 1.9
Height is measured in centimeters
Savings is measured in Indian Rupees
All other covariates are binary indicators

Table 1.2: Summary statistics of subjective survival expectations in own versus hypothetical person survival

	Own Mortality			**Hypothetical Person Mortality**		
Stats	1 years	5 years	10 years	1 years	5 years	10 years
Mean	0.65	0.61	0.55	0.63	0.61	0.56
p25	0.60	0.60	0.50	0.60	0.60	0.50
p50	0.50	0.40	0.30	0.50	0.40	0.40
p75	0.90	0.80	0.80	0.90	0.80	0.80
p-values of unpaired t-test for equality of means*				0.32	0.89	0.73
N	407	410	405	423	420	420
Response Rate	0.87	0.88	0.87	0.91	0.90	0.90
p-values of unpaired t-test for equality of response rates**				0.10	0.30	0.13

*Unpaired t-test for equality of means between own versus hypothetical survival
**Unpaired t-test for equality of response rates between own versus hypothetical survival

Table 1.3: Summary Statistics of subjective survival expectations in difference between own survival and hypothetical person survival

stats	1 years	5 years	10 years
Proportion with different responses	53.96	55.03	55.89
mean	0.03	0.01	-0.01
p50	0.10	0.10	0.10
p25	-0.20	-0.20	-0.20
p75	0.20	0.20	0.20
N	252	257	261
p-values of unpaired t-test for mean different from zero	0.91	0.90	0.38

Table 1.4: Difference between subjective survival expectations in own survival and hypothetical person survival

	Basic Reg			Baisc Reg + self health controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	1-year survival	**5-year survival**	**10-year survival**	**1-year survival**	**5-year survival**	**10-year survival**
Male						
Female	-0.012	0.051	0.048	-0.01	0.052	0.049
	(0.052)	(0.036)	(0.037)	(0.049)	(0.036)	(0.039)
45-54 years						
55-64 years	0.005	0.029	0.018	0.018	0.032	0.023
	(0.047)	(0.044)	(0.051)	(0.047)	(0.045)	(0.050)
65-74 years	0.035	0.019	-0.026	0.076	0.051	-0.022
	(0.063)	(0.063)	(0.061)	(0.068)	(0.065)	(0.064)
Over 75 years	0.068	-0.087	-0.019	0.139**	-0.05	-0.005
	(0.065)	(0.066)	(0.070)	(0.067)	(0.059)	(0.072)
No Schooling						
Primary/Middle School	-0.01	0.007	-0.003	-0.006	0.02	-0.003
	(0.061)	(0.051)	(0.054)	(0.056)	(0.052)	(0.053)
High School or more	0.059	0.062	0.048	0.052	0.046	0.055
	(0.087)	(0.079)	(0.067)	(0.082)	(0.077)	(0.067)
Other Caste						
Schedule Caste	-0.04	0.05	0.043	-0.044	0.041	0.039
	(0.084)	(0.081)	(0.073)	(0.069)	(0.081)	(0.072)
Schedule Tribe	-0.041	-0.021	0.023	-0.026	0.01	0.032
	(0.077)	(0.073)	(0.068)	(0.071)	(0.070)	(0.069)
Other Backward Caste	0.061	-0.019	-0.03	0.06	-0.029	-0.033
	(0.058)	(0.056)	(0.053)	(0.057)	(0.053)	(0.054)
Mortality Format						
Survival Format	-0.003	0.025	0.017	-0.018	0.021	0.012
	(0.044)	(0.033)	(0.039)	(0.043)	(0.031)	(0.037)
Both parents alive						
One or both parents are dead	-0.074	-0.024	-0.035	-0.109	-0.043	-0.028
	(0.094)	(0.087)	(0.076)	(0.094)	(0.074)	(0.070)
Income-Well below Average						
Income - Below Average	0.154**	-0.003	-0.021	0.174**	-0.002	-0.003
	(0.064)	(0.067)	(0.069)	(0.058)	(0.071)	(0.068)
Income - About Average	0.088*	-0.033	-0.089	0.111**	-0.035	-0.072
	(0.051)	(0.072)	(0.058)	(0.043)	(0.074)	(0.060)
Income - Well Off	-0.021	-0.033	-0.065	-0.033	-0.046	-0.06
	(0.064)	(0.093)	(0.077)	(0.057)	(0.094)	(0.077)
Punjab						
Rajasthan	-0.061	0.033	-0.073	-0.034	0.026	-0.074
	(0.069)	(0.066)	(0.053)	(0.070)	(0.065)	(0.058)
Kerala	0.023	0.055	-0.021	0.11	0.087	-0.022
	(0.074)	(0.055)	(0.063)	(0.076)	(0.073)	(0.070)
Karnataka	-0.058	-0.084	-0.132**	-0.028	-0.074	-0.122**
	(0.062)	(0.064)	(0.052)	(0.066)	(0.066)	(0.056)
Self Health-Very Good						
Self Health-Good				-0.510***	-0.11	-0.229**
				(0.061)	(0.127)	(0.093)
Self Health-Fair				-0.520***	-0.09	-0.219**
				(0.069)	(0.142)	(0.100)
Self Health-Poor				-0.743***	-0.336**	-0.248*
				(0.117)	(0.160)	(0.140)
Self Health-Very poor				-0.768***	-0.216	-0.055
				(0.187)	(0.176)	(0.126)
Cons	0.015	-0.004	0.092	0.518***	0.117	0.291**
	(0.102)	(0.143)	(0.116)	(0.090)	(0.154)	(0.140)
N	243	247	253	243	247	253

Regressions weighted by the pooled individual weights to provide survey design adjusted standard errors. Robust standard errors clustered at state level in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All covariates are coded as binary indicators.
The dependent variable is defined as the difference between subjective expectation of survival reported for oneself versus that reported for a hypothetical person.

Table 1.5: Summary statistics of subjective survival expectations in mortality format versus survival format

Mortality Format						
	Own Mortality			**Hypothetical Person Mortality**		
Stats	1 years	5 years	10 years	1 years	5 years	10 years
Mean	0.64	0.58	0.50	0.62	0.60	0.51
p25	0.60	0.50	0.50	0.60	0.50	0.50
p50	0.40	0.40	0.30	0.40	0.40	0.40
p75	0.90	0.80	0.70	0.90	0.80	0.70
N	205	203	203	214	214	213
Response Rate	0.86	0.85	0.85	0.90	0.90	0.89
Survival Format						
	Own Survival			**Hypothetical Person Survival**		
Stats	1 years	5 years	10 years	1 years	5 years	10 years
Mean	0.67	0.65	0.61	0.65	0.62	0.61
p25	0.70	0.70	0.60	0.60	0.60	0.60
p50	0.50	0.50	0.50	0.50	0.50	0.50
p75	0.10	0.90	0.80	0.90	0.80	0.80
N	202	207	202	209	206	207
Response Rate	0.89	0.91	0.89	0.92	0.90	0.91
p-values of unpaired t-test for equality of means*	0.21	0.01	0.00	0.35	0.34	0.00
p-values of unpaired t-test for equality of response rates**	0.84	0.79	0.95	0.74	0.60	0.69

*Unpaired t-test for equality of means between mortality and survival format of the questionnaire

**Unpaired t-test for equality of response rates between mortality and survival format of the questionnaire

Table 1.6: Mean Subjective expectations of survival (weighted)

		Own Survival			Hypothetical Person Survival		
		1-year survival	5-year survival	10-year survival	1-year survival	5-year survival	10-year survival
ALL		0.65	0.61	0.55	0.63	0.61	0.56
men		0.65	0.62	0.56	0.63	0.62	0.58
women		0.65	0.61	0.55	0.64	0.60	0.55
age	45 54	0.65	0.63	0.57	0.63	0.63	0.58
	55 64	0.69	0.64	0.58	0.67	0.62	0.56
	65 74	0.64	0.59	0.48	0.60	0.56	0.53
	75+	0.60	0.51	0.48	0.55	0.57	0.50
caste	scheduled caste	0.60	0.59	0.53	0.60	0.59	0.51
	scheduled tribe	0.59	0.59	0.56	0.61	0.59	0.59
	other backward class	0.66	0.61	0.55	0.62	0.62	0.58
	Other Caste	0.68	0.63	0.56	0.66	0.62	0.55
education	no schooling	0.60	0.58	0.54	0.58	0.58	0.55
	primary/middle schooling	0.68	0.62	0.53	0.67	0.64	0.54
	High school or more	0.73	0.70	0.63	0.69	0.65	0.61
Income	well below average	0.57	0.60	0.59	0.60	0.59	0.58
	below average	0.69	0.63	0.56	0.63	0.63	0.56
	about average	0.67	0.61	0.53	0.65	0.61	0.55
	well off	0.56	0.57	0.54	0.62	0.59	0.55
state	Punjab	0.63	0.58	0.52	0.6	0.57	0.48
	Rajasthan	0.67	0.65	0.61	0.65	0.63	0.64
	Kerala	0.80	0.71	0.58	0.76	0.67	0.57
	Karnataka	0.51	0.52	0.51	0.52	0.57	0.55
N		*407*	*410*	*405*	*423*	*420*	*420*

Means weighted by the pooled individual weights to provide survey design adjusted standard errors.
All variables are coded as binary indicators.

Table 1.7: State Life table summary statistics

	State life table estimates									
	Overall		**Punjab**		**Rajasthan**		**Kerala**		**Karnataka**	
stats	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival
mean	0.84	0.68	0.82	0.66	0.82	0.66	0.83	0.66	0.88	0.73
p50	0.91	0.78	0.89	0.74	0.92	0.81	0.88	0.73	0.92	0.80
p25	0.79	0.56	0.79	0.55	0.73	0.47	0.79	0.56	0.87	0.69
p75	0.94	0.86	0.93	0.84	0.95	0.87	0.94	0.85	0.95	0.87
N	410	405	103	103	95	95	98	98	104	104
	Subjective Survival Expectations									
stats	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival	5-year survival	10-year survival
mean	0.61	0.56	0.58	0.52	0.65	0.62	0.71	0.59	0.52	0.51
p50	0.60	0.50	0.60	0.50	0.60	0.60	0.80	0.55	0.50	0.50
p25	0.40	0.30	0.40	0.30	0.40	0.40	0.50	0.40	0.40	0.35
p75	0.80	0.80	0.80	0.70	0.90	0.80	1.00	0.80	0.60	0.60
N	410	405	103	103	95	95	98	98	104	104
p-value of t-test*	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.05	0.00	0.00

*Unpaired t-test for equality of means between state life table estimates and subjective survival expectations

Table 1.8: Basic Regressions of socio demographic characteristics and self -reported health on subjective survival expectations

	Basic Reg			Basic Reg + Self rated health		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	1-year survival	5-year survival	10-year survival	1-year survival	5-year survival	10-year survival
Male						
Female	0.020	0.011	-0.005	0.030	0.022	0.001
	(0.030)	(0.024)	(0.027)	(0.032)	(0.025)	(0.027)
45-54 years						
55-64 years	0.054*	0.024	0.011	0.070**	0.041	0.022
	(0.032)	(0.031)	(0.038)	(0.030)	(0.029)	(0.036)
65-74 years	-0.008	-0.045	-0.091*	0.037	0.001	-0.061
	(0.039)	(0.046)	(0.046)	(0.038)	(0.040)	(0.042)
Over 75 years	-0.035	-0.122**	-0.095*	0.018	-0.064	-0.057
	(0.051)	(0.043)	(0.055)	(0.045)	(0.047)	(0.061)
No Schooling						
Primary/Middle School	0.043	0.046	-0.004	0.046	0.053	0.000
	(0.041)	(0.037)	(0.041)	(0.040)	(0.036)	(0.040)
High School or more	0.086	0.103*	0.075	0.078	0.095*	0.07
	(0.053)	(0.053)	(0.055)	(0.051)	(0.050)	(0.053)
Other Caste						
Schedule Caste	-0.007	0.028	0.008	0.002	0.032	0.011
	(0.048)	(0.042)	(0.050)	(0.048)	(0.044)	(0.051)
Schedule Tribe	-0.034	-0.008	-0.031	-0.015	0.013	-0.017
	(0.070)	(0.054)	(0.055)	(0.069)	(0.056)	(0.058)
Other Backward Caste	0.017	0.008	-0.004	0.029	0.013	0.001
	(0.035)	(0.033)	(0.035)	(0.036)	(0.033)	(0.036)
Mortality Format						
Survival Format	0.028	0.057*	0.125***	0.019	0.047	0.119**
	(0.031)	(0.032)	(0.034)	(0.031)	(0.032)	(0.035)
Both parents alive						
One or both parents are dead	-0.136**	-0.102*	-0.048	-0.148**	-0.115**	-0.055
	(0.049)	(0.054)	(0.046)	(0.050)	(0.047)	(0.042)
Income-Well below Average						
Income - Below Average	0.090**	0.028	-0.033	0.102**	0.035	-0.026
	(0.044)	(0.040)	(0.045)	(0.043)	(0.040)	(0.045)
Income - About Average	0.049	-0.006	-0.073*	0.071*	0.012	-0.061
	(0.041)	(0.041)	(0.041)	(0.038)	(0.043)	(0.043)
Income - Well Off	-0.019	0.002	-0.028	-0.011	0.001	-0.027
	(0.065)	(0.057)	(0.046)	(0.066)	(0.060)	(0.048)
Punjab						
Rajasthan	0.043	0.085	0.119**	0.046	0.089	0.121**
	(0.068)	(0.060)	(0.051)	(0.068)	(0.060)	(0.051)
Kerala	0.124**	0.097**	0.068	0.191***	0.152**	0.105**
	(0.050)	(0.047)	(0.045)	(0.052)	(0.050)	(0.050)
Karnataka	-0.124***	-0.080**	-0.032	-0.119**	-0.073**	-0.029
	(0.035)	(0.037)	(0.043)	(0.036)	(0.036)	(0.043)
Self Health-Very Good						
Self Health-Good				-0.061	-0.116	-0.081
				(0.105)	(0.108)	(0.085)
Self Health-Fair				-0.149	-0.167	-0.119
				(0.107)	(0.114)	(0.096)
Self Health-Poor				-0.240*	-0.368**	-0.214*
				(0.126)	(0.129)	(0.110)
Self Health-Very poor				-0.355**	-0.309**	-0.282*
				(0.137)	(0.137)	(0.143)
N	391	391	391	391	391	391

Regressions weighted by the pooled individual weights to provide survey design adjusted standard errors. Robust standard errors clustered at state level in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1-year survival, 5-year survival and 10-year survival indicate the respondent's subjective survival expectations reported for the three time frames.

All covariates are coded as binary indicators

Table 1.9: Activity of Daily Living

ADL Dimensions	**Percentage of Respondents who report difficulty**	**N**
Difficulty with dressing	5.64	319
Difficulty walking	6.58	319
Difficulty bathing or showering	4.08	319
Difficulty eating	5.33	319
Difficulty getting in or out of bed	8.15	319
Difficulty using toilet	5.02	319

Table 1.10: Association between subjective survival expectations and objective measures of health

	1-year survival	**5-year survival**	**10-year survival**
Activity of Daily Life	0.030**	0.027*	0.008
	(0.014)	(0.016)	(0.021)
N	319	318	313
High Blood Pressure	-0.041	-0.001	-0.035
	(0.038)	(0.034)	(0.038)
N	309	307	304
Height	0.005**	0.005**	0.005**
	(0.002)	(0.002)	(0.002)
N	320	319	314

Each cell in the above table reports the results of separate estimations with all the control variables used in columns 1 to 3, Table 1.8

Regressions weighted by the pooled individual weights to provide survey design adjusted standard errors. Robust standard errors clustered at state level in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$

ADL is the first component of a principle component analysis with mean of 0 and standard deviation of 1 using the components shown in Table 1.9

High blood pressure is a binary indicator with 1 indicating the indicende of high blood pressure

Height is measured in centimeters

Table 1.11: Association of subjective survival probablities with haemoglobin concentrations

	Female			**Male**		
	1-year survival	**5-year survival**	**10-year survival**	**1-year survival**	**5-year survival**	**10-year survival**
Low Haemoglobin	0.058	0.051	0.072	-0.138**	-0.167**	-0.024
	(0.036)	(0.030)	(0.050)	(0.062)	(0.068)	(0.074)
N	152	148	145	141	144	142

Each cell in the above table reports the results of separate estimations with all the control variables used in columns 1 to 3, Table 1.8 Low haemoglobin is a binary indicator with 1 indicating haemoglobin levels below 12.0g/dl for women and below 13.0g/dl for men. Regressions weighted by the pooled individual weights to provide survey design adjusted standard errors. Robust standard errors clustered at state level in parenthesis $***p < 0.01$,

Table 1.12: Association of subjective survival probablities with savings and outstanding bank loans

	Basic Reg				Basic Reg	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Loans	**Loans**	**Loans**	**Savings**	**Savings**	**Savings**
1-year survival	0.116**			-0.39		
	(0.052)			(0.356)		
5-year survival		-0.045			0.322	
		(0.054)			(0.361)	
10-year survival			-0.061			0.142
			(0.038)			(0.317)
Male						
Female	-0.021	-0.02	-0.021	0.339	0.354*	0.351
	(0.040)	(0.039)	(0.039)	(0.206)	(0.208)	(0.218)
45-54 years						
55-64 years	-0.05	-0.044	-0.045	0.607**	0.591**	0.582**
	(0.038)	(0.038)	(0.038)	(0.280)	(0.286)	(0.285)
65-74 years	0.015	0.009	0.006	0.487*	0.584**	0.552**
	(0.047)	(0.047)	(0.047)	(0.256)	(0.262)	(0.256)
Over 75 years	-0.046	-0.058	-0.058	0.289	0.376	0.336
	(0.058)	(0.057)	(0.058)	(0.567)	(0.542)	(0.546)
No Schooling						
Primary/Middle School	0.071**	0.079**	0.077**	0.241	0.224	0.236
	(0.035)	(0.037)	(0.037)	(0.265)	(0.263)	(0.271)
High School or more	0.031	0.046	0.045	0.757*	0.722*	0.746*
	(0.058)	(0.061)	(0.060)	(0.393)	(0.373)	(0.390)
Other Caste						
Schedule Caste	0.078	0.077	0.076	-1.051**	-0.990*	-0.981*
	(0.048)	(0.047)	(0.047)	(0.500)	(0.545)	(0.541)
Schedule Tribe	-0.005	-0.011	-0.012	-0.051	-0.222	-0.135
	(0.048)	(0.048)	(0.048)	(1.400)	(1.351)	(1.336)
Other Backward Caste	0.033	0.034	0.033	-0.175	-0.18	-0.161
	(0.065)	(0.066)	(0.066)	(0.403)	(0.397)	(0.402)
Income-Well below Average						
Income - Below Average	0.081*	0.094**	0.091**	-1.093**	-1.174**	-1.142**
	(0.042)	(0.043)	(0.043)	(0.388)	(0.407)	(0.413)
Income - About Average	-0.012	-0.005	-0.01	-0.541*	-0.559*	-0.545*
	(0.043)	(0.041)	(0.042)	(0.297)	(0.289)	(0.283)
Income - Well Off	-0.028	-0.029	-0.031	-0.688	-0.713	-0.706
	(0.047)	(0.046)	(0.046)	(0.460)	(0.474)	(0.467)
Punjab						
Rajasthan	0.009	0.018	0.021	-0.383	-0.365	-0.391
	(0.039)	(0.040)	(0.039)	(0.334)	(0.341)	(0.340)
Kerala	0.342***	0.360***	0.360***	-0.071	-0.149	-0.135
	(0.059)	(0.058)	(0.059)	(0.496)	(0.473)	(0.477)
Karnataka	0.042	0.025	0.026	-1.454***	-1.350***	-1.392***
	(0.046)	(0.046)	(0.045)	(0.355)	(0.368)	(0.377)
N	391	391	391	129	129	129

Regressions weighted by the pooled individual weights to provide survey design adjusted standard errors. Robust standard errors clustered at state level in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1-year survival, 5-year survival and 10-year survival indicate the respondent's subjective survival expectations reported for the three time frames.

Loans are a binary indicator with 1 denoting the existence of an outstanding loan. Savings are coded in Indian Rupees with a mean of INR 35,521 and standard deviation of INR 66,923

All other covariates are coded as binary indicators

Chapter 2

The impact of a negative health information shock on survival expectations in England

2.1 Introduction

Individuals make numerous inter-temporal decisions based on their expected probability of survival, such as decisions about investment, remittances, bequest, saving, and healthcare. There is a growing body of literature that measures individual survival probabilities and estimates their effects on economic actions.

While previous research suggests that subjective survival expectations are able to measure varied private information and are predictive of actual survival (Bloom et al., 2006; Delavande and Rohwedder, 2011; Elder, 2007; Hurd and McGarry, 2002; Perozek, 2008), we do not yet fully comprehend the formation of these survival expectations. We have little evidence to establish whether individuals update their expectations with new information in meaningful ways. The mechanisms through which individuals update their survival expectations are important to understanding individual actions. I use a survey induced variation in the form of a negative health information shock and test its impact on survival expectations. Consistent with rational behaviour, I find that individuals who received a negative health information shock report lower survival probabilities as compared to individuals who did not receive the negative health information shock.

I also extend the literature by systematically analysing the relationship between having been previously diagnosed by a doctor with a chronic or cardiovascular condition and subsequent survival expectations in England. My results suggest that individuals with pre-existing health conditions on average report significantly lower survival probabilities than individuals who do not. These findings suggest that individuals are able to use private information to meaningfully update their survival probabilities and thereby highlight the predictive power of survival expectations.

The rest of the paper is organised as follows. Section 2.2 and Section 2.3 review previous research that is relevant to this work. Section 2.4 describes the English Longitudinal Study of Ageing (ELSA) dataset and the various components used in this study. Section 2.5 outlines the empirical strategy. Section 2.6 presents the main results and Section 2.7 concludes.

2.2 Do Survival Expectations Matter?

Several large-scale longitudinal surveys now elicit survival probabilities from respondents including the Health and Retirement Study (HRS) in the United States, the English Longitudinal Survey of Ageing (ELSA) in the England and the Survey of Health, Ageing and Retirement (SHARE) in Europe to name a few. Longitudinal household surveys in Asian countries like the Longitudinal Ageing study in India (LASI), Korean Longitudinal Study of Ageing (KLoSA) and the Japanese Study on Ageing and Retirement (JSTAR) also elicit detailed survival probabilities.

There is a growing body of literature investigating the best methods to elicit survival probabilities in different country contexts. For example, Manski (2004) and Hurd (2009) review the literature on the elicitation of probabilistic expectations in developed countries and Delavande (2014) and Delavande et al. (2011) conclude that it is feasible to elicit subjective probabilities from individuals with varied socio-economic backgrounds in diverse developing country contexts as well.

Survival expectations in the ELSA dataset have been previously studied in the context of association with actual survival (Adams et al., 2014). The authors find that anticipated survival was positively associated with actual survival and that this relationship was strongest in younger individuals and in those with higher household incomes. ELSA data has also been used as a part of a larger cross-country study comparing differential survival probabilities between Europe and the United States (Delavande and Rohwedder, 2011). They find that wealthier people have higher survival expectations as is the case with actual survival. Survival expectations in ELSA have been used to test the timing of retirement (O'Donnell et al., 2008). The authors find that men who are extremely pessimistic about their survival chances are the least likely to retire but the propensity to retire falls as survival expectations improve. While the impact of survival expectations from the ELSA dataset on health, retirement and actual survival are therefore clearly of importance, there has only been limited investigation (Section 2.3 provides an overview of this literature) into the formation of survival expectations. It is the objective of this paper to close this gap.

2.3 How are Survival Expectations formed?

Notwithstanding the necessity of an examination of elicitation techniques of survival probabilities and their applicability to individual economic action, there exists relatively little scholarly work on the actual formation of survival probabilities in the UK. The little empirical evidence available regarding the formation of survival probabilities mostly stems from studies using the Health and Retirement Study in the United States.

For instance, Smith et al. (2001) evaluate the effect of an exogenous health shock to survival expectations in the Health & Retirement Survey and find that smokers update their survival expectations differently than non-smokers. They find that smokers only react to smoking related heath shocks while non-smokers react to all health shocks. Hurd and McGarry (2002), using the Health and Retirement Study, find that survival expectations decline with the death of a parent and that survival probabilities predict actual mortality. They also find that respondents update survival expectations when they have a new cancer diagnosis. Hurd et al. (2001) find that individuals modify their survival probabilities with the onset of adverse health conditions such as high blood pressure, cancer and diabetes. Liu et al. (2007) use a two-period panel data from the national Health Insurance Program in Taiwan (NHI) and find that the number of abnormal test items in a physical examination and the number of recommendations received from the doctor reduce survival probabilities. The applicability of this evidence to England and English data has not yet been tested.

2.4 Data Description

2.4.1 English Longitudinal Study of Ageing (ELSA)

I use data from the English Longitudinal Study of Ageing (ELSA)[1], which is a representative sample of the English population aged 50 and over. It is a comprehensive dataset, which in addition to socio-demographic information also collects extensive subjective and objective health measures. The study began in 2002 and the sample is resurveyed every 2 years. Survival expectations to age 80 was asked of ELSA respondents for the first time in wave 3. For this analysis I use biomarker data from wave 2 which was administered in 2004 and sociodemographic and

[1]Details about the dataset can be found in Marmot et al. (2010)

survival expectations data from wave 3 which was administered in 2006. I also use data from wave 0 to wave 3 to construct indicators of pre-diagnosed health conditions.

2.4.2 Analytical Sample

Data from the nurse visits of ELSA (detailed in Section 2.4.5) define our analytical sample as they are an integral part of this study. Only Cohort 1 Core Members, hereafter referred to as C1CM, who had an in-person interview (i.e. not by proxy) were eligible for a nurse visit. Table 2.8.1 presents the selection criteria for the wave 2 nurse data. 7666 individuals had a productive nurse visits out of whom 6228 individuals had a full or partially viable blood sample. The survival expectations questions were only asked of individuals less than 70 years of age and so we retain these. 4092 out of these respondents are less than 70 years of age.

I restrict the analysis to data from waves 2 and 3 for two reasons. Firstly, ELSA has a high rate of sample attrition. ELSAs attrition rate by wave 3 is nearly 23% as compared to the rate of 6% in the Health and Retirement Study in the US. In addition there has been a declining response rate of core members to participate in the nurse visit. Secondly disease thresholds are being periodically re-evaluated (Smith, 2007). This means that the rates of disease diagnosis may vary between the biennial waves and the estimates on morbidity and disease incidence would be imprecise.

Table 2.8.1 also shows our selection criteria for wave 3 data. The total number of respondents interviewed at wave 3 is 7535 out of whom 4368 individuals are less than 70 years of age. 3280 out of these have matched blood data from wave 2 and 3158 individuals have full information on the socio-demographic and health covariates used in our estimations.

Given that respondents may have died or otherwise attrited between wave 2 and wave 3, I test the balance on observed sociodemographic characteristics between the analytical sample from wave 3 and the sample with blood information from wave 2 and present the results in Table 2.8.2. Column 1 presents the means on covariates for the wave 2 sample and column 2 presents the means on covariates for the analytical sample. The number of individuals that have died or otherwise attrited is 934 which translates to an attrition rate of 23%.[2] We find

[2]The number of individuals who have died or otherwise attrited for all cohort1 1 core members from wave 2 (N=8780) to wave 3 (N=7535) is 1245 respondents which translates to 14%

a statistically significant difference in means between the samples on four of the covariates a shown by the p-value of an unpaired t-test reported in column 3. The analytical sample is on average of a higher income and of lower age groups than the wave 2 sample. Given that descriptive statistics suggest that the samples are of significantly different age and income tertiles, I use these covariates to estimate tightened Lee bounds (Lee, 2009) of the main results in Section 2.6.1.

Net income in ELSA are comprehensive weekly values and are a sum of employment income, state pension income, state benefit income, private pension income and asset income. For ease of interpretation these have been recoded to income tertiles of low, middle and high income. As shown in Column 2 of Table 2.8.2, 32% of the sample is from a low income household while 31% and 37% are from a middle income and high income household respectively. 41% of our sample are less than 59 years of age and 32% are between ages 60 and 64 years. At 74% a majority of the sample are either married, cohabiting or have a civil partner while 8% of the sample is widowed. 54% of the sample are women. Since parental death has been previously shown to be predictive of both survival expectations (Hurd and McGarry, 2002), we condition on this in our estimations. The proportion of individuals who have experienced at least one parental death is 9%. 62% of the sample were previous smokers or are current smokers and 66% of the sample report having consumed alcohol in the 7 days preceding the survey. At wave 3 individuals were asked to report their highest educational qualification obtained which is classified into National Vocational Qualification (NVQ).[3] For the purpose of this paper, this is classified into 3 categories; No Qualification, NVQ 1 - NVQ 2 which is equivalent to GCSE grades A*- C and NVQ 3 and above which is equivalent to A level and above.[4] This information was not asked of respondents in wave 2 and so I am unable to check for balance on these variables but nonetheless include them as covariates in the estimations. 37% of the sample has an NVQ 3 qualification or above which translates into A-levels or higher education. 42% have an NVQ Level 1 or NVQ Level 2 qualification which are equivalent to GCSE qualifications. 21% of the sample report having no educational qualifications at all.

[3]NVQs are based on statements of performance that describe what competent people in a particular occupation are expected to be able to do.

[4]Full NVQ classifications and its comparison to different qualifications is available at https://www.gov.uk/what-different-qualification-levels-mean/compare-different-qualification-levels

2.4.3 Survival expectations of older respondents

ELSA respondents below 70 years of age were asked about their survival expectations in wave 3. All respondents were privately asked about the probability that they would survive up to a target age of 85 or more. Respondents were first given the following introductory text

"Now I have some questions about how likely you think various events might be. When I ask a question I'd like you to give me a number from 0 to 100, where 0 means that you think there is absolutely no chance an event will happen, and 100 means that you think the event is absolutely certain to happen."

Subsequently, respondents were asked about the probability that they would survival up to a target age of 85 years or more. The specific wording of the question was *"What are the chances you will live to be 85 or more?"* The responses were recorded on a scale of 0 to 100. Figure 2.8.1 is a histogram of the subjective survival expectations to target age 85 years among our sample. The mean subjective expectation of survival to age 85 years is 49.39. Previous studies have shown that expectations of 0.5 may be indicative of epistemic uncertainty (e.g., de Bruin et al., 2000). Similarly, survival expectations in ELSA also tend to cluster at responses of 0.5.

Since actual mortality is known to vary over socio-economic characteristics, I investigate whether the subjective probability of survival also varies over these characteristics in meaningful ways in Table 2.8.3. Means are weighted by the wave 3 cross-sectional weights provided in the ELSA core dataset to provide survey design adjusted standard errors. There is a clear statistically significant income and educational gradient in survival expectations. For instance, respondents from the high income tertile report a 51.7 percent probability of survival to age 85 while respondents form the low income tertile report a 44.8 percent probability of survival. Similarly, respondents with an NVQ level 3 or above qualification report a 51 percent probability of survival while individuals with no education report 43.67 percent probability of survival. Women report a higher survival probability as compared to men which is consistent with women having a higher life expectancy at birth than men in England. Life expectancy at birth in Britain is 82.9 years for women and 79.2 years for men (Organisation for Economic Co-operation and Development, 2015b). Life expectancy at the age of 65 is also higher for women at 20.9 years while for men it is 18.6 years (Organisation for Economic Co-operation and Development, 2015a). The average survival expectations reported by current or past smokers is approximately 7 percent less than the probability reported by individuals who never smoked and this difference is significant suggesting that smokers are aware of the negative health outcomes of smoking. Interestingly, individuals who consumed alcohol in the 7 days preceding the survey report a 5% significantly higher probability of survival than individuals who did not.

2.4.4 Doctor Diagnosed Health Conditions

ELSA implements a Health module as a part of its main questionnaire in all waves. This includes information about long-standing illness and specific diagnosis and symptoms. The first time a respondent takes part in a survey, starting from the Health Survey of England (HSE), they are asked if they have ever been diagnosed by a doctor with a particular condition, separately for each diagnosis. The question is framed as *"Did a doctor ever tell you that you had..."* In each subsequent wave they are asked if this condition continues and if any new conditions have been diagnosed. I created binary indicators for whether a condition has ever been diagnosed in previous waves using all previous waves for wave 2 and wave 3 and for and each condition.

Table 2.8.4 reports the summary statistics of these variables indicating pre-existing health conditions as diagnosed by a doctor classified into cardiovascular conditions in Panel A and chronic conditions in Panel B. Cardiovascular conditions include the following pre-diagnosed conditions; high blood pressure, angina, heart attack, chronic heart failure, heart murmer, abnormal heart rhythm, diabetes, stroke and cholesterol while chronic conditions includethe following pre-diagnosed conditions; lung disease, asthma, arthiritis, osteoporosis, cancer, parkinsons and psychiatric disorder. I report the conditions reported at wave 2 (Column 1) and at wave 3 (Column 2) and the percentage change between both waves (Column 3). Pre-existing conditions which are reported at wave 2 are disease prevalence in the analytical sample while the change is indicative of new disease incidence. The percentage change between waves is reflective of new disease incidence in the sample. The most prevalent cardiovascular disease at wave 2 is high blood pressure with 35% of individuals having been diagnosed by wave 2. High cholesterol is the highest newly reported disease with a 16% increase in new diagnoses at wave 3. Similarly the most prevalent diagnosed chronic condition is arthritis with 32% of the sample having been diagnosed with it by wave 2. Arthritis is also the highest newly reported disease with a 4.62% increase in new diagnoses at wave 3. I collapse these into two binary indicators; one for the presence of any cardiovascular conditions and one for the presence of any chronic conditions and use these indicators measured in wave 2 as a measure of morbidity in the regressions. 50% reported having been diagnosed with a cardiovascular condition and 51% report having been diagnosed with a chronic condition by wave 2.

In Table 2.8.5, I examine whether subjective probability of survival varies over the cardiovascular and chronic conditions described earlier. We would expect that on average individuals who have been diagnosed of a condition by wave 3 would report lower survival probabilities as compared to individuals who have not been diagnosed with a condition. I find this to be true for all of the cardiovascular and chronic conditions reported. Column 1 presents the mean of survival probabilities weighted by the longitudinal weights provided in the ELSA Wave 3 dataset for individuals who have not been diagnosed with the corresponding health condition while column 2 presents the weighted mean of survival probabilities for individuals who have been diagnosed with the corresponding health condition. As Column 3 reports the percentage difference in the mean between the two groups, for all conditions, the difference reported is a negative. Column 4 reports the p-values of an unpaired t-test of equality of means between the two groups. The differences in survival probabilities are statistically significant for almost all conditions. These descriptive statistics seem to suggest that individuals who have been diagnosed with a pre-existing condition do report a lower survival expectations than individuals who do not.

2.4.5 Blood Analytes

ELSA also included a nurse visit to conduct biomedical tests once every two waves starting at wave 2. (See Figure 2.8.6 for timeline). All sample members who gave consent for the nurse data were also eligible for a blood sample to be taken. Sample members who gave consent also had their blood sample test results sent to their General Practitioner (GP) and or themselves. The proportion of respondents who consented to having their blood sample results sent to their GP was 97.7%. The letter sent to the GP included the exact results of the blood analysis and the GP's were also informed of the normal range used for the analysis. In addition to sending a letter to the GP, respondents who gave consent were also sent a letter informing them of the results. 96.6% of the sample consented to having their blood results sent to themselves. This letter showed whether the result of each of the analyses conducted on the blood sample was within or outside the normal range. If the results were out of range, the respondents were advised to contact their GP.

I use four blood analytes in this study:

1. Blood glycated haemoglobin:
   Blood glycated haemoglobin measures the amount of sugar build up in the blood in the 8 to 12 weeks before the measurement is taken. It is used to highlight individuals who are not managing their diabetes well or flag individuals who might be suffering from diabetes which is associated with an increased risk of heart disease. Blood glycated haemoglobin can be expressed as a value in millimoles per mole (mmol/mol) or as a percentage. The ELSA wave 2 dataset records it in percentage form. Since mmol/mol has been the default unit to use in the UK since 2009 we report both statistics. Figure 2.8.2 graphs the kernel densities of blood glycated haemoglobin level of our sample for men and women. The mean blood glycated haemoglobin level of the analytical sample is 37.6 mmol/mol (5.59%) for men and 36.8 mmol/mol (5.52%) for women. These values are slightly lower in comparison to the levels of blood glycated haemoglobin in the Health Survey of England (2011), where the level was 38.7 mmol/mol (5.7%) for men and 38.3 mmol/mol (5.7%) for women. I use a cut-off point of >48mmol/mol (>=6.5%) which is the established cut-off for a diagnosis of Type 2 diabetes (Association et al., 2010; Organization et al., 2011).

2. C - Reactive Protein (CRP):
   CRP is a marker of inflammatory load and higher values are associated with adverse cardiovascular disease and mortality (Libby et al., 2011). It rises in response to inflammation and is a defence against harmful stimulus. Figure 2.8.3 graphs the kernel densities of c - reactive protein level of our sample for men and women in milligrams per litre (mg/l). The mean CRP level in the analytical sample is 2.17 mg/l for men and 2.49mg/l for women and the comparative statistics for the Health Survey of England (2009) are 2.5 mg/l and 3.1 mg/l. The clinical cut-off point is 3 mg/l and values over this cut-off are considered a risk factor for CVD (Pearson et al., 2003). Values over 10 mg/L have been removed from the analysis as they are an indicator of recent infection.

3. Ferritin:
   Ferritin levels measured in microgram per litre (ug/l) are the size of iron stores in the body. Higher levels of ferritin suggest excess body iron. Figure 2.8.4 graphs the kernel densities of blood ferritin levels for men and women. The mean blood ferritin level for the analytical sample is 153.2 ug/l for men and 89.5 ug/l for women. These values are slightly lower to the mean values obtained from the Health Survey of England (2009) with 158.9

ug/l for men in the same age group as our sample and 102 ug/l for women. Ferritin levels greater than 300 ug/l may indicate iron overload in men and postmenopausal women. We use this cut-off to create a binary indicator of high ferritin levels.

4. Haemoglobin:
   Low levels of haemoglobin are usually indicative of an iron deficiency and anaemia. This is prevalent in women and increases with advancing age and increases the risk of cardiovascular disease and mortality (Culleton et al., 2006; Patel et al., 2009). Figure 2.8.5 graphs the kernel densities of blood haemoglobin level measured in grams (g) per decilitre (dL) of our sample for men and women. The mean blood haemoglobin level is 15.28 g/dl for men and 13.98 g/dL for women. These are close to the mean values by gender obtained from the Health Survey of England (2009) where the mean blood haemoglobin level is 14.9g/dL and 13.3g/dL for men and women respectively. I use clinical cut-off points of <12 g/dL for women and <13 g/dL for men to define anaemia based on the World Health Organisation criteria (UNICEF et al., 2001).

I create binary indicators for an out-of-range result using the clinical cut-offs described for each of the 4 biomarkers with 1 indicating an out-of-range result. Table 2.8.6 presents the proportion of the sample with out of range results for the full sample and by gender. At 24%, c-reactive protein has the highest out-of-range result and haemoglobin has the lowest at 3%. A significantly higher proportion of men have out-of-range blood glycated haemoglobin and ferritin result while a significantly higher proportion of women have an out-of-range c-reactive protein result.

2.5 Empirical Specification

This paper examines the effect of a survey induced negative health information shock on subjective survival expectation. I hypothesise that individuals who have received a negative health information in the form of out-of-range biomarker results post wave 2 will on average report a lower probability of survival to age 85 years as compared to individuals who have not received a negative health information.

Figure 2.8.6 outlines the timeline of the main surveys and the health interventions. The dependent variable of survival expectation to age 85 or more is measured at wave 3. The independent variable of interest, a negative health information intervention in the form GP's and/or respondents themselves being notified of blood assay results occurs at post wave 2. Nurse visits took place in the weeks following the main survey and the results were sent to the GP and/or to the respondent within approximately 3 months following the nurse visit. Cardiovascular and chronic conditions are recorded at both wave 2 and wave 3. I use the cardiovascular and chronic conditions recorded at wave 2 as an indicator for morbidity or disease prevalence in the analytical sample and condition on this in the estimations.

I estimate a least square regression of the following functional form with robust standard errors clustered at the primary sampling unit level:

$$Y_{it} = \beta_1 X_{i,(t-1)} + \beta_2 Z_{i,(t-1)} + \beta_3 C_i + \beta_4 A_i t + \varepsilon_{it}$$

where,

Y_{it} is the expectation of surviving over 85 years of age for individual i at time t ;
$X_{i,(t-1)}$ is an indicator for an out-of-range analyte result at time t-1
$Z_{i,(t-1)}$ is pre diagnosed health conditions for individual i at time t-1;
C_i are socio demographic controls;
$A_i t$ are age fixed effects at time t;
ε_{it} is the error term.

All the results presented in this paper are weighted by wave 3 cross-sectional weights provided in the ELSA dataset. Re-estimating the results using the weight for blood sample analysis provided in the wave 2 nurse visit documentation does not alter the results significantly.

Similar to the approach used in this paper, Rodriguez-Lesmes (2014) use ELSA data to analyse the impact of a health check that provides advice to individuals with mildly raised blood pressure using a regression discontinuity design and find that there is an increase in the average use of medication for thoese individuals who were advised to visit their doctor. In contrast, this paper estimates the impact of several negative health information shocks on survival expectations.

Data for up to wave 6 of ELSA is now publicly available. ELSA also conducted a further 2 nurse visits in wave 4 and wave 6. I do not use the out-of-range biomarker data from wave 6 as we do not have survival expectations measured post wave 6. I also restrict the analysis to the first nurse visit conducted in wave 2 for several reasons. Firstly ELSA has a high rate of sample attrition. ELSAs attrition rate by wave 3 is nearly 23% as compared to the rate of 6% in the Health and Retirement Study in the US. Previous analysis on attrition in ELSA has shown that least educated individuals and individuals with lower numerical abilities are more likely to drop out of the survey. In addition there has been a declining response rate of core members to participate in the nurse visit. Secondly this analysis suffers from mortality selection. This sample selection problem is greater at later waves. A discussion and life table comparison is given in Banks et al. (2010). The authors show that after age 65, the mortality in ELSA is lower than the mortality in the English national death index (GAD). The authors attribute this to ELSA not having reached a population representative steady state. Thus by restricting the analysis to the earlier waves of ELSA we are reducing this selection bias. Thirdly disease thresholds are being periodically re-evaluated (Smith, 2007). This means that the rates of disease diagnosis may vary between the biennial waves and the estimates on morbidity and disease incidence would be imprecise. Fourthly ELSA only started recording detailed prescribed medication use of respondents from wave 6 which means that had I used wave 4 in this analysis I would have been unable to separate the normal biomarker results in wave 4 from natural causes or due to drug dependency and good disease management. This would induce measurement error in longitudinal estimates. Once data for Wave 7 is available, there is scope to test the results in this paper with the results of the nurse visit in wave 6 conditional on medication use.

2.6 Out-of-range biomarker results and subjective survival expectations

Table 2.8.7 presents the association between an out-of-range blood glycated haemoglobin result in wave 2 and survival expectations in wave 3. Column 1 presents the unconditional estimates. Column 2 conditions on demographic characteristics described in Section 2.4.2. Column 3 additionally conditions on any pre-existing cardiovascular conditions reported in wave 2 and column 4 additionally conditions on any pre-existing chronic conditions reported in wave 2. A high blood glycated haemoglobin result in wave 2 is associated with a 13.7% decrease in survival expectations in wave 3 as shown in Column 1. The magnitude of this estimate reduces in each subsequent specification to 8.3% and continues to be significant.

Columns 2, 3 and 4 in Table 2.8.7 also presents the associations between known sociodemographic predictors and subjective survival expectations. As seen in the descriptive statistics (Section 2.4.3), low income is negatively associated with survival expectations with a magnitude of 3.9%. Divorced/Separated respondents and widowed respondents report a lower survival expectation as compared to never married respondents. Women report higher survival expectations as compared to men with a magnitude of 8.4%. Respondents without any formal educational qualifications report lower survival probabilities. Respondents who were previous smokers or who are current smokers at the time of the wave 3 data collection report a 4% lower probability of survival while respondents who consumed alcohol in the 7 days preceding the survey report higher survival expectations with a magnitude of 3.3%. The magnitudes of these estimates and their significance remain stable in the remaining specifications irrespective of the blood analyte used. These estimates are in line with existing literature on the associations between sociodemographics and survival expectations in other countries. For instance, Hurd and McGarry (1995) find that a higher income, wealth and education are positively related to survival expectations and that smokers report lower survival probabilities in the US. As expected both doctor diagnosed chronic and cardiovascular diseases have a negative relationship with survival expectations with magnitudes of 5.6% and 3.1% respectively and these are significantly different from zero.

Similarly Table 2.8.8 presents the association between an out-of-range c-reactive protein result and survival expectations. An out-of-range c-reactive protein result is negatively associated with survival expectations with a magnitude of 1.9%. This relationship is statistically significant in the first three specifications (Columns 1, 2 & 3) but is less precisely estimated when conditional on doctor diagnosed chronic conditions in Column 4. Table 2.8.9 presents the estimates of the association between an out-of-range ferritin result and survival expectations. I find a negative relationship between them with a magnitude of 2.28% but this is only significantly difference from zero in the first bivariate specification. I find a negative relationship between an out-of-range haemoglobin result and survival expectations and present these results in Table 2.8.10. Individuals who received an out-of-range blood haemoglobin results on average report 8.9% lower survival probabilities than individuals who do not and the magnitude is statistically significant in all specifications.

Since the negative health information shocks may have a different impact on different respondents, there is no straightforward way to combine them together. For instance, a high blood glycated haemoglobin would have a different impact on a respondent who has previously been diagnosed with diabetes as compared to a respondent who has not previously been diagnosed with diabetes. Nonetheless, as an additional robustness check, Table 2.8.11 presents the results when including all binary negative health information shocks on subjective survival expectations. The analytical sample in this table is restricted to those respondents who had a viable blood sample to test for low levels of haemoglobin. As shown in Column (4) all but Ferritin continue to have a negative association with survival expectations.

2.6.1 Sample Selection

The results presented in the previous section rests on the assumption that treatment in the form of a negative health information shock is random. The analytical sample used in the main results suffers from sample selection in that only individuals who are observed in wave 3 are included in the estimations. Individuals who have died between wave 2 and wave 3 or who have otherwise attrited are not observed. It is plausible that individuals who received an out-of-range biomarker result post wave 2 are more likely to not be observed in wave 3 due to mortality or due to being institutionalized leading to a non-random selection. So for any given individual, the assignment to treatment makes selection less likely.

To account for potential sample selection bias in the previous estimates, I determine an interval for effect size which correspond to extreme assumptions about the impact of selection using Lee (2009) bounds. The sample is trimmed from above and below such that the share of observed individuals is equal for both groups. The results presented in this section also use age covariates which have explanatory power for mortality related attrition to yield tighter bounds. The two samples compared are individuals who have a biomarker result for the relevant biomarker in wave 2 to the subsample of these individuals who also have a non-missing value of survival expectations in wave 3. The rate of attrition between wave 2 and wave 3 is 23%.

Table 2.8.12 presents the tightened bounds for the estimates of an out-of-range blood glycated haemoglobin information shock on survival expectations. Both the lower (β = -15.63) and the upper bound estimates (β = -11.54) are negative and statistically significant and the effect confidence interval does not cover zero. Similarly, Table 2.8.13 presents the tightened bounds for the estimates of an out-of-range c-reactive protein result and survival expectations. Again both, lower (β = -3.7) and upper (β = -3) bounds are negative and statistically significant, the effect confidence interval does not include a zero and the bounds cover the OLS point estimate from the main results. Table 2.8.14 presents the tightened bounds for an out-of-range ferritin result and survival expectations. We find a negative and statistically significant lower bound (β = -6.13) but find that the upper bound (β = -3.69) is not significantly different from zero. Also the effect confidence interval covers a zero. Table 2.8.15 presents the tightened bounds for an out-of-range haemoglobin result and survival expectations. I find that both the lower (β = -8.78) and upper (β = -9.66) bounds estimate are negative and statistically different from zero and the bounds cover our OLS point estimate. The effect confidence interval does not cover a zero.

Overall, accounting for sample selection seems to validate our main results of a negative and significant relationship between an out-of-range biomarker result for blood glycated haemoglobin, c-reactive protein and haemoglobin and a negative but insignificant result for ferritin.

2.6.2 Heterogeneous Effects

In this section we further investigate our previous estimates by two heterogeneous sub groups; educational qualification and smoking status. Previous research has established an association between education and smoking and actual mortality. I examine if these relationships persist when examining survival expectations instead of actual mortality. Given that previous research into the association between survival expectations and smoking has established that smokers update their expectations differently than non-smokers (Smith et al., 2001), this is of particular interest.

Firstly I examine if the direction and magnitude of the main estimates are different for individuals who are current/previous smokers versus non-smokers and present these estimates for the specification conditional on all pre-existing conditions in Table 2.8.16. The negative relationship between an out-of-range biomarker result and survival probabilities persists in all cases, but remains statistically different from zero only in the blood glycated haemoglobin and haemoglobin estimations. The magnitudes of these effects are interesting. For the blood glycated haemoglobin result, the marginal reduction in survival probabilities is higher amongst non-smokers as compared to smokers. In contrast, the marginal reduction in survival probabilities for the haemoglobin specification is higher amongst smokers as compared to non-smokers. An out-of-range indicator of blood glycated haemoglobin is defined as 1 for values $>$ 48mmol/mol while for haemoglobin this is defined as 1 for values $<$ 12g/dL. Biomedical literature has previously established that smokers on average have a higher distribution of both blood glycated haemoglobin (Sargeant et al., 2001; Urberg et al., 1989) and haemoglobin (Nordenberg et al., 1990). There is also an ongoing discussion to possibly define different cut-off points for smokers versus non-smokers (Milman and Pedersen, 2009). These results seem to indicate that smokers are possibly either aware of this information regarding differential results for smokers or are perhaps made aware of this by their GP's. In any case further research is needed to understand the previously established differences in updating behaviour between these groups.

Secondly I investigate if the magnitude and direction of our estimates are different for individuals that do not have any educational qualifications as compared to individuals who have some educational qualifications. I find larger effects for individuals without any educational qualifications as shown in Table 2.8.17, suggesting that survival expectations for this group are more sensitive to negative health information shocks as compared to individuals with at least an NVQ 1 qualification.

2.7 Conclusion

I examine whether survival expectations respond to a negative health information shock in the older English population. I find that on average individuals who have received a negative health information shock report lower survival probabilities than individuals who do not. I also find that individuals who have been diagnosed with any chronic or cardiovascular condition report significantly lower survival expectations. This suggests individuals are able to rationally update their expectations based on information received, lending evidence to the predictive power of subjective survival expectations. Given that respondents to surveys in general, and health surveys in particular, are usually a healthier sample of the general population (Cohen and Duffy, 2002; Kypri et al., 2011), it is likely that the magnitude of these results are an underestimation of the actual effect in the general population. These findings support previous research using the HRS in the US and suggest its applicability to the UK.

I also find that the magnitude of this estimate varies by smoking status and the specific blood analyte indicator used in the estimation. This confirms previous research that smokers update their survival probabilities differently as compared to non-smokers and that there may be a need to define different clinical cut-off points for smokers.

These results therefore contribute to the validity of subjective survival expectations and support the idea that they are able to elicit private information that are crucial to understanding and anticipating individual actions. Survival expectations can thus provide information over and above mortality predictions, which are merely based on extrapolations of historical data. The increasing prominence of survival expectations within longitudinal surveys and the use of survival probabilities in empirical work thus seems justified.

2.8 Figures and Tables

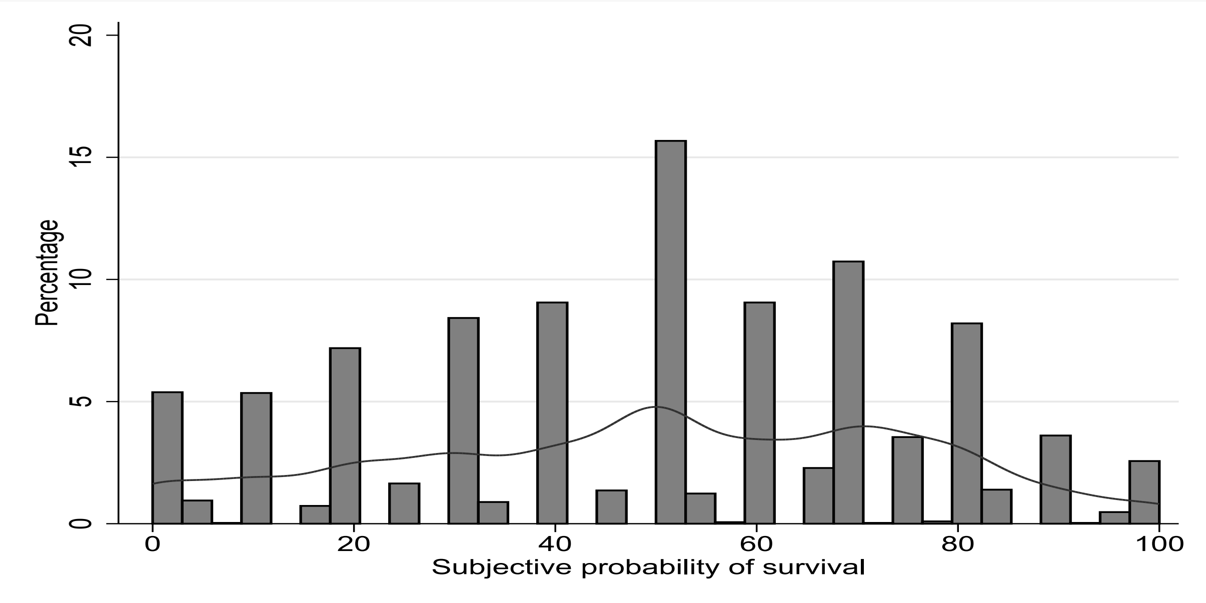

Figure 2.8.1: Distribution of subjective expectations of survival

Notes: This figure plots distribution of the subjective expectations of survival to target age 85 years in the analytical sample. The specific wording of the question was "What are the chances you will live to be 85 or more?" The response was recorded on a scale of 0 to 100

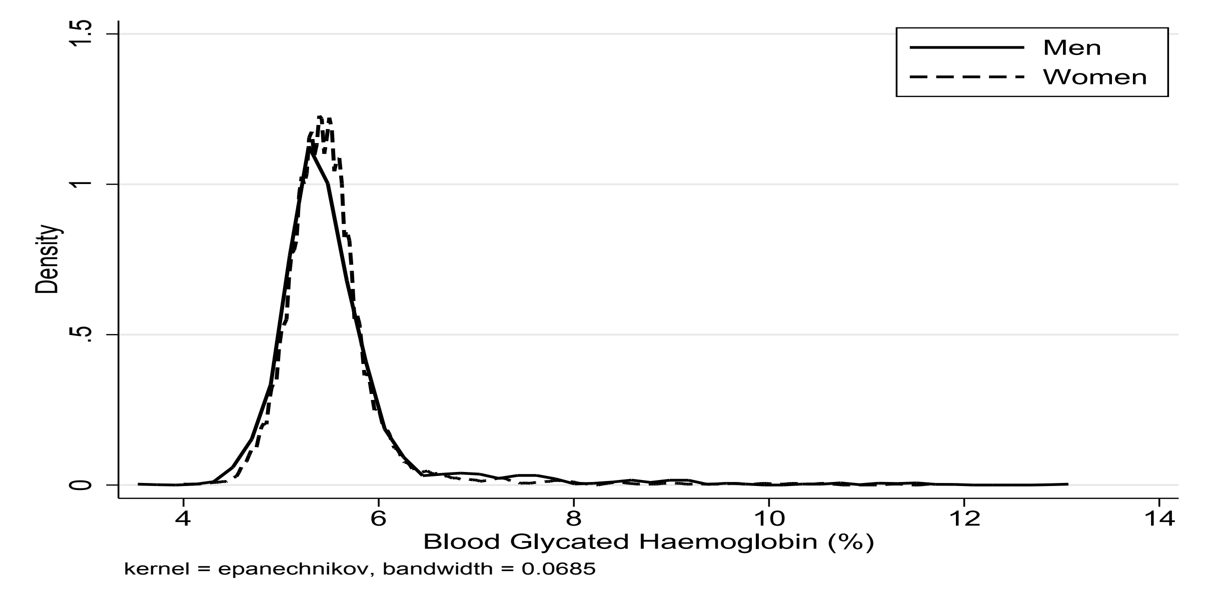

Figure 2.8.2: Kernel Density Estimate of Blood Glycated Haemoglobin

Notes: This figure graphs the kernal densities of blood glycated haemoglobin level of the analytical sample for men and women measured in millimoles per mole (mmol/mol). Mean blood glycated haemoglobin is 37.6mmol/mol for men and 36.8 mmol/mol for women

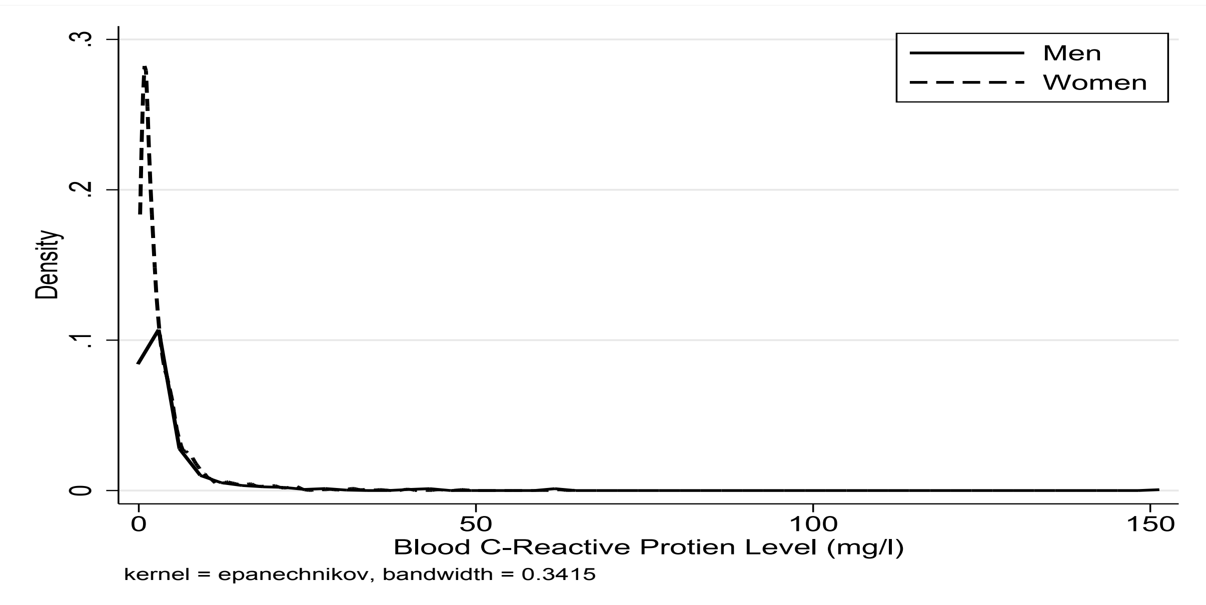

Figure 2.8.3: Kernel Density Estimate of C- Reactive Protein

Notes: This figure graphs the kernal densities of c-reactive protein levels of the analytical sample for men and women measured in milligrams per litre (mg/l). Mean c-reactive protien is 2.17 mg/l for men and 2.49 mg/l for women

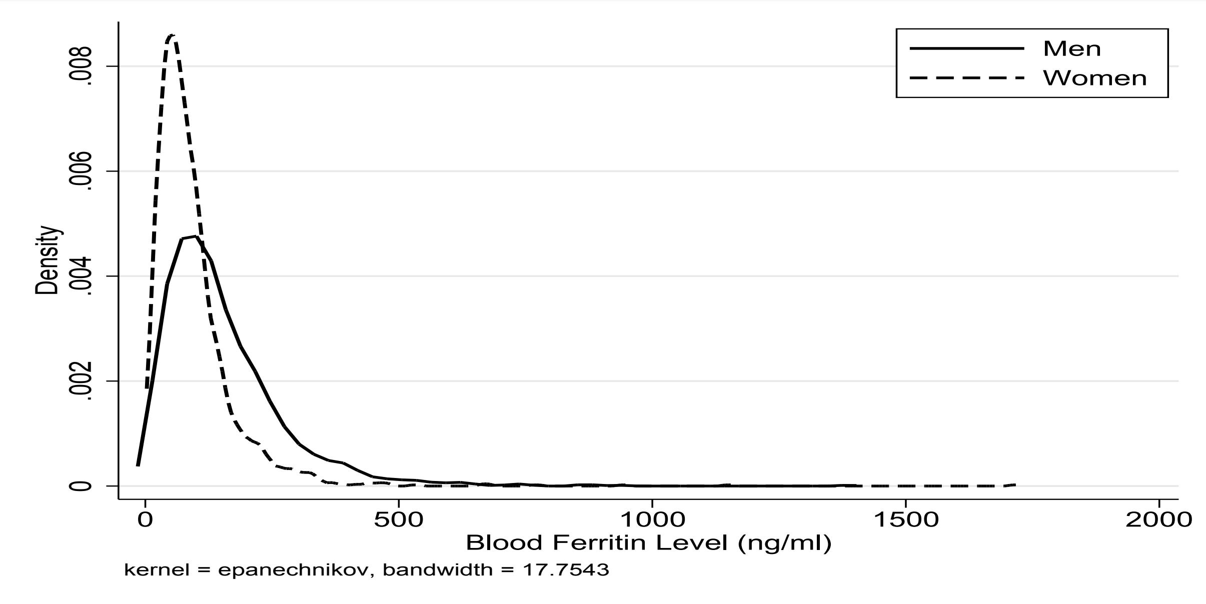

Figure 2.8.4: Kernel Density Estimate of Ferritin

Notes: This figure graphs the kernal densities of ferritin levels of the analytical sample for men and women measured in microgram per litre (ug/l). Mean ferritin is 153.2ug/l for men and 89.5ug/l for women

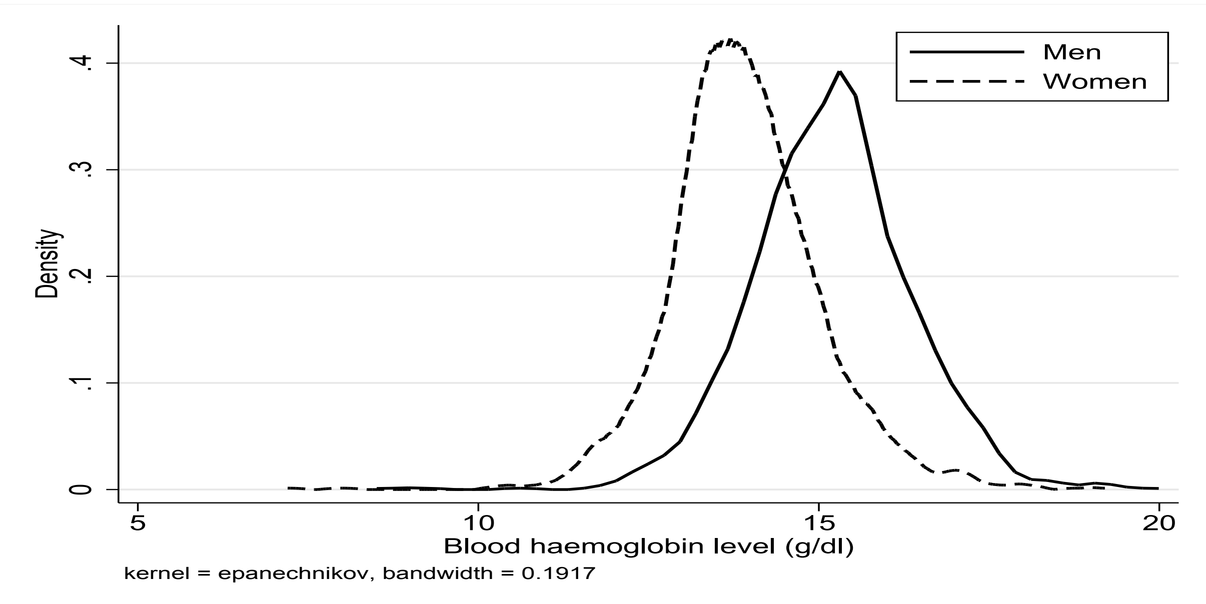

Figure 2.8.5: Kernel Density Estimate of Haemoglobin

Notes: This figure graphs the kernal densities of haemoglobin levels of the analytical sample for men and women measured in grams (g) per decilitre (dL). Mean haemoglobin is 15.28 g/dL for men and 13.98 g/DL for women

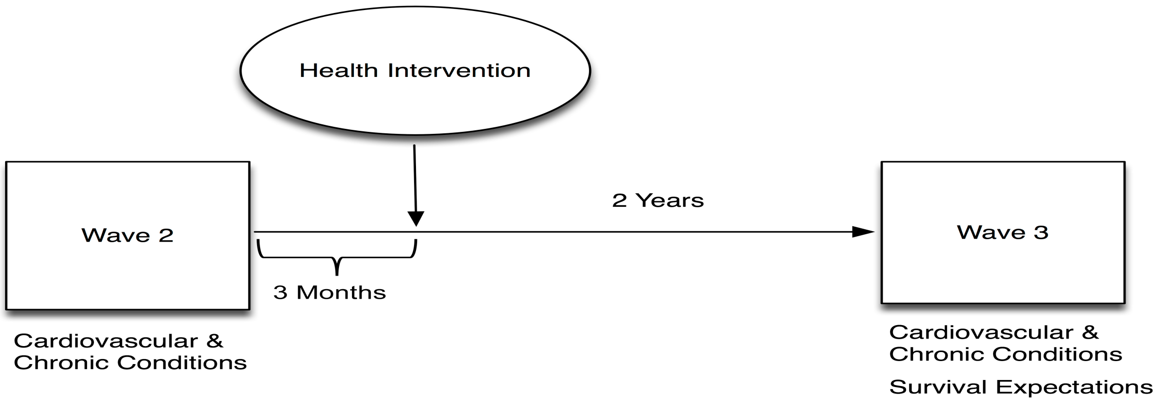

Figure 2.8.6: Timeline of Survey

Notes: This figure presents a time of the survey waves used in this analysis. Wave 2 was conducted in the year 2004 and wave 3 was conducted in the year 2006. Cardiovascular and chronic conditions were collected in both waves and all preceeding waves while survival expectations was collected only in wave 3

Table 2.8.1: Analytical Sample

Wave 2 Sample Criteria	N	Wave 3 Sample Criteria	N
Total number of C1CM respondents	8780		
Respondents who consented to nurse visit	7666		
Individuals with partial or full blood samples	6228	Total number of C1CM respondents	7535
Individuals < 70 years of age*	4092	Individuals <70 years of age	4368
Matched Analytical Sample** N = 3158			

* Sample with viable blood information in wave 2 used to estimate lee bounds in Section 2.6.1
** Analytical sample used in main estimations consisting of respondents aged less than 70 years of age with viable blood information at wave 2
C1CM refers to Cohort 1 Core Members who are ELSA sample members from Wave 0 onwards. This means that the analysis in this paper does not include respondents from any of the ELSA refreshment samples.

Table 2.8.2: Test of balance between demographics of wave 2 and wave 3 samples

	(1) Wave 2 Sample	(2) Analytical Sample from Wave 3	(3) t-test*
Low Income	0.34	0.32	0.10
Middle Income	0.33	0.31	0.53
High Income	0.33	0.37	0.02
Age <59	0.39	0.41	0.00
Age 60-64	0.29	0.32	0.01
Age 65-69	0.32	0.27	0.00
Single/Never Married	0.05	0.05	0.93
Married or Cohabiting	0.74	0.74	0.72
Divorced/Separated	0.13	0.13	0.93
Widowed	0.08	0.08	0.42
Male	0.46	0.46	0.78
Female	0.54	0.54	0.78
Parental Death - Mother	0.06	0.06	0.81
Parental Death - Father	0.03	0.03	0.81
Current/Previous Smoker	0.62	0.62	0.97
Consumed alcohol in last 7 days	0.65	0.66	0.25
Education**: NVQ 3 & above		0.37	
Education: NVQ 1 - NVQ 2		0.42	
Education: No Education		0.21	
N	4092	3158	

Column (1) presents the means for the respondents in the Wave 2 sample that have full/partial blood information.
Column (2) presents the means for the analytical sample used in this paper. The way this sample has been selected is shown in 2.8.1.
*p-values of a t-test of unpaired equality of means between Column(1) & Column(2)
** Educational Qualification information is only available from Wave 3 onwards.

Table 2.8.3: Means of survival expectations by demographics (weighted)

Variable	Categories	Survival to age 85 years	t-test*
Income	Low Income	44.80	
	Middle Income	49.80	
	High Income	51.70	0.00
Age	Age: $<$ 59 years	49.26	
	Age: 60 - 64 years	47.73	
	Age: 65 - 69 years	49.48	0.61
Marital Status	Never Married	48.92	
	Married or Civil Partnership	49.77	
	Divorced/Separated	45.66	
	Widowed	44.58	0.13
Gender	Men	45.50	
	Women	52.06	0.13
Education	Education: NVQ 3 & above	51.00	
	Education: NVQ 1 - NVQ 2	49.94	
	Education: No Education	43.67	0.04
Parental Death	Mother is alive	48.75	
	Mother not alive	50.32	0.99
	Father is alive	48.76	
	Father is not alive	51.51	0.91
Previous/Current Smoker	No	52.75	
	Yes	46.50	0.00
Consumed alcohol in last 7 days	No	45.76	
	Yes	50.49	0.00
All		49.39	
N		3158	

*p-value of an ANOVA test to analyse the differences in group means
Age is shown as a categorical variable here for descriptive purposes. Age is included as age fixed effects in all the regressions.
For the purpose of this paper, educational qualification is classified into 3 categories; No Qualification, NVQ 1 - NVQ 2 which is equivalent to GCSE grades A*- C and NVQ 3 and above which is equivalent to A level and above

Table 2.8.4: Means of disease prevalence by waves

	(1) Mean at Wave 2	(2) Mean at Wave 3	(3) Percentage Change
Panel A - Cardiovascular Conditions			
High Blood Pressure	0.35	0.40	4.94
Angina	0.06	0.06	0.51
Heart Attack	0.04	0.04	0.47
Chronic Heart Failure	0.00	0.00	0.06
Heart Murmur	0.04	0.04	0.38
Abnormal Heart Rhythm	0.05	0.06	1.08
Diabetes	0.04	0.04	0.00
Stroke	0.02	0.02	0.54
High Cholesterol	0.18	0.34	16.02
Any Cardiovascular Disease	0.50	0.62	11.34
Panel B - Chronic Conditions			
Lung Disease	0.06	0.06	0.73
Asthma	0.13	0.14	0.60
Arthritis	0.32	0.37	4.62
Osteoporosis	0.05	0.06	1.23
Cancer	0.06	0.07	1.27
Parkinson's	0.00	0.00	0.10
Psychiatric Disorder	0.12	0.13	1.04
Any Chronic Disease	0.51	0.55	4.69
N	3158	3158	

Column (1) is the proportion of the analytical sample with the corresponding condition at wave 2
Column (2) is the proportion of the analytical sample with the corresponding condition at wave 3
Column (3) is the percentage change in disease prevalence in the analytical sample between wave 2 and wave 3

Table 2.8.5: Means of survival expectations by disease prevalence (weighted)

	(1) No	(2) Yes	(3) Change	(4) t-test*
Panel A - Cardiovascular Conditions				
High Blood Pressure	50.67	46.07	-4.60	0.00
Angina	49.57	37.91	-11.66	0.00
Heart Attack	49.38	36.60	-12.78	0.00
Chronic Heart Failure	48.93	22.25	-26.68	0.00
Heart Murmur	48.93	46.45	-2.48	0.16
Abnormal Heart Rhythm	49.26	42.40	-6.86	0.00
Diabetes	49.36	35.63	-13.73	0.00
Stroke	49.13	36.45	-12.68	0.00
High Cholesterol	50.06	46.50	-3.56	0.00
Any Cardiovascular Disease	52.17	46.76	-5.41	0.00
Panel B - Chronic Conditions				
Lung Disease	49.47	39.37	-10.10	0.00
Asthma	49.62	43.72	-5.90	0.00
Arthritis	50.21	46.43	-3.78	0.00
Osteoporosis	49.12	44.30	-4.82	0.00
Cancer	49.27	43.03	-6.24	0.00
Parkinsons	48.89	34.64	-14.25	0.03
Psychiatric Disorder	50.06	45.01	-5.05	0.00
Any Chronic Disease	52.39	45.87	-6.52	0.00

Column (1) presents the weighted mean survival expectations for respondents without the corresponding health condition.
Column (2) presents the weighted mean survival expectations for respondents with the corresponding health condition.
Column (3) is the percentage change in survival expectations between respondents with the corresponding health condition and without.
*p-value of an unpaired t-test for equality of means between Column (1) & Column (2)

Table 2.8.6: Summary Statistics of out-of-range biomarker indicators

	(1) Total	(2) Men	(3) Women	(4) t-test*
Blood Glycated Haemoglobin	0.04	0.05	0.03	0.00
	[N= 3158]	[N=1455]	[N=1703]	
C - Reactive Protein	0.24	0.21	0.26	0.00
	[N= 3158]	[N=1455]	[N=1703]	
Ferritin	0.05	0.09	0.02	0.00
	[N= 3158]	[N=1455]	[N=1703]	
Haemoglobin	0.03	0.02	0.03	0.12
	[N= 2968]	[N=1372]	[N=1596]	

*p-value of an unpaired t-test for equality of means between men & women
All 4 of the blood analytes are coded are binary indicators with 1 indicating an out-of-range result. High blood glycated haemoglobin is defined using a cut-off point of >48mmol/mol, c-reactive protein is defined using a cut-off point of >3 mg/l, ferritin is defined using a cut-off point of >300ug/l and haemoglobin is defined using a cut-off point of <12g/dL for women and <13g/dL for men.
Test for low levels of haemoglobin is only done if there remains viable blood sample after the tests for blood glycated haemoglobin, c-reactive protien and ferritin. 2968 respondents out of the 3158 respondents had a viable test result for the haemoglobin biomarker.

Table 2.8.7: Association between out-of-range Blood Glycated Haemoglobin and Survival Expectations

	(1) Survival Expectations	(2) Survival Expectations	(3) Survival Expectations	(4) Survival Expectations
Blood Glycated Haemoglobin	-13.722***	-9.701***	-8.448***	-8.291***
	(2.567)	(2.594)	(2.619)	(2.600)
High Income				
Middle Income		-1.217	-1.273	-0.979
		(1.169)	(1.167)	(1.164)
Low Income		-4.419***	-4.262***	-3.875***
		(1.310)	(1.308)	(1.299)
Never Married				
Married or Civil Partner		-2.835	-2.829	-2.837
		(2.301)	(2.311)	(2.269)
Divorced/Separated		-5.257**	-5.249**	-5.001*
		(2.610)	(2.616)	(2.566)
Widowed		-6.991**	-7.098**	-7.022**
		(2.921)	(2.922)	(2.880)
Male				
Female		7.629***	7.605***	8.480***
		(1.009)	(1.008)	(1.003)
NVQ 3 and above				
NVQ 1 - NVQ 2		-0.980	-0.942	-0.892
		(1.090)	(1.089)	(1.084)
No Qualification		-6.178***	-6.043***	-5.746***
		(1.441)	(1.440)	(1.436)
Parental Death - Mother		1.581	1.300	1.384
		(2.042)	(2.049)	(2.053)
Parental Death - Father		1.749	1.450	1.274
		(2.793)	(2.805)	(2.761)
Current/Previous Smoker		-4.607***	-4.588***	-4.201***
		(0.983)	(0.982)	(0.973)
Consumed alcohol in the last 7 days		3.721***	3.574***	3.269***
		(1.066)	(1.063)	(1.060)
Pre-existing Cardiovascular Conditions			-3.662***	-3.105***
			(0.970)	(0.968)
Pre-existing Chronic Conditions				-5.621***
				(0.967)
Age fixed effects		✓	✓	✓
Observations	3158	3158	3158	3158

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Survival expectations are recorded on a scale of 0 - 100. Blood Glycated Haemoglobin is a binary indicator with 1 indicating a negative health information shock of high levels of blood glygated haemoglobin in the previous wave.

Column (1) presents the unconditional association between an out-of-range result and survival probability

Column (2) conditions on socio-demographic covariates

Column (3) additionally conditions on pre-existing doctor diagnosed cardiovascular conditions

Column (4) additionally conditions on pre-existing doctor diagnosed chronic conditions

Table 2.8.8: Association between out-of-range CRP and Survival Expectations

	(1) Survival Expectations	(2) Survival Expectations	(3) Survival Expectations	(4) Survival Expectations
C-Reactive Protein	-3.380***	-2.466**	-2.251**	-1.858
	(1.150)	(1.137)	(1.137)	(1.133)
High Income				
Middle Income		-1.225	-1.279	-1.006
		(1.166)	(1.165)	(1.161)
Low Income		-4.463***	-4.275***	-3.919***
		(1.317)	(1.314)	(1.306)
Never Married				
Married or Civil Partner		-2.737	-2.739	-2.761
		(2.313)	(2.322)	(2.279)
Divorced/Separated		-5.178**	-5.179**	-4.938*
		(2.625)	(2.629)	(2.579)
Widowed		-6.818**	-6.954**	-6.895**
		(2.924)	(2.925)	(2.883)
Male				
Female		8.029***	7.953***	8.796***
		(1.009)	(1.007)	(1.002)
NVQ 3 and above				
NVQ 1 - NVQ 2		-0.911	-0.876	-0.836
		(1.091)	(1.089)	(1.084)
No Qualification		-6.153***	-6.001***	-5.730***
		(1.445)	(1.445)	(1.442)
Parental Death - Mother		1.574	1.272	1.346
		(2.040)	(2.048)	(2.049)
Parental Death - Father		1.603	1.293	1.135
		(2.821)	(2.832)	(2.782)
Current/Previous Smoker		-4.635***	-4.606***	-4.235***
		(0.986)	(0.985)	(0.976)
Consumed alcohol in the last 7 days		3.966***	3.765***	3.474***
		(1.068)	(1.064)	(1.060)
Pre-existing Cardiovascular Conditions			-3.995***	-3.452***
			(0.956)	(0.955)
Pre-existing Chronic Conditions				-5.556***
				(0.971)
Age fixed effects		✓	✓	✓
Observations	3158	3158	3158	3158

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Survival expectations are recorded on a scale of 0 - 100. C-reactive protien is a binary indicator with 1 indicating a negative health information shock of high levels of c-reactive protien in the previous wave.

Column (1) presents the unconditional association between an out-of-range result and survival probability

Column (2) conditions on socio-demographic covariates

Column (3) additionally conditions on pre-existing doctor diagnosed cardiovascular conditions

Column (4) additionally conditions on pre-existing doctor diagnosed chronic conditions

Table 2.8.9: Association between out-of-range Ferritin and Survival Expectations

	(1) Survival Expectations	(2) Survival Expectations	(3) Survival Expectations	(4) Survival Expectations
Ferritin	-5.000**	-2.737	-2.317	-2.282
	(2.313)	(2.346)	(2.321)	(2.307)
High Income				
Middle Income		-1.351	-1.394	-1.096
		(1.167)	(1.165)	(1.162)
Low Income		-4.649***	-4.443***	-4.051***
		(1.312)	(1.309)	(1.301)
Nver Married				
Married or Civil Partner		-2.732	-2.741	-2.751
		(2.315)	(2.324)	(2.280)
Divorced/Separated		-5.082*	-5.099*	-4.853*
		(2.626)	(2.630)	(2.579)
Widowed		-6.781**	-6.930**	-6.856**
		(2.927)	(2.927)	(2.884)
Male				
Female		7.731***	7.693***	8.570***
		(1.020)	(1.018)	(1.012)
NVQ 3 and above				
NVQ 1 - NVQ 2		-0.996	-0.952	-0.901
		(1.089)	(1.088)	(1.083)
No Qualification		-6.326***	-6.157***	-5.855***
		(1.443)	(1.442)	(1.439)
Parental Death - Mother		1.483	1.188	1.274
		(2.031)	(2.039)	(2.041)
Parental Death - Father		1.737	1.409	1.234
		(2.797)	(2.808)	(2.760)
Current/Previous Smoker		-4.745***	-4.704***	-4.313***
		(0.987)	(0.986)	(0.977)
Consumed alcohol in the last 7 days		4.054***	3.844***	3.533***
		(1.067)	(1.062)	(1.059)
Pre-existing Cardiovascular Conditions			-4.034***	-3.467***
			(0.957)	(0.955)
Pre-existing Chronic Conditions				-5.654***
				(0.969)
Age fixed effects		✓	✓	✓
Observations	3158	3158	3158	3158

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Survival expectations are recorded on a scale of 0 - 100. Ferritin is a binary indicator with 1 indicating a negative health information shock of high levels of ferritin in the previous wave.
Column (1) presents the unconditional association between an out-of-range result and survival probability
Column (2) conditions on socio-demographic covariates
Column (3) additionally conditions on pre-existing doctor diagnosed cardiovascular conditions
Column (4) additionally conditions on pre-existing doctor diagnosed chronic conditions

Table 2.8.10: Association between out-of-range Haemoglobin and Survival Expectations

	(1) Survival Expectations	(2) Survival Expectations	(3) Survival Expectations	(4) Survival Expectations
Haemoglobin	-10.006***	-9.678***	-9.338***	-8.850***
	(3.064)	(2.695)	(2.672)	(2.656)
High Income				
Middle Income		-1.022	-1.050	-0.719
		(1.191)	(1.189)	(1.183)
Low Income		-4.039***	-3.850***	-3.454***
		(1.334)	(1.332)	(1.323)
Never Married				
Married or Civil Partner		-3.355	-3.334	-3.350
		(2.353)	(2.364)	(2.322)
Divorced/Separated		-5.291**	-5.310**	-5.107*
		(2.681)	(2.688)	(2.634)
Widowed		-8.363***	-8.470***	-8.384***
		(3.027)	(3.027)	(2.984)
Male				
Female		8.138***	8.082***	9.003***
		(1.028)	(1.026)	(1.022)
NVQ 3 and above				
NVQ 1 - NVQ 2		-1.191	-1.154	-1.147
		(1.116)	(1.115)	(1.109)
No Qualification		-7.171***	-7.024***	-6.749***
		(1.462)	(1.462)	(1.457)
Parental Death - Mother		1.941	1.651	1.751
		(2.128)	(2.135)	(2.141)
Parental Death - Father		0.399	0.157	-0.002
		(2.922)	(2.934)	(2.880)
Current/Previous Smoker		-4.590***	-4.543***	-4.083***
		(1.002)	(1.001)	(0.993)
Consumed alcohol in the last 7 days		3.797***	3.607***	3.256***
		(1.091)	(1.087)	(1.083)
Pre-existing Cardiovascular Conditions			-3.665***	-3.064***
			(0.982)	(0.981)
Pre-existing Chronic Conditions				-5.885***
				(0.991)
Age fixed effects		✓	✓	✓
Observations	2968	2968	2968	2968

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Survival expectations are recorded on a scale of 0 - 100. Haemoglobin is a binary indicator with 1 indicating a negative health information shock of low levels of haemoglobin in the previous wave.
Column (1) presents the unconditional association between an out-of-range result and survival probability
Column (2) conditions on socio-demographic covariates
Column (3) additionally conditions on pre-existing doctor diagnosed cardiovascular conditions
Column (4) additionally conditions on pre-existing doctor diagnosed chronic conditions

Table 2.8.11: Linear Probability Results - All negative health information shocks

	(1) Survival Expectations	(2) Survival Expectations	(3) Survival Expectations	(4) Survival Expectations
Blood Glycated Haemoglobin	-12.886***	-9.290***	-8.273***	-8.214***
	(2.572)	(2.598)	(2.626)	(2.597)
C-Reactive Protein	-3.242***	-2.502**	-2.356**	-1.954*
	(1.165)	(1.149)	(1.149)	(1.143)
Ferritin	-4.794**	-2.697	-2.398	-2.302
	(2.319)	(2.365)	(2.347)	(2.330)
Haemoglobin	-9.462***	-9.362***	-9.117***	-8.619***
	(2.963)	(2.658)	(2.645)	(2.621)
High Income				
Middle Income		-0.774	-0.818	-0.515
		(1.190)	(1.189)	(1.183)
Low Income		-3.578***	-3.459***	-3.108**
		(1.334)	(1.333)	(1.324)
Never Married				
Married or Civil Partner		-3.152	-3.151	-3.187
		(2.326)	(2.337)	(2.298)
Divorced/Seperated		-5.149*	-5.178*	-4.992*
		(2.659)	(2.665)	(2.615)
Widowed		-8.182***	-8.287***	-8.224***
		(3.018)	(3.017)	(2.976)
Male				
Female		7.739***	7.741***	8.631***
		(1.035)	(1.034)	(1.031)
NVQ 3 and above				
NVQ 1 - NVQ 2		-1.157	-1.126	-1.129
		(1.117)	(1.116)	(1.110)
No Qualification		-6.906***	-6.804***	-6.560***
		(1.460)	(1.461)	(1.456)
Parental Death - Mother		2.013	1.765	1.850
		(2.131)	(2.138)	(2.147)
Parental Death - Father		0.392	0.185	0.043
		(2.920)	(2.934)	(2.884)
Current/Previous Smoker		-4.424***	-4.399***	-3.963***
		(0.999)	(0.998)	(0.990)
Consumed alcohol in the last 7 days		3.309***	3.196***	2.876***
		(1.093)	(1.090)	(1.087)
Pre-existing Cardiovascular Conditions			-3.079***	-2.516**
			(0.997)	(0.993)
Pre-existing Chronic Conditions				-5.741***
				(0.993)
Constant	50.807***	56.364***	57.493***	57.842***
	(0.582)	(4.740)	(4.780)	(4.688)
Observations	2968	2968	2968	2968

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Survival expectations are recorded on a scale of 0 - 100. Blood Glycated Haemoglobin, C-reactive protien, Ferritin and Haemoglobin are all binary indicators with 1 indicating a negative health information shock of the corresponding condition in the previous wave.

Column (1) presents the unconditional association between all out-of-range results and survival probability

Column (2) conditions on socio-demographic covariates

Column (3) additionally conditions on pre-existing doctor diagnosed cardiovascular conditions

Column (4) additionally conditions on pre-existing doctor diagnosed chronic conditions

Table 2.8.12: Tightened Lee treatment effect bounds for blood glycated haemoglobin

Number of obs. = 4092						
Number of selected obs. = 3158						
Number of cells = 16						
Overall trimming proportion = 0.0533						
Effect 95% conf. interval* : [-22.9813	-2.9173]					
Survival Expectations	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
Blood Glycated Haemoglobin						
lower	-15.6326	4.307943	-3.63	0.000	-24.076	-7.18917
upper	-11.5375	5.053337	-2.28	0.022	-21.44189	-1.63317

Estimates are tightened using age fixed effects and weighted by wave 3 cross-sectional weights
Survival expectations are recorded on a scale of 0 - 100. Blood Glycated Haemoglobin is a binary indicator with 1 indicating a negative health information shock of high levels of blood glygated haemoglobin in the previous wave.
*Confidence interval for treatment effects

Table 2.8.13: Tightened Lee treatment effect bounds for C - Reactive Protein

Number of obs. = 4092						
Number of selected obs. = 3158						
Number of cells = 16						
Overall trimming proportion = 0.0028						
Effect 95% conf. interval* : [-6.2852	-0.5320]					
Survival Expectations	Coef.	Std. Err.	z	$P > z$	[95% Conf.	Interval]
C - Reactive Protein						
lower	-3.73384	1.443205	-2.59	0.01	-6.562466	-0.90521
upper	-3.00224	1.3973	-2.15	0.032	-5.740893	-0.26358

Estimates are tightened using age fixed effects and weighted by wave 3 cross-sectional weights
Survival expectations are recorded on a scale of 0 - 100. C-reactive protien is a binary indicator with 1 indicating a negative health information shock of high levels of c-reactive protien in the previous wave.
*Confidence interval for treatment effects

Table 2.8.14: Tightened Lee treatment effect bounds for Ferritin

Number of obs. = 4092						
Number of selected obs. = 3158						
Number of cells = 16						
Overall trimming proportion = 0.0471						
Effect 95% conf. interval* : [-10.6636	0.7607]					
Survival Expectations	Coef.	Std. Err.	z	$P > z$	[95% Conf.	Interval]
Ferritin						
lower	-6.1283	2.680673	-2.29	0.022	-11.38229	-0.87424
upper	-3.6915	2.631574	-1.4	0.161	-8.849302	1.466278

Estimates are tightened using age fixed effects and weighted by wave 3 cross-sectional weights
Survival expectations are recorded on a scale of 0 - 100. Ferritin is a binary indicator with 1 indicating a negative health information shock of high levels of ferritin in the previous wave.
Column (1) presents the unconditional association between an out-of-range result and survival probability
*Confidence interval for treatment effects

Table 2.8.15: Tightened Lee treatment effect bounds for Haemoglobin

Number of obs. = 3834						
Number of selected obs. = 2968						
Number of cells = 16						
Overall trimming proportion = 0.0873						
Effect 95% conf. interval* : [-15.7102	-2.7919]					
Survival Expectations	Coef.	Std. Err.	z	$P > z$	[95% Conf.	Interval]
Haemoglobin						
lower	-8.7795	3.536317	-2.48	0.013	-15.71055	-1.84844
upper	-9.6571	3.502891	-2.76	0.006	-16.52262	-2.79154

Estimates are tightened using age fixed effects and weighted by wave 3 cross-sectional weights
Survival expectations are recorded on a scale of 0 - 100. Haemoglobin is a binary indicator with 1 indicating a negative health information shock of low levels of haemoglobin in the previous wave.
*Confidence interval for treatment effects

Table 2.8.16: Current/Past Smokers versus Non-Smokers -

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Smokers	Non-Smokers	Smokers	Non-Smokers	Smokers	Non-Smokers	Smokers	Non-Smokers
Ferritin	-1.987	-3.641						
	(3.099)	(3.422)						
Blood Glycated Haemoglobin			-7.166**	-11.389**				
			(3.096)	(5.182)				
C-Reactive Protein					-1.915	-1.875		
					(1.432)	(1.838)		
Haemoglobin							-10.749***	-6.553*
							(3.841)	(3.674)
Demographics	✓	✓	✓	✓	✓	✓	✓	✓
Pre-existing Cardiovascular Conditions	✓	✓	✓	✓	✓	✓	✓	✓
Pre-existing Chronic Conditions	✓	✓	✓	✓	✓	✓	✓	✓
Age Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1960	1198	1960	1198	1960	1198	1840	1128
Mean (Dependent Variable)	47.06	53.22	47.06	53.22	47.06	53.22	47.06	53.22
sd. (Dependent Variable)	26.50	24.50	26.50	24.50	26.50	24.50	26.50	24.50

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Survival expectations are recorded on a scale of 0 - 100. Blood Glycated Haemoglobin, C-reactive protien, Ferritin and Haemoglobin are all binary indicators with 1 indicating a negative health information shock of the corresponding condition in the previous wave.
Columns (1), (3), (5) & (7) present the estimates for individuals who are current or previous smokers.
Columns (2), (4), (6) & 87) present the estimates for individuals who do not smoke.

Table 2.8.17: No Education versus At least NVQ 1 -

	(1) No Education	(2) Education	(3) No Education	(4) Education	(5) No Education	(6) Education	(7) No Education	(8) Education
Ferritin	-1.233	-2.428						
	(5.473)	(2.583)						
Blood Glycated Haemoglobin			-11.058**	-7.278**				
			(4.477)	(3.241)				
C-Reactive Protein					-3.151	-1.446		
					(2.353)	(1.292)		
Haemoglobin							-19.652***	-4.284
							(4.187)	(3.113)
Demographics	✓	✓	✓	✓	✓	✓	✓	✓
Pre-existing Cardiovascular Conditions	✓	✓	✓	✓	✓	✓	✓	✓
Pre-existing Chronic Conditions	✓	✓	✓	✓	✓	✓	✓	✓
Age Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	664	2494	664	2494	664	2494	628	2340
Mean (Dependent Variable)	44.01	50.83	44.01	50.83	44.01	50.83	44.01	50.83
sd. (Dependent Variable)	27.28	25.37	27.28	25.37	27.28	25.37	27.28	25.37

Standardized beta coefficients; Robust standard errors in parentheses clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Survival expectations are recorded on a scale of 0 - 100. Blood Glycated Haemoglobin, C-reactive protien, Ferritin and Haemoglobin are all binary indicators with 1 indicating a negative health information shock of the corresponding condition in the previous wave.
Columns (1), (3), (5) & (7) present the estimates for individuals who have no educational qualifications.
Columns (2), (4), (6) & 87) present the estimates for individuals who have NVQ 1 or above.

Chapter 3

The effect of domestic violence on neonatal & infant mortality

3.1 Introduction

According to the World Health Organisation classification, India falls within a region characterised by high mortality. Even within this region, with a child mortality rate of 48 per 1000 live births, India has higher child mortality than most of its neighbours (United Nations Population Division (UN), 2015). Over the years, India has established several programmes aimed at reducing child mortality, from the Family Welfare Programme in 1977 to the National Rural Health Mission in 2005.[1] These programmes rely on equitable healthcare and improved access to public health services with a distinct focus on rural areas and low socio economic status groups. Despite these efforts, India failed to achieve its former objective to reduce the Under Five Mortality Rate(U5MR) to less than 100 per 1000 births by the year 2000 (UNICEF, 2012). UN data also suggests India is likely to miss its Millennium Development Goal to reduce the child mortality rate to 42 per 1,000 live births by the end of 2015. Several social and economic factors beyond access to healthcare can influence child mortality. One such factor in the context of India is domestic violence.

A number of studies have investigated the association between domestic violence against the mother and child mortality (See Section 3 for a review). However, a limitation thus far is that none of them account for the potential endogeneity of domestic violence. Endogeneity in this instance may arise due to three concerns. The first of these is the high plausibility of systematic underreporting of domestic violence in India. For instance, women from rural areas may systematically underestimate their experience of violence and thus unknowingly underreport it. Similarly, women who do not work outside the house may have a systematically higher threshold of privacy and so withhold information about the existence and extent of domestic violence leading to measurement error in the data generating process. The second concern arises due to the problem of omitted variable bias in linear probability estimation procedures of child mortality. For example a low level of confidence of the mother, which is not captured in the data, may make her more vulnerable to domestic violence and simultaneously restrict the level of childcare she chooses to access. This non-ignorable selection can lead to inconsistent estimation in standard linear probability techniques (Clarke and Windmeijer, 2012). The third is the possibility of reverse causality of domestic violence and child mortality. Since the timing of the violence is not available in the data, it is impossible to ascertain whether the experience

[1]A comprehensive list of initiatives can be found in Infant and Child Mortality in India Levels, Trends and Determinants, NIMS, ICMR and UNICEF, 2012.

of violence caused child mortality or the loss of a child subsequently led to domestic violence within the marriage. Thus the outcome variable may impinge on the independent variable, causing linear probability estimates of the impact to be biased.

This paper empirically tests if the positive association between domestic violence and mortality is just indicative of an association or represents a causal mechanism by exploiting exogenous variation in domestic violence through the real price of gold in India. A higher price of gold at the time of marriage of the mother reduces the share of gold jewellery in the dowry basket. This reduces the bargaining power of the woman within the household, which in turn exposes her to a higher risk of domestic violence. Domestic violence is measured as a factor score with a mean of 0 and standard deviation of 1. The instrumental variable estimates suggest that a one standard deviation increase in domestic violence against the mother increases the likelihood of both neonatal mortality and infant mortality by 17.6 and 16.9 percentage points respectively. The positive effect of domestic violence on child mortality persists through alternative specifications conditioning on child, mother and household socio-demographic factors and several temporal and spatial controls. I find no evidence of an alternative mechanism that may be driving these effects.

The rest of this paper is organised as follows. Section 3.2 provides a background on domestic violence. Section 3.3 outlines the potential mechanisms through which domestic violence affects child mortality. Section 3.4 describes the data and analytical sample. Section 3.5 outlines the identification strategy and presents a detailed discussion on the instrument and the exclusion restriction. Section 3.6 reports the main results. Section 3.7 tackles several robustness checks and investigates heterogeneous effects and Section 3.8 concludes.

3.2 Background on Domestic Violence

Violent behaviour is a recognised multifaceted problem with negative consequences for the individual, the economy and for the society as a whole. Unfortunately women face an increased threat of violent behaviour. Gendered violence is present without boundaries in every country irrespective of diverse social, economic and political backgrounds.

3.2.1 Externalities and Costs of Domestic Violence

Domestic violence has far reaching consequences not just for the victim of abuse and for the household, but also for the economy of the country as a whole. The medical, policing and judicial costs due to violence have been quantified in a few developed countries as staggering amounts, such as 1.1 billion Dollars (Canadian) in Canada (Zhang, 2012) to 23 billion British Pounds per annum in Great Britain (Walby, 2004). Over the years, a number of in-depth analyses have been conducted in Western countries to determine the causes and quantify the effects of domestic violence (Tauchen et al., 1985) (Farmer and Tiefenthaler, 1997) (Iyengar, 2009). In contrast, the number of studies done in developing countries is extremely limited due to the lack of nationally representative data of good quality. This void in the literature is of particular significance as the negative effects of violence have a multiplier effect in these countries due to the continued persistence of adverse social and economic conditions. Predictably, the estimated rate of violent death in low and middle income countries is twice that of a high income country (Waters et al., 2004). In developing countries, violence against women causes more death and disability than cancer, malaria, traffic accident and wars combined (Morrison and Orlando, 1999).

3.2.2 Prevalence and Risk Factors

Although most societies look down upon gendered violence, in India the reality is that they are often endorsed under the garb of cultural practices, collective norms or religious beliefs. The prevalence rate of domestic violence in itself occupies a large variation among differing reports, from 17% (Martin et al., 1999) to 41% (Peedicayil et al., 2004). One possible explanation for this is the non-standardisation of survey questions regarding violence in the various reports and differences in the subjective interpretation of violence. However, a more likely explanation is the under-reporting of incidences due to the social stigma attached to violence (Kishor and Johnson, 2004) and/or the underestimation of violence in itself. Actual prevalence of violence in India is therefore at a risk of underestimation and is thought by experts to be much higher than reported.

Peedicayil et al. (2004) estimate the prevalence of physical violence and determines the factors associated with the violence during pregnancy in India. Overall, 41% of the sample had experienced some form of physical violence, out of which 12.9% also experienced violence during pregnancy. Factors that are associated with the risk of domestic violence are having husbands who consume alcohol, husbands having an affair, dowry harassment and husbands accusing the wife of having an extramarital affair. Other significant risk factors include a husband's low education, a husband's substance abuse, no social support, three or more children and household crowding. Recently, several studies have also identified financial stress faced by the household as a significant risk factor in determining domestic violence. Being an agrarian society, local precipitation shocks, too, have shown to have a significant effect on domestic violence. In periods of drought, husbands may attempt to extract more surplus from the wife to smooth their own consumption and thereby increase domestic violence and dowry deaths (Sekhri and Storeygard, 2014).

3.3 Domestic Violence and Child Mortality

One of the detrimental effects of domestic violence arises from the fact that women may be abused during pregnancy. In addition to the apparent and well researched reduction in women's welfare, violence during pregnancy can have an impact on child mortality through various mechanisms. The most direct mechanism is through the effect of blunt physical trauma and the resulting harm caused to the foetus (Nasir and Hyder, 2003). A second mechanism is through the deterrent effect that violence has on women's access to pre-natal healthcare (Petersen et al., 2001). Third, persistence of post natal domestic violence has a negative impact on child care, especially in terms of restricted access to post-natal healthcare and inadequate maternal nutrition (Newberger et al., 1992). Fourth, women who experience violence also tend to have higher levels of psychological stress, which is associated with low birth weight or pre-term delivery and are well known risk factors for neonatal and infant mortality (Campbell et al., 1999).

There has been previous work attempting to ascertain a causal estimate of the negative externalities of domestic violence in other countries. For instance Aizer (2011) uses variation in the enforcement of laws in the United States and finds that hospitalisation for an assault while pregnant reduces birthweight by 163 grams. Agüero (2013) uses spatial and temporal variation in the creation of shelters for women in Peru to show that domestic violence against the mother has a causal effect on the short term health outcomes of their children.

A few studies have established an association between domestic violence and child mortality in India. Rawlings and Siddique (2014) examine the effects of domestic abuse on child mortality across 30 different developing countries one of which is India. They find that children born to mothers who have experienced violence are 0.4 percentage points more likely to die in the neonatal stage and 0.9 percentage points more likely to die in the infancy stage of life. Jejeebhoy (1998) explores the link between wife beating during pregnancy and foetal and infant death using data from a community based survey during 1993-94 in Uttar Pradesh in the North and Tamil Nadu in the South of India. It allowed the authors to test for regional and religious differences within India. For the sample as a whole, 40% of women experienced violence. They highlight the association between women's experiences of wife beating and infant and foetal loss, even when conditioning on several social, economic and geographical factors. The paper concludes that these associations are stronger and more significant in Uttar Pradesh than in Tamil Nadu, as women in Tamil Nadu have some measure of autonomy due to the state's egalitarian setting and kinship patterns (Dyson and Moore, 1983). A relatively recent investigation conducted by Koenig et al. (2010) is based upon a 2002-2003 follow-up study of a cohort selected from the 1998-99 National Family and Health Survey (NFHS 2) in four Indian states. The authors find that births to mothers who experienced multiple incidents of domestic violence had a 68% higher risk of perinatal and neonatal mortality. No differences in mortality rates were observed for births where the mother had experienced only one episode of violence. Similarly research by Ackerson and Subramanian (2009) analyses the effect of domestic violence on child mortality using the 2005-06 National Family Health Survey (NFHS 3). They find that maternal experience of physical violence increased mortality rates among all children and these associations do not differ according to the child's gender. Sexual and psychological violence were less strongly associated.

3.4 Data

Data from the National Family and Health Survey (NFHS 3), India's version of the Demographic Health Survey (International Institute for Population Sciences, 2007), is used for this study. The survey was fielded between November, 2005 and August, 2006, and is the third of a series of cross-sectional NFHS surveys. It is based on a sample of households which is representative both at national and state level; 124,385 eligible women of reproductive age (15-49) have completed interviews. The dataset contains a rich variety of information, including background characteristics, reproductive histories, antenatal, delivery and post natal care and husband's background. Within reproductive histories, births and deaths for children were recorded with the total number of births recorded at 256,782. A Status of Women and Spousal Violence module was also carried out in each of India's 29 states. Mother level data has been merged with birth level data in order to evaluate the mother's socio economic characteristics at the birth level.

3.4.1 Domestic Violence

The focus of this paper is to analyse the effect of physical violence on child mortality and the results primarily focus on this physical aspect of domestic violence. Physical violence is measured in much greater detail in the NFHS 3 (discussed below) as compared to other forms of violence. As an additional robustness check I also check if the results remain robust to the inclusion of sexual and emotional aspects of domestic violence in Section 3.7.2.

In order to minimise measurement error, one woman was selected at random from each household for the domestic violence module subject to privacy being obtained. The domestic violence module was asked at the end of the interview so that the interviewer had built a rapport with the respondent. Violence in the NFHS 3 is measured using the modified Conflict Tactics Scale (CTS) (Straus et al., 1973) using the following set of questions: *(Does/Did) your (last) husband ever do any of the following things to you?*

1. *Slap you?*
2. *Twist your arm or pull your hair?*
3. *Push you, shake you, or throw something at you?*
4. *Punch you with his fist or something that could hurt you?*
5. *Kick you, drag you or beat you up?*
6. *Threaten to attack you with a knife, gun or any other weapon?*
7. *Try to choke you or burn you on purpose?*

Out of a total sample of 124,385 mothers, 40,682 were not eligible for the domestic violence module as there was more than one eligible respondent within the household. 477 eligible respondents were not surveyed as privacy could not be obtained. Subject to these criteria, 83,703 eligible respondents were interviewed. Each of the above questions was allowed five responses as follows:

- *No*
- *Yes, but currently a widow or timing missing*

- *Not in the last 12 months*
- *Sometimes during the last 12 months*
- *Often during the last 12 months*

The 5 response options to the 7 questions are coded as ordinal variables ranging from 0 to 4 in the order listed above with 0 being *No*. The summary statistics of the responses to each of these questions is provided in Table 3.9.1. Given the ordinal scale of the responses, for the main specification I used the first component of a polychoric principle component analysis on all of the 5 responses to the 7 dimensions of violence. The first principle component accounts for 82% of the overall variability of domestic violence.

I also check if the results remain consistent in the case when a simple factor score of the 7 components is used as an index for domestic violence. Alternatively, I created an ordinal measure of domestic violence that is equal to the number of kinds of physical violence the respondent is exposed to. This is an index (0, 7) which is 0 if domestic violence does not exist in the household and progressively adds 1 for a non-zero response to each of the 7 questions mentioned above. Approximately 25% of the analytical sample have reported having experienced at least one form of physical violence. Table 3.9.2 provides the summary statistics of the components of domestic violence of this index.

3.4.2 Analytical Sample

The analysis is restricted to ever-married women who were a part of the domestic violence module and have had at least one birth. In addition, the sample is restricted to single live births and only includes marriages that occurred after 1991 to improve the strength of the identification of the instrument (See discussion in Section 3.5.2).

The analysis conditions on various socio-demographic factors that have been previously shown to have an association with child mortality in the given context. These are classified into child, mother, household and father level characteristics.

Table 3.9.3 presents the summary statistics at the parent and household level. The sample mean of number of years of education of the mother is 7.5 years, which translates into secondary education. The mean age of the mother at the time of the survey is 27 years. The average height of mothers in the sample is 15.2 metres and is representative of the national average. Mother's height is used as an indicator for the mother's overall stock of health (Strauss and Thomas, 2007) in the estimations. The average age of the father at the time of survey is 32.6 years. Fathers are on average more educated than mothers with 8.7 years of schooling. The sample has a equal proportion of households from rural and urban locations. The religious distribution of the analytical sample is indicative of the national representation with a 74% of the sample being from a Hindu household. Other major religions represented in the sample are Muslims at 9% and Christians at 11% . Due to low numbers the remaining religions such as Jain, Sikh, Jews etc. are combined into an other religion category which makes up 6% of the analytical sample. NFHS 3 constructs a wealth index using a principle component analysis based on information regarding ownership of household items, dwelling characteristics, home contruction materials and access to a bank or post office. This score is then divided into population quintiles with each quiltile given a rank from 1 (poorest) to 5 (richest). 8% of the analytical sample used in this paper is from the poorest wealth quintile while 34% are from the richest quintile. 38% of the sample of mothers belong to a high caste.

Table 3.9.4 presents the summary statistics at the birth level. The analytical sample contains more boys than girls, which is unsurprising given the permeation of ultrasound technology and the rampant sex selection in India (Bhalotra and Cochrane, 2010). On average, mothers in the sample have two births and mother's mean age at the time of birth of the index child is 23 years.

3.4.3 Mortality

The dependent variable is a binary variable with 1 indicating mortality and 0 indicating survival of the child. Sample selection problems are eliminated since the birth histories are retrospective and inclusion in the sample is not restricted to survival at survey date. I created subsets of the data based on age at the time of death and 2 models are estimated on the following dependent variables:

1. Neonatal mortality: All deaths from the 1^{st} day of life to 30 days of life, conditional on children who were born 30 days before the date of the survey to allow for full exposure to the risk of neonatal mortality. So children who are less than 30 days old at the time of the survey have been excluded from the model.

2. Infant mortality: All deaths from the 1^{st} day of life up to the 1^{st} year of life, conditional on children who were born 1 year before the date of the survey to allow for full exposure

to the risk of infant mortality. So children who are less than 1 year old at the time of the survey have been excluded from the model.

The number of births that resulted in a death in the neonatal model and the infant model is 1,356 (3.10%) and 1,463 (3.71%) respectively as shown in Table 3.9.5. Sample means by violence reported versus violence not reported are shown in Table 3.9.6. The proportion of births that resulted in a death in the neonatal and infant model is lower when no violence was reported in both definitions of mortality. Even in a descriptive manner, the proportion of births that resulted in a death is higher for both outcomes when violence was reported. There is a statistically significant difference in the means of mortality between respondents who report the presence of domestic violence and respondents who do not report domestic violence in both the neonatal and infant stages of life.

3.5 Identification Strategy

A valid instrument that is strongly correlated with the domestic violence and is uncorrelated with child mortality is needed to overcome the potential endogeneity of domestic violence. This section first outlines the empirical specification and proceeds to discuss the validity of the real price of gold in India as a plausible source of exogenous variation.

3.5.1 Econometric Specification

Since the measure of domestic violence in the data is at the mother level, I estimate a mother level regression in the first stage of the instrumental variable estimation and a child level regression in the second stage. The first stage equation for mother *m* in state *s* is:

$$X_{ms} = \beta_1 Z_m + \beta_2 C_{ms} + \beta_3 V_s + \varepsilon_{ms}$$

where,
Z_m is the real price of gold at the month of marriage of the mother ;
X_{ms} is the domestic violence experienced by the mother m in state s ;
C_{ms} are socio demographic controls at the mother level;
V_s are state fixed effects;
ε_{ms} is the error term.

Second stage equation for index child *i*, to mother *m*, in state *s*, year of birth *t* and month of birth *j* is:

$$Y_{imjst} = \beta_1 \widehat{X}_{ms} + \beta_2 C_{imjst} + \beta_3 V_{jst} + \varepsilon_{imjst}$$

where,
Y_{imjst} is the outcome for each birth i to mother m in month j in state s and year t ;
$\widehat{X}_{ms}$ is the fitted values of domestic violence from the first stage equation ;
C_{imjst} are socio domegraphic controls at the mother level and at the child level;
V_{jst} are fixed effects for state, month of birth and year of birth;
ε_{imjst} is the error term.

Due to the potential endogeneity of domestic violence, this paper estimates an instrumental variable regression where violence is instrumented by the price of gold at the time of marriage of the mother. β_1 in the second stage equation is the coefficient of interest that defines the causal relationship between the fitted values of domestic violence and mortality. All estimations are weighted using the domestic violence weights provided in the NFHS 3 which account for differential probability to being selected into the domestic violence module.

India's socioeconomic conditions vary considerably, especially between northern and southern states. State fixed effects have been included in all models to capture state specific effects. Fixed effects for the year of birth of the index child is included in all second stage estimations to ensure that the estimated effects are not driven by shocks occurring in particular child birth years. Further, fixed effects for the month of birth have been included to account for seasonal variations in mortality (Brainerd and Menon, 2014), which is of particular significance in an agrarian society. An interaction term for state specific trends has also been included to account for the differential trends of mortality rates within each of the states.

3.5.2 Instrument for Domestic Violence

A number of studies have documented the effect of economic independence on domestic violence, both in developed and developing countries. A majority of them find that greater economic independence of the wife increases her options outside marriage, thereby reducing the risk of domestic violence (Farmer and Tiefenthaler, 1997) (Tauchen et al., 1985). Economic independence of an individual can be enhanced through several mechanisms. While perhaps the most extensively researched mechanism is through the employment of the woman.[2], the examination of effects of alternative variables that enhance the bargaining power of women within the household has been relatively limited.

Several recent studies in the context of India have also shown an effect of unearned income on domestic violence. For instance, Heath and Tan (2014) show that unearned income improves a woman's autonomy within the household and can raise her labour supply. In a similar vein Amaral (2014) find that women who become eligible for inheritance are 17 percentage points less likely to be victims of domestic violence. In contrast, Siwan and Garance (2014) find that increase in property rights for women increases the incidence of wife-beating while Anderson and Genicot (2014) find that it increases suicide rates of both men and women in India.

One such factor that would affect a woman's intra-household bargaining power, which has not been explored in the context of India is dowry. Dowry practices continue to be widespread in spite of being prohibited by law. [3] Historically, dowry was given as a voluntary gift to the bride. The groom or the groom's family had no claims to the dowry even after the death of the bride.[4] (Bühler, 1964). It was anticipated to be an economic safety net for the bride. At present, the value of the dowry given is dependent upon the financial capacity of the bride's family and increases with the positive attributes of the prospective groom (Becker, 1991)(Anderson, 2007). Cash and gold are two of the most prevalent forms of dowry in India in addition to silver, land, car, house etc.

Recent research into dowry in South Asia has led to two distinct theories of dowry motives: bequest as a pre-mortem inheritance and groomprice as a price that clears the marriage market. Although scholars have documented an increasing transformation of dowry from bequest to groomprice (Srinivas, 1984)(Banerjee, 1999), a dowry basket characteristically has elements of both. Research has also found these different regimes of dowry to have heterogeneous effects on women's welfare (Arunachalam and Logan, 2006). Bequest dowries improve the bargaining power of the woman within the household and may thus mitigate domestic violence

[2]Although employment does improve the bargaining power of the woman within the household, research shows divergent effects of women's employment on domestic violence in developing countries Chin (2012)

[3]The Dowry Prohibition Act, 1961.

[4]What (was given) before the (nuptial) fire, what (was given) on the bridal procession, what was given in token of (Such Property), as well as a gift subsequent and what was given (to her) by her affectionate husband, shall go to her offspring (even) if she dies in the lifetime of her husband. The Laws of Manu (c. 200 AD)

against women within the household (Brown, 2009). This has also been cited as the reason why Indian women continue to support dowry practices, despite it being against the law.

Dowry elements with the motive of groomprice are usually a direct transfer of assets to the groom or the groom's family in the form of cash, land, residential property etc. Bequest dowries are less likely to involve cash only transfers as brides have limited control on cash only transfers. In contrast, dowry elements with the motive of bequest are usually a direct transfer of assets to the bride in the form of property or jewellery. India holds 11% of the world's gold stock. 75% of this stock is in the form of jewellery (Grubb, 2015). Gold, one of the primary elements of dowry in India, is almost always given in the form of jewellery to the bride. Studies show that 75% of women in India claimed that their jewellery remained with them after marriage (Basu, 1999). Often a woman has her own locked trunk or a locker at a bank in which she stores her jewels (Hershman, 1981).

I speculate that a high price of gold at the time of marriage reduces the share of gold jewellery in the dowry basket. This reduces the value of assets the bride has direct control over, and in turn exposes her to a higher risk of domestic violence by reducing her bargaining power within the marriage. As a result, the real price of gold at the time of marriage of the mother is used as an instrument in this study. Since there exist no data that observe dowry amounts, I am unable to formally test this hypothesis at the individual level. Nonetheless gold demand statistics in India seem to justify this assumption. Firstly the demand for gold in India is higher during the traditional wedding months from September through January (Wood and Wachman, 2010)(Jadhav, 2015). Thus in spite of wealth accumulation for the purpose of dowry from birth of the daughter, families do purchase gold at the time of marriage with an approximate purchase of 20 to 2000 grams per wedding (Bhandari, 2013)(Afonso, 2014). Secondly, the demand for gold in India is negatively related to the price of gold. This means that at any point in time and even during the marriage driven high demand months of September to January, the amount of gold purchased is lower when the price of gold is high (Ghosal, 2015).

India imports 92% of its gold demand. The price of gold is determined by the London Price Fix twice a day and is external to the country. The national demand for gold in India is then determined through the interplay of this international gold price, share prices (rate of return on alternative financial assets), GDP, the exchange rate and rate of household financial savings (Vaidyanathan, 1999). I use data on the monthly price of gold in Indian Rupees (INR) per troy ounce from the World Gold Council. So although there is no spatial variation in the instrument as the price of gold is set outside India, there is considerable temporal variation for identification with 168 data points (12 months by 14 years). This series is shown in Figure 3.9.1. The price of gold being set outside India also strengthens the exogeneity of this instrument.

Post-independence India had a closed economy characterised by a desire for self-sufficiency. Rigid control of gold sales and taxation led to an extensive black market of gold through smuggling. Gold prices within the country were thus not determined through previously mentioned global market forces, but by local market forces (Vaidyanathan, 1999). In 1991, on the verge of bankruptcy, India's economy changed drastically when it adopted more liberal economic policies. Due to this structural break, the estimation is restricted to mother's who were married post 1991, after which we would expect the world price of gold to be a more accurate measure of the Indian price of gold.

Figure 3.9.2 is a histogram representing the mean of births within each year of marriage in the sample range broken down by whether domestic violence was reported or not weighted by the national domestic violence weight provided with the DHS. There are sufficient numbers of observations within each year to enable identification of the first stage instrumental variable regression.

3.5.3 Exclusion Restriction

The identification rests on the concept of unearned income, much in the same vein as inheritance rights for women. We would expect the price of gold at the time of birth/death of the child to not satisfy the exclusion restriction as the mother could possibly use the gold to buy healthcare services for the child. This is avoided by using the price of gold a time point well before the conception of the child. Thus the assumption being made is that the price of gold effects the economic resources and bargaining power of the woman within the household at the time of marriage.

Given that gold market analysts often fail to accurately predict gold price trends due to the inherent uncertainty in gold price fluctuations, there is no compelling reason to believe that the average individual is able to accurately do so. Consequently, there is no reason to believe that the price of gold at marriage would influence pre-marital investments in prospective brides or grooms in anticipation of future gold prices. On the other hand women may delay marriage if the price of gold is too high. This is examined by confirming that the price of gold is not associated with the age at marriage of the women in Section 7.5.

The price of gold at the time of marriage may affect spousal match quality. I investigate this possibility by regressing the match characteristics on several key variables between the bride and groom on the price of gold and report the results in Section 7.6.

3.6 Results

Linear probability estimations have a small but insignificant effect on child mortality as shown in Table 3.9.7 . These estimates are conditional on the full set of controls used in the second stage of the instrumental variable specification and include state fixed effects, year of birth fixed effects, month of birth fixed effects, state fixed time trends and birth order fixed effects.

Next, I present a discussion of the instrumental variable first stage statistics followed by the second stage estimation results.

Table 3.9.8 presents the first stage statistics of the two-stage least square estimation. The independent variable of interest is the monthly price of gold (in 1000's). Column 1 presents the effect of the instrument on domestic violence conditional on mother's characteristics only. Column 2 additionally conditions on father's age and education. Column 3 additionally conditions on state fixed effects. I find a positive and significant effect in the first stage in all three specifications. The columns also presents the Kleibergen-Paap Wald test robust to heteroskedasticity for the strength of the instrument in the first stage. Price of gold exceeds the Stock and Yogo critical value with a test statistic of 19.74 in Column 3 which is the full speciifcation used for the second stage results presented in Table 3.9.9.

Column 1 and Column 2 in Table 3.9.9 present the second stage results of the two-stage least squares estimation for neonatal and infant mortality respectively. Domestic violence is a significant predictor of child mortality in both stages of the child's life. In the first 30 days a one standard deviation increase in domestic violence increases the likelihood of both neonatal mortality and infant mortality by 17.6 and 16.9 percentage points relative to the mean respectively.

The effects of socio-demographic controls largely follow the existing literature on the subject. Previous results (Oster, 2009) find that although mortality is high from 0 to 6 months, it does not explain excess female mortality. Similarly, in these results, girls have lower risk of mortality in the neonatal model. A possible explanation is that since sex-selective abortion is rampant in India, the girls that are eventually born are only desired births (Bhalotra and Cochrane, 2010). Households located in rural areas have a higher the risk of death in both models. This is congruent with the well documented lack of adequate health and post-natal care available for the inhabitants of rural India. Muslim children have a slight advantage in the infant mortality model as has been established previously (Bhalotra et al. 2010). Christian children also have an advantage in both neonatal and infant mortality models. Births to scheduled castes relative to high caste have a significantly lower likelihood of mortality.

3.7 Robustness Checks

In this section I test the robustness of the estimates based on alternative specifications of domestic violence and by exploring alternative mechanisms.

3.7.1 Alternative Measures of Violence

Table 3.9.10 presents the first stage results for violence coded as a simple factor score. This method does not account for the ordinal categories in the 7 components of domestic violence. Table 3.9.11 presents the second stage results of the effect of domestic violence coded as a factor score on neonatal and infant mortality. I find that not accounting for the ordering of the domestic violence responses leads to higher point estimates of 0.39 percentage points and 0.38 percentage points for neonatal and infant models respectively. Coding violence as a factor score also leads to higher standard errors and lowers the strength of the instrument in the first stage with a Kleibergen-Paap Wald test statistic of 15.78.

Table 3.9.12 report the first stage statistics for estimations when domestic violence has been coded as a cumulative ordinal index with 0 indicating that the respondent did not report any form of physical violence. Coding violence as a cumulative index also reduces the strength of the instrument in the first stage with a lower F-statistic at 18.58 as compared to the baseline model. Table 3.9.13 presents the second stage results of the effect of the fitted values of domestic violence on neonatal and infant mortality. The coefficients are consistent with the baseline instrumental variable estimation with higher standard errors and a higher magnitude of 0.32 percentage points increase in risk of mortality in both models.

3.7.2 Alternative Types of Violence

The NFHS 3 also includes 2 questions regarding emotional and sexual violence. Respondents were asked if they ever experienced any sexual or emotional violence with two responses allowed for yes and no. The last 2 panels in Table 3.9.1 shows the responses to these questions. I created a polychoric principal component analysis including these two variables to check if the results remained consistent. The first principle component accounts for 75% of the overall variability of domestic violence which includes physical, sexual and emotional violence. Table 3.9.14 presents the first stage results of the instrumental variable estimation when using domestic violence inclusive of emotional and sexual violence. The Kleibergen-Paap Wald test statistic is lower at 17.96 than when using only the physical components. Table 3.9.15 presents the second stage results of coding violence as inclusive of sexual and emotional violence. The point estimates have higher standard errors and higher magnitudes at 0.47 and 0.48 percentage points for neonatal and infant models respectively.

3.7.3 Age at Marriage

A possible concern with using the price of gold at the time of marriage as an instrument for domestic violence post marriage could be that the price of gold may affect the age of the mother at which the marriage occurs. It is conceivable that a higher price of gold at the time of marriage induces families to postpone the marriage in anticipation of a future drop in the price. I test this relationship in our analytical sample and find no significant association between the price of gold and the age at marriage. (Table 3.9.16)

3.7.4 Marital Matching

Since dowry is the price at which the marriage market clears in India, at equilibrium, the amount of dowry may have an effect on groom characteristics. Since I only expect the amount of gold in the dowry basket to be affected and not the value of the dowry basket as a whole (which would almost definitely have a cash component if not other physical asset components), I do not expect the exogenously set price of gold to have an effect on groom or match characteristics.

However, to address this possibility, I test this in the data by looking at the association between the price of gold and key spousal match characteristics such as differences in age and education between the bride and groom in Table 3.9.17. Column 1 presents the difference in age between the husband and wife, column 2 presents the difference in education (measured in single years). I do not find any significant effects of the price of gold at the time of marriage on the difference in age or the difference on education between the mother and the father.

3.7.5 Mother's financial autonomy

Unearned income may affect other measures of bargaining outcomes. One such measure that is of relevance to the identification strategy is that an increase in the price of gold at the time of marriage could decrease the woman's financial decision making power within the household and thereby affect the fraction of household income that is allocated to health and education expenditure (Duflo, 2003). All the estimations in this paper are conditional on household wealth. Nonetheless, I test for this possibility by using a variable in the dataset that elicits the self-reported extent of the financial decision making authority of the mother. The respondents were asked *"Who usually decides how the money you earn will be used: mainly you, mainly your husband/partner, or you and your husband/partner jointly?"* The response options given to them were *respondent alone*, *respondent and husband*, *husband alone* or *someone else*. I coded this as a binary indicator to be 1 if the respondent chose the *respondent alone* option and to be 0 otherwise. I then test the association between this indicator and the price of gold at the time of marriage and report the results in Table 3.9.18. I do not find a significant association between the price of gold at the time of marriage and wife's financial decision making autonomy.

This suggests that the amount of gold in the dowry basket does not change the dynamics of household financial decision making authority but may be acting as a credible threat point for the wife to leave the marriage.

3.7.6 Restricting births to children born in the last few years

The data used in this analyses was collected in the years 2005 and 2006 from mothers who were between 15 and 49 years of age. The births used in the analysis are retrospective data of all births post the year 1991. As one goes back in time, the children born to younger mothers may become an unrepresentative sample and may bias the estimated effects. (Arulampalam and Bhalotra, 2006). I test for biases in the estimates by restricting the sample to children born in the 8 years, 6 years and 4 years preceding the survey. Columns 1 and 2 in Table 3.9.19 presents the estimates for children born in the 8 years preceding the survey for the neonatal and infant mortality models. Although less precisely estimated, I continue to find a positive effect of domestic violence on neonatal and infant mortality in all specifications.

3.8 Conclusion

This study constitutes a significant first step towards establishing a causal link between domestic violence and infant mortality. I find a significant positive relationship between domestic violence and both neonatal and infant mortality. I circumvent the problem of endogeneity by using the real price of gold as a source of exogenous variation in domestic violence. The results remained consistent through alternative measures of domestic violence and through several robustness tests. The instrument for domestic violence, although specific to India could potentially be used to estimate the causal effects of other externalities of domestic violence not explored in this paper, such as women's labour market outcomes and other birth outcomes.

The analytical sample is affected by sample selection as I am unable to include foetal deaths. So the births in the sample are foetuses that came to full term and are therefore likely to be stronger foetuses. Thus the magnitude of effects reported in this paper are likely to be a lower bound of actual effect sizes.

Since the timing of the violence is not available in the data, I am unable to ascertain whether the effects we estimate are the results of violence specifically during pregnancy or of the long-term existence of violence in the marriage. This could be enhanced by more extensive data on the timing of violence and the cash values of various kinds of dowries. Given the prevalence of both domestic violence and dowry practices in India, there is an inherent need for this data. However, the illegality of dowry and domestic violence and subsequent underreporting of each could make further accurate data collection difficult and must be addressed methodologically for precision in future analysis, for example by changing key survey parameters to overcome underreporting and systematic measurement errors.

Concerted policy initiatives directed at the identification and eradication of domestic violence can effectively reduce neonatal and infant mortality levels in India. This could set helpful examples for developing countries where public health funding dedicated to the lowering of child mortality is frequently limited. Public policy addressing key aspects of improving absolute levels of gender equality tend to be relatively inexpensive and, if incorporating mechanics aimed at the reduction of domestic violence, should induce a reduction in child mortality.

3.9 Figures and Tables

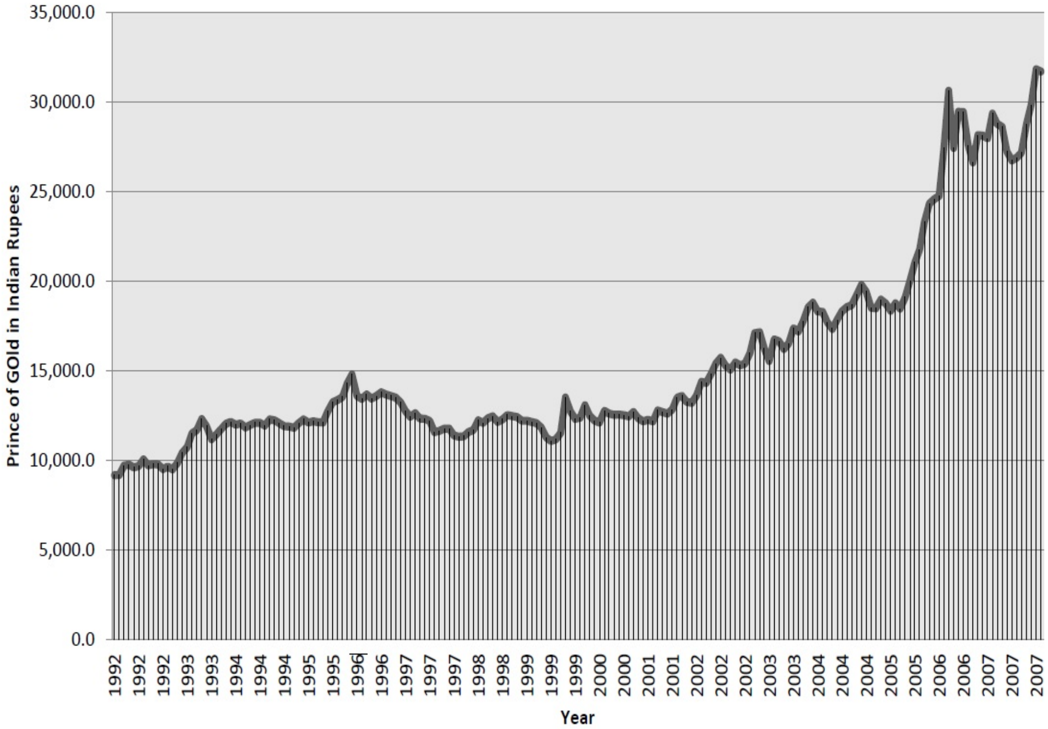

Figure 3.9.1: Time series of the price of gold

Notes: This figure plots the time series of the price of gold in Indian Rupees over the year of marriage for the analytical sample. Price of gold is obtained from the World Gold Council and is matched to the month and year of marriage of the mother

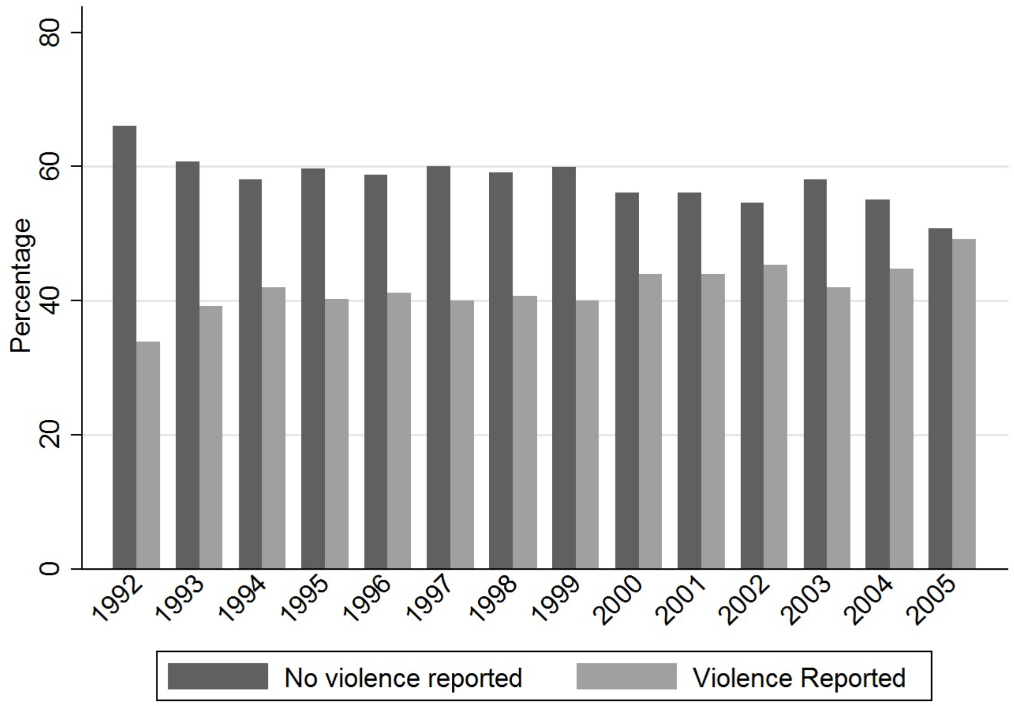

Figure 3.9.2: Percentage of domestic violence by year of marriage

Notes: This graph shows the percentage of mothers who reported domestic violence versus those mothers that did not by the year of marriage.

Table 3.9.1: Responses to the components of Domestic Violence

Domestic Violence Questions			
spouse ever slapped	**Freq.**	**Percent**	**Cum.**
No	16,355	76.11	76.11
Yes, but timing missing	18	0.08	76.2
Not during the last 12 months	1,501	6.99	83.18
Sometimes during the last 12 months	3,091	14.38	97.57
Often during the last 12 months	523	2.43	100
Total	21,488	100	
spouse ever twisted her arm or pull her hair	**Freq.**	**Percent**	**Cum.**
No	19,560	91.03	91.03
Yes, but timing missing	9	0.04	91.07
Not during the last 12 months	491	2.28	93.35
Sometimes during the last 12 months	1,169	5.44	98.79
Often during the last 12 months	259	1.21	100
Total	21,488	100	
spouse ever pushed, shook or threw something	**Freq.**	**Percent**	**Cum.**
No	19,743	91.88	91.88
Yes, but timing missing	9	0.04	91.92
Not during the last 12 months	427	1.99	93.91
Sometimes during the last 12 months	1,088	5.06	98.97
Often during the last 12 months	221	1.03	100
Total	21,488	100	
spouse ever punched with fist or something harmful	**Freq.**	**Percent**	**Cum.**
No	20,270	94.33	94.33
Yes, but timing missing	4	0.02	94.35
Not during the last 12 months	286	1.33	95.68
Sometimes during the last 12 months	738	3.43	99.12
Often during the last 12 months	190	0.88	100
Total	21,488	100	
spouse ever kicked or dragged	**Freq.**	**Percent**	**Cum.**
No	20,085	93.47	93.47
Yes, but timing missing	4	0.02	93.49
Not during the last 12 months	351	1.63	95.12
Sometimes during the last 12 months	841	3.91	99.04
Oftern during the last 12 months	207	0.96	100
Total	21,488	100	
spouse ever threatened or attacked with knife/gun or other	**Freq.**	**Percent**	**Cum.**
No	21,327	99.25	99.25
Yes, but timing missing	3	0.01	99.26
Not during the last 12 months	32	0.15	99.41
Sometimes during the last 12 months	88	0.41	99.82
Often during the last 12 months	38	0.18	100
Total	21,488	100	
spouse ever tried to strangle or burn	**Freq.**	**Percent**	**Cum.**
No	21,249	98.89	98.89
Yes, but timing missing	3	0.01	98.9
Not during the last 12 months	42	0.2	99.1
Sometimes during the last 12 months	134	0.62	99.72
Often during the last 12 months	60	0.28	100
Total	21,488	100	
Ever experienced any emotional violence	**Freq.**	**Percent**	**Cum.**
No	19,134	89.05	89.05
Yes	2,354	10.95	100
Total	21,488	100	
Ever experienced any sexual violence	**Freq.**	**Percent**	**Cum.**
No	20,068	93.39	93.39
Yes	1,420	6.61	100
Total	21,488	100	

This table presents the responses of the analytical sample of mothers to each of the domestic violence components. The first 7 panels present the statistics for physical components of violence while the last 2 panels present the statistics of emotional and sexual violence

Table 3.9.2: Cumulative Ordinal Index of violence

Cumulative ordinal index of violence	**Freq.**	**Percent**	**Cum.**
No Violence	16,101	74.93	74.93
spouse ever slapped	2,589	12.05	86.98
spouse ever twisted her arm or pull her	1,068	4.97	91.95
spouse ever pushed, shook or threw some	701	3.26	95.21
spouse ever punched with fist or something else	412	1.92	97.13
spouse ever kicked or dragged	413	1.92	99.05
spouse ever threatened or attacked with	142	0.66	99.71
spouse ever tried to strangle or burn	62	0.29	100
Total	21,488	100	

The 7 physical components of violence presented above are sequentially added to create a cumulative ordinal index of domestic violence.

Table 3.9.3: Summary Statistics - Mother level

Variable	Obs	Mean	Std. Dev.
Mother's years of schooling	21488	7.51	5.04
Mother's age - st time of survey	21488	27.21	5.1
Mother's height	21488	15.25	.59
Husband's age in years	21488	32.62	6
Husband's years of schooling	21488	8.74	4.89
Location of household - Urban	21488	.5	.5
Location of household - Rural	21488	.5	.5
Religion - Hindu	21488	.74	.44
Religion - Muslim	21488	.09	.29
Religion - Christian	21488	.11	.31
Religion - Other Religions	21488	.06	.23
Wealth Category - Poorest	21488	.08	.26
Wealth Category - Poorer	21488	.12	.33
Wealth Category - Middle	21488	.19	.39
Wealth Category - Richer	21488	.27	.44
Wealth Category - Richest	21488	.34	.47
Caste - Schedule Caste	21488	.16	.37
Caste - Schedule Tribe	21488	.15	.35
Caste - Other Backward Caste	21488	.31	.46
Caste - High Caste	21488	.38	.49

This table presents the summary statistics for the variables that are measured at the mother level and are used in both stages of the instrumental variable estimations.

Table 3.9.4: Summary Statistics - Child level

Variable	Obs	Mean	Std. Dev.
Gender - Boy	43736	.52	.5
Gender - Girl	43736	.48	.5
Birth order number	43736	1.77	.96
Mother's age at birth	43736	22.87	4.28

This table presents the summary statistics of the variables that are meaured at the child level and are included in the second stage of the instrumental variable estimations.

Table 3.9.5: Summary Statistics of Neonatal & Infant Mortality

Variable	N	Mean	Standard Deviation
Neonatal Mortality	43,736	0.031	0.173
Infant Mortality	39,470	0.037	0.189

Both neonatal and infant mortality are binary variables with 1 indicating that the child is dead.

Table 3.9.6: Difference in mortality by domestic violence

	Domestic Violence	**N**	**mean**	**sd**	**t-test***
Neonatal Mortality	Violence not reported	31659	0.029	0.166	
	Violence reported	12077	0.037	0.188	
	Total	43736	0.031	0.173	0.00
Infant Mortality	Violence not reported	28369	0.034	0.181	
	Violence reported	11101	0.044	0.206	
	Total	39470	0.037	0.189	0.00

Both neonatal and infant mortality are binary variables with 1 indicating that the child is dead.
Violence not reported indicates that the mother has not reported the presence of any of the 7 physical components of domestic violence.
Violence reported indicated that the mother has reported at least one form of the 7 components of physical violence.
*p-values of unpaired t-test for equality mortality between violence reported and violence not reported.

Table 3.9.7: Linear Regression Results

	(1) Neonatal Mortality	(2) Infant Mortality
Domestic Violence	0.001	0.002
	(0.003)	(0.003)
Boy		
Girl	-0.006**	-0.003
	(0.003)	(0.003)
Mother's age at birth	0.005	0.005
	(0.003)	(0.003)
Mothers height <14m		
Mothers height 14m-15m	-0.004	-0.008**
	(0.004)	(0.004)
Mothers height >15m	-0.011**	-0.011**
	(0.003)	(0.004)
Poorest		
Poorer	0.004	-0.011
	(0.006)	(0.007)
Middle	-0.000	-0.007
	(0.006)	(0.007)
Richer	-0.009	-0.019**
	(0.006)	(0.007)
Richest	-0.013**	-0.025**
	(0.007)	(0.008)
Urban		
Rural	0.007**	0.003
	(0.003)	(0.004)
Hindu		
Muslim	-0.002	-0.007
	(0.005)	(0.005)
Christian	-0.011**	-0.011**
	(0.005)	(0.006)
Other Religions	-0.010**	-0.013*
	(0.005)	(0.007)
High Caste		
Scheduled caste	-0.007	-0.005
	(0.004)	(0.005)
Scheduled tribe	-0.005	0.006
	(0.006)	(0.007)
Other backward class	0.001	0.003
	(0.004)	(0.004)
Husband's age in years	-0.001	0.000
	(0.000)	(0.000)
Year of birth FE	✓	✓
State FE	✓	✓
Month of birth FE	✓	✓
State FE*Year	✓	✓
N	43736.000	39470.000
Mean (dep Var)	0.031	0.037

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Domestic violence is the first component of a polychoric principle component analysis using the 7 physical components presented in Table 3.9.1. Neonatal and Infant mortality are binary variables with 0 indicating that the child is alive and 1 indicating that the child is dead. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Table 3.9.8: Instrumental Variable - 1^{st} stage results

	(1) Domestic Violence	(2) Domestic Violence	(3) Domestic Violence
Price of Gold	0.003*	0.007***	0.007***
	(0.001)	(0.002)	(0.002)
Mother's Height <14m			
Mothers height 14m-15m	-0.018**	-0.019**	-0.018**
	(0.009)	(0.009)	(0.009)
Mothers height >15m	-0.028***	-0.028***	-0.026**
	(0.008)	(0.008)	(0.008)
Poorest			
Poorer	-0.049**	-0.048**	-0.028
	(0.021)	(0.022)	(0.022)
Middle	-0.110***	-0.098***	-0.073***
	(0.020)	(0.020)	(0.021)
Richer	-0.168***	-0.153***	-0.117***
	(0.020)	(0.020)	(0.021)
Richest	-0.244***	-0.226***	-0.178***
	(0.021)	(0.021)	(0.022)
Urban			
Rural	-0.074***	-0.066***	-0.050***
	(0.008)	(0.008)	(0.009)
Hindu			
Muslim	0.017	0.014	0.025**
	(0.013)	(0.012)	(0.013)
Christian	-0.057***	-0.050***	-0.000
	(0.013)	(0.013)	(0.015)
Other Religions	0.003	0.005	0.014
	(0.015)	(0.015)	(0.017)
High Caste			
Scheduled caste	0.090***	0.091***	0.084***
	(0.012)	(0.012)	(0.012)
Scheduled tribe	-0.005	-0.005	0.015
	(0.014)	(0.014)	(0.015)
Other backward class	0.023**	0.023**	0.002
	(0.008)	(0.008)	(0.009)
Husband's age in years		-0.001	-0.001
		(0.001)	(0.001)
N	21488	21488	21488
Kleibergen-Paap Wald test	72.52	18.33	19.74

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Domestic violence is the first component of a polychoric principle component analysis using the 7 physical components presented in Table 3.9.1.
Column 2 additionally conditions on Husband's years of schooling.
Column 3 additionally includes state fixed effects.

Table 3.9.9: Instumental Variable - 2^{nd} stage results

	(1) Neonatal Mortality	(2) Infant Mortality
Fitted values	0.176***	0.169**
	(0.053)	(0.065)
Boy		
Girl	-0.006**	-0.003
	(0.003)	(0.003)
Mother's age at birth	0.005	0.006
	(0.003)	(0.004)
Mother's Height $<$ 14m		
Mothers height 14m-15m	0.001	-0.004
	(0.004)	(0.004)
Mothers height $>$15m	-0.003	-0.004
	(0.004)	(0.005)
Poorest		
Poorer	0.004	-0.010
	(0.006)	(0.007)
Middle	0.010	0.003
	(0.007)	(0.008)
Richer	0.008	-0.002
	(0.008)	(0.010)
Richest	0.014	0.002
	(0.011)	(0.013)
Rural	0.015***	0.011**
	(0.004)	(0.004)
Hindu		
Muslim	-0.006	-0.011**
	(0.005)	(0.005)
Christian	-0.015**	-0.016**
	(0.005)	(0.006)
Other Religions	-0.013**	-0.016**
	(0.005)	(0.006)
High Caste		
Scheduled caste	-0.025***	-0.023**
	(0.007)	(0.008)
Scheduled tribe	-0.012*	-0.001
	(0.006)	(0.007)
Other backward class	-0.002	0.001
	(0.004)	(0.004)
Husband's age in years	-0.000	0.000
	(0.000)	(0.000)
Year of birth FE	✓	✓
State FE	✓	✓
Month of birth FE	✓	✓
State FE*Year	✓	✓
N	43736.000	39470.000
Mean (dep Var)	0.031	0.037

Robust standard errors clustered at month and year of marriage in parenthesis $^{***}p < 0.01$, $^{**}\ p < 0.05$, $^{*}\ p < 0.1$. Domestic violence is the first component of a polychoric principle component analysis using the 7 physical components presented in Table 3.9.1. Neonatal and Infant mortality are binary variables with 0 indicating that the child is alive and 1 indicating that the child is dead. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Table 3.9.10: Instrumental Variable - 1^{st} stage results with domestic violence as a factor score

	(1) Scores for factor 1
Price of Gold	0.010***
	(0.003)
Mother's Height < 14m	
Mothers height 14m-15m	-0.035**
	(0.014)
Mothers height >15m	-0.041**
	(0.014)
Poorest	
Poorer	-0.042
	(0.035)
Middle	-0.106**
	(0.034)
Richer	-0.176***
	(0.034)
Richest	-0.269***
	(0.036)
Urban	
Rural	-0.082***
	(0.014)
Hindu	
Muslim	0.037*
	(0.021)
Christian	0.010
	(0.025)
Other Religions	0.016
	(0.028)
High Caste	
Scheduled caste	0.123***
	(0.019)
Scheduled tribe	0.021
	(0.025)
Other backward class	0.006
	(0.014)
Husband's age in years	-0.003
	(0.002)
N	21488.000
Kleibergen-Paap Wald test	15.78

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Domestic violence is the first factor of a factor score using the 7 physical components presented in Table 3.9.1. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and state fixed effects.

Table 3.9.11: Instrumental Variable - 2^{nd} stage results with domestic violence as a factor score

	(1) Neonatal Mortality	(2) Infant Mortality
Domestic Violence	0.397**	0.388**
	(0.134)	(0.120)
Boy		
Girl	-0.007	-0.004
	(0.007)	(0.007)
Mother's age at birth	0.008	0.009
	(0.009)	(0.009)
Mother's Height < 14m		
Mothers height 14m-15m	0.015	0.010
	(0.011)	(0.011)
Mothers height >15m	0.017	0.016
	(0.013)	(0.014)
Poorest		
Poorer	0.004	-0.010
	(0.018)	(0.019)
Middle	0.036*	0.028
	(0.021)	(0.021)
Richer	0.054**	0.043*
	(0.027)	(0.026)
Richest	0.083**	0.070**
	(0.037)	(0.034)
Urban		
Rural	0.034**	0.030**
	(0.013)	(0.013)
Hindu		
Muslim	-0.013	-0.018
	(0.015)	(0.014)
Christian	-0.021	-0.021
	(0.016)	(0.016)
Other Religions	-0.023	-0.025
	(0.016)	(0.018)
High Caste		
Scheduled caste	-0.069**	-0.066**
	(0.024)	(0.023)
Scheduled tribe	-0.023	-0.012
	(0.018)	(0.019)
Other backward class	-0.007	-0.005
	(0.010)	(0.010)
Husband's age in years	0.000	0.001
	(0.001)	(0.001)
Year of birth FE	✓	✓
State FE	✓	✓
Month of birth FE	✓	✓
State FE*Year	✓	✓
N	43736.000	39470.000
Mean (dep Var)	0.031	0.037

Robust standard errors clustered at month and year of marriage in parenthesis $***p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. Domestic violence is the first factor of a factor score using the 7 physical components presented in Table 3.9.1. Neonatal and Infant mortality are binary variables with 0 indicating that the child is alive and 1 indicating that the child is dead. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Table 3.9.12: Instrumental Variable - 1^{st} stage results with domestic violence as a cumulative ordinal index

	(1) Cumulative ordinal index of violence
Price of Gold	0.013***
	(0.003)
Mother's Height $<$ 14m	
Mothers height 14m-15m	-0.032*
	(0.017)
Mothers height $>$15m	-0.045**
	(0.017)
Poorest	
Poorer	-0.101**
	(0.043)
Middle	-0.195***
	(0.041)
Richer	-0.297***
	(0.041)
Richest	-0.465***
	(0.043)
Urban	
Rural	-0.141***
	(0.017)
Hindu	
Muslim	0.031
	(0.024)
Christian	0.010
	(0.033)
Other Religions	0.044
	(0.034)
High Caste	
Scheduled caste	0.153***
	(0.023)
Scheduled tribe	0.025
	(0.031)
Other backward class	-0.011
	(0.017)
Husband's age in years	-0.001
	(0.002)
N	21488.000
Kleibergen-Paap rk Wald test	18.58

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Domestic violence is a cumulative ordinal index randing from 0 to 7 using the 7 physical components presented in Table 3.9.1. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and state fixed effects.

Table 3.9.13: Instrumental Variable - 2^{nd} stage results with domestic violence as a cumulative ordinal index

	(1) Neonatal Mortality	(2) Infant Mortality
Domestic Violence	0.323**	0.315***
	(0.098)	(0.092)
Boy		
Girl	-0.007	-0.005
	(0.008)	(0.009)
Mother's age at birth	0.014	0.015
	(0.011)	(0.011)
Mother's Height < 14m		
Mothers height 14m-15m	0.012	0.007
	(0.012)	(0.012)
Mothers height >15m	0.014	0.013
	(0.013)	(0.014)
Poorest		
Poorer	0.031	0.016
	(0.022)	(0.023)
Middle	0.064**	0.056**
	(0.028)	(0.027)
Richer	0.092**	0.081**
	(0.036)	(0.035)
Richest	0.137**	0.123**
	(0.050)	(0.047)
Urban		
Rural	0.047**	0.043**
	(0.017)	(0.016)
Hindu		
Muslim	-0.015	-0.020
	(0.016)	(0.016)
Christian	-0.027	-0.028
	(0.019)	(0.018)
Other Religions	-0.031*	-0.033
	(0.018)	(0.021)
High Caste		
Scheduled caste	-0.068**	-0.065**
	(0.023)	(0.022)
Scheduled tribe	-0.026	-0.015
	(0.021)	(0.021)
Other backward class	-0.000	0.002
	(0.011)	(0.011)
Husband's age in years	-0.001	-0.000
	(0.001)	(0.001)
Year of birth FE	✓	✓
State FE	✓	✓
Month of birth FE	✓	✓
State FE*Year	✓	✓
N	43736.000	39470.000
Mean (dep Var)	0.031	0.037

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Domestic violence as a cumulative ordinal index as shown in Table . Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Table 3.9.14: Instrumental Variable - 1^{st} stage results with domestic violence as an index inclusive of emotional and sexual violence

	(1) Domestic Violence
Indianrupee	0.008***
	(0.002)
Mother's Height <14m	
Mothers height 14m-15m	-0.023**
	(0.010)
Mothers height >15m	-0.027**
	(0.010)
Poorest	
Poorer	-0.030
	(0.025)
Middle	-0.072**
	(0.024)
Richer	-0.127***
	(0.024)
Richest	-0.199***
	(0.026)
Urban	
Rural	-0.057***
	(0.010)
Hindu	
Muslim	0.039**
	(0.015)
Christian	0.008
	(0.019)
Other Religions	0.001
	(0.020)
High Caste	
Scheduled caste	0.095***
	(0.014)
Scheduled tribe	0.030*
	(0.018)
Other backward class	0.005
	(0.010)
Husband's age in years	-0.001
	(0.001)
N	21488.000
Kleibergen-Paap Wald test	17.96

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Domestic violence is the first component of a polychoric principle component analysis using all the 9 components presented in Table 3.9.1. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and state fixed effects.

Table 3.9.15: Instrumental Variable - 2^{nd} stage results with domestic violence as an index inclusive of emotional and sexual violence

	(1) Neonatal Mortality	(2) Infant Mortality
Domestic Violence	0.472**	0.476***
	(0.161)	(0.144)
Boy		
Girl	-0.006	-0.003
	(0.006)	(0.007)
Mother's age at birth	0.006	0.007
	(0.008)	(0.008)
Mother's Height $<$ 14m		
Mothers height 14m-15m	0.010	0.006
	(0.009)	(0.010)
Mothers height $>$15m	0.011	0.010
	(0.011)	(0.012)
Poorest		
Poorer	0.005	-0.009
	(0.016)	(0.017)
Middle	0.029*	0.023
	(0.018)	(0.019)
Richer	0.044*	0.035
	(0.023)	(0.022)
Richest	0.073**	0.062**
	(0.033)	(0.030)
Urban		
Rural	0.030**	0.026**
	(0.011)	(0.011)
Hindu		
Muslim	-0.013	-0.018
	(0.014)	(0.013)
Christian	-0.023	-0.023
	(0.015)	(0.015)
Other Religions	-0.022	-0.025
	(0.014)	(0.017)
High Caste		
Scheduled caste	-0.061**	-0.060**
	(0.021)	(0.020)
Scheduled tribe	-0.030*	-0.020
	(0.017)	(0.018)
Other backward class	-0.005	-0.002
	(0.009)	(0.009)
Husband's age in years	-0.000	0.000
	(0.001)	(0.001)
Year of birth FE	✓	✓
State FE	✓	✓
Month of birth FE	✓	✓
State FE*Year	✓	✓
N	43736.000	39470.000
Mean (dep Var)	0.031	0.037

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Domestic violence is the first component of a polychoric principle component analysis and includes physical, emotional and sexual violence using all the 9 components presented in Table 3.9.1. Neonatal and Infant mortality are binary variables with 0 indicating that the child is alive and 1 indicating that the child is dead. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Table 3.9.16: Effect of price of gold on age at marriage of the mother

	(1) Mother'a age at time of marriage b/se
Price of Gold	0.000
	(0.001)
Mother's Height $<$ 14m	
Mothers height 14m-15m	0.016**
	(0.007)
Mothers height $>$15m	0.011
	(0.007)
Poorest	
Poorer	0.006
	(0.013)
Middle	-0.001
	(0.013)
Richer	0.011
	(0.013)
Richest	0.014
	(0.015)
Urban	
Rural	0.017**
	(0.007)
Hindu	
Muslim	-0.001
	(0.010)
Christian	-0.005
	(0.014)
Other Religions	-0.008
	(0.014)
High Caste	
Scheduled caste	0.007
	(0.009)
Scheduled tribe	-0.005
	(0.012)
Other backward class	-0.009
	(0.007)
Husband's age in years	0.002**
	(0.001)
N	21488.000

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Mother's age at marriage is the age of the mother in years at the time of marriage. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and state fixed effects.

Table 3.9.17: Effect of price of gold on marital match characteristics

	(1) Difference in Age	(2) Difference in Education
Price of Gold	-0.012	-0.013
	(0.013)	(0.012)
Mother's Height $<$ 14m		
Mothers height 14m-15m	0.077	0.125*
	(0.066)	(0.065)
Mothers height $>$15m	0.163**	0.161**
	(0.068)	(0.066)
Poorest		
Poorer	-0.061	0.432***
	(0.142)	(0.128)
Middle	-0.046	0.655***
	(0.138)	(0.121)
Richer	0.053	0.630***
	(0.141)	(0.120)
Richest	0.109	0.365**
	(0.153)	(0.127)
Urban		
Rural	0.032	0.412***
	(0.066)	(0.066)
Hindu		
Muslim	0.398***	-0.492***
	(0.098)	(0.100)
Christian	-0.795***	-0.351**
	(0.150)	(0.123)
Other Religions	-0.168	-0.330**
	(0.138)	(0.129)
High Caste		
Scheduled caste	0.091	0.413***
	(0.081)	(0.086)
Scheduled tribe	-0.279**	0.665***
	(0.119)	(0.111)
Other backward class	0.049	0.291***
	(0.067)	(0.071)
N	21488.000	21488.000

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees.
Dependent variable in Column (1) is the difference in age in years between the Mother and the Father. Column (1) additionally controls for Mother's years of schooling, Father's years of schooling and state fixed effects.
Dependent variable in Column (2) is the difference in years of schooling between the Mother and the Father. Column (1) additionally controls for Mother's age at time of survey and state fixed effects.

Table 3.9.18: Effect of price of gold on Mother's Financial Autonomy

	(1) Mother's Financial Autonomy
Price of Gold	-0.002
	(0.003)
Mother's Height <14m	
Mothers height 14m-15m	0.026*
	(0.015)
Mothers height >15m	-0.016
	(0.015)
Poorest	
Poorer	0.002
	(0.023)
Middle	0.040*
	(0.024)
Richer	0.046*
	(0.027)
Richest	0.111***
	(0.033)
Urban	
Rural	-0.059***
	(0.015)
Hindu	
Muslim	0.133***
	(0.031)
Christian	0.009
	(0.027)
Other Religions	-0.005
	(0.032)
High Caste	
Scheduled caste	-0.007
	(0.020)
Scheduled tribe	0.004
	(0.025)
Other backward class	-0.022
	(0.018)
Husband's age in years	0.001
	(0.002)
N	10654.000

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Price of gold in 1000 Indian Rupees. Mother's financial autonomy is a binary variable with 1 indicating that the Mother usually decides how to spend the money earned and 0 otherwise. Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and state fixed effects.

Table 3.9.19: 2^{nd} stage results - Restricting births to children born in the last 8 years to 4 years

	(1) 8 years Neonatal Mortality	(2) 8 years Infant Mortality	(3) 6 years Neonatal Mortality	(4) 6 years Infant Mortality	(5) 4 years Neonatal Mortality	(6) 4 years Infant Mortality
Domestic Violence	0.168**	0.148*	0.174**	0.187**	0.215**	0.310**
	(0.061)	(0.077)	(0.071)	(0.095)	(0.089)	(0.144)
Boy						
Girl	-0.005*	-0.000	-0.004	0.002	-0.010**	-0.003
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.007)
Mother's age at birth	0.007*	0.009**	0.008*	0.010**	0.006	0.006
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)
Poorest						
Poorer	0.004	-0.013	0.017**	0.003	0.023**	-0.003
	(0.007)	(0.008)	(0.007)	(0.009)	(0.010)	(0.015)
Middle	0.010	0.002	0.017**	0.010	0.023**	0.016
	(0.007)	(0.009)	(0.008)	(0.011)	(0.010)	(0.017)
Richer	0.006	-0.005	0.013	0.007	0.020	0.018
	(0.009)	(0.011)	(0.010)	(0.013)	(0.012)	(0.021)
Richest	0.013	-0.003	0.022	0.013	0.033*	0.032
	(0.012)	(0.015)	(0.014)	(0.018)	(0.017)	(0.028)
Urban						
Rural	0.013**	0.010*	0.015**	0.012*	0.014**	0.014
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.010)
Hindu						
Muslim	-0.007	-0.011*	-0.007	-0.011	-0.006	-0.001
	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)	(0.013)
Christian	-0.007	-0.010	-0.009	-0.014*	-0.006	-0.011
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.012)
Other Religions	-0.009	-0.019**	-0.005	-0.014	-0.003	-0.017
	(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(0.015)
High Caste						
Scheduled caste	-0.029***	-0.025**	-0.028**	-0.025**	-0.036**	-0.041**
	(0.008)	(0.009)	(0.009)	(0.012)	(0.012)	(0.018)
Scheduled tribe	-0.014**	-0.006	-0.020**	-0.017*	-0.021**	-0.019
	(0.007)	(0.008)	(0.008)	(0.010)	(0.010)	(0.015)
Other backward class	-0.008*	-0.004	-0.007	-0.006	-0.008	-0.010
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.009)
Husband's age in years	-0.001	0.000	-0.001	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Year of birth FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Month of birth FE	✓	✓	✓	✓	✓	✓
State FE*Year	✓	✓	✓	✓	✓	✓
N	30909.000	26580.000	23089.000	18716.000	14744.000	10313.000
Mean (dep var)	0.029	0.035	0.027	0.034	0.028	0.033

Robust standard errors clustered at month and year of marriage in parenthesis ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Domestic violence is the first component of a polychoric principle component analysis of the 7 physical violence components presented in Table 3.9.1 . Additional controls: Mother's years of schooling, Father's years of schooling, Mother's age at time of survey and birth order fixed effects.

Chapter 4

Conclusion

This thesis examined two aspects of demographic change in the form of survival expectations among the elderly and child mortality. Demographic change in the form of an increasing elderly dependency ratio plays a crucial role in national policy choices especially regarding pensions and healthcare. At the current growth rate of the ageing populations vis-á-vis youth populations, it is likely that pension and social care systems in both developed and developing countries will be at risk.

Recent research into the field of subjective survival expectations has shown that they hold valuable information about individualistic beliefs and that they covary in expected ways with socio-economic predictors such as income, age, wealth and occupational status. Eliciting survival expectations in a developing country context involves several methodological considerations. For instance these populations typically have lower numeracy skills and are more likely to be prone to question framing effects. There has been a debate on whether it would be feasible to elicit survival expectations in these countries where respondents are less likely to understand the concept of probability and the rules of nested events.

In Chapter 1, I examine several methodological considerations in eliciting survival probabilities in India. Individuals understand the concept of probabilities on average and are able to respond in meaningful ways, but I find that responses are sensitive to framing effects of own versus hypothetical person effects. I compare the individual survival probabilities to state specific life table estimates and find that overall individuals are relatively pessimistic about their survival probabilities. I then examine socioeconomic gradients in survival expectations and find that they vary with previously established socioeconomic predictors as expected. Younger individuals, those with more education and those from a higher caste report higher survival probabilities and these differences persist even when conditioning on self-reported health. To our knowledge this is the first attempt to compare survival probabilities to objective measures of health using biomarker data. I find that shorter respondents and those with decreased haemoglobin concentrations report lower survival probabilities. Also, respondents who have a higher survival probability are more likely to have an outstanding loan which is consistent with the idea that they are making an investment for their future.

Overall these results suggest that it is feasible to ask survival expectations of older respondents in India, the probabilities vary in meaningful ways with individual characteristics and they seem to have predictive power even conditional on self-reported health. So it is realistic to elicit survival expectations in India and they do contain valuable information which can be used to inform policy decisions. Moreover, the evidence shows that these subjective survival expectations have information over and above what is forecast by life table estimates. This is of particular relevance as India moves towards an affordable healthcare system for its citizens.

Having examined methodological considerations in eliciting survival probabilities, in Chapter 2, I focus on their formation. Previous research into subjective survival expectations have been limited with regards to the intricacies of how these expectations are formed. We have little knowledge of how these expectations are formed and whether these expectations are updated in response to new information in meaningful ways. I hypothesise that individuals who have been exposed to a survey induced negative health information shock would on average subsequently report lower survival probabilities. I examine this by using the longitudinal aspect of ELSA to analyse the effect of an out-of-range biomarker result on survival probabilities. I find that individuals who are exposed to a negative health information shock do report lower survival probabilities on average and the results do not seem sensitive to sample selection. I also extend the literature by systematically analysing the association between a pre-existing doctor diagnosed cardiovascular and chronic condition on survival probabilities and find that individuals who have been previously diagnosed report significantly lower survival expectations. These are the first estimates for England which confirm previous analysis on the relationship between doctor-diagnosed conditions and survival probabilities in the United States using the Health and Retirement Study. I also find that smokers update their survival probabilities differently as compared to non-smokers. These findings suggest that individuals rationally update their survival probabilities in response to new information lending evidence to their predictive power.

From wave 6 onwards, ELSA data has started recording medication use of respondents. It is possibly worthwhile to test if the relationships observed in this paper persist conditional on medication use. Furthermore, whilst evidence suggests that individuals do update their survival probabilities in the US and UK, this has as yet not been investigated in India. Once more waves of the LASI data are available, this would also be an interesting aspect for further research.

In Chapter 3, I investigate the variation in child mortality caused by domestic violence. Although a number of studies in India have examined the association between domestic violence and child mortality, a limitation thus far is that they do not account for the potential endogeneity of such an estimation. I attempt to overcome this by using an instrumental variable approach and use the price of gold at the time of marriage and dowry mechanics as an instrument for domestic violence. I find a positive and significant relationship between domestic violence and neonatal and infant mortality. This relationship persists through several specifications and do not seem to be driven by alternative mechanisms. These results confirm that domestic violence has negative effects not just for the victim of the abuse but also for the children of these victims.

Curbing domestic abuse would have major benefits not just for the health of mothers: it could lower child mortality rates and reduce violence within wider society. Government initiatives aimed at improving gender equality will be helpful in achieving this, but fundamentally Indians need to reassess how their society values women in general. This chapter hopes to induce a necessary reconsideration of women's role in Indian society.

Bibliography

Ackerson, L. K., Subramanian, S., 2009. Intimate partner violence and death among infants and children in india. Pediatrics 124 (5), e878–e889.

Adams, J., Stamp, E., Nettle, D., Milne, E. M., Jagger, C., 2014. Socioeconomic position and the association between anticipated and actual survival in older english adults. Journal of epidemiology and community health 68 (9), 818–825.

Afonso, S., 2014. Surprise end to india gold controls boosts wedding demand. Bloomberg.
URL `http://www.bloomberg.com/news/articles/2014-11-30/surprise-end-to-india-gold-controls-boosts-wedding-demand`

Agüero, J., 2013. Causal estimates of the intangible costs of violence against women in latin america and the caribbean.

Aguila, E., Borges, A., Castillejos, C. M., Pierson, A., Weidmer, B. A., 2014. Mortality expectations of older mexicans.

Aizer, A., 2011. Poverty, violence, and health the impact of domestic violence during pregnancy on newborn health. Journal of Human Resources 46 (3), 518–538.

Amaral, S., 2014. Do improved property rights decrease violence against women in india? Available at SSRN 2504579.

Anderson, S., 2007. The economics of dowry and brideprice. The Journal of Economic Perspectives, 151–174.

Anderson, S., Genicot, G., March 2014. Suicide and property rights in india. Working Paper 19978, National Bureau of Economic Research.
URL `http://www.nber.org/papers/w19978`

Arokiasamy, P., Bloom, D., Lee, J., Feeney, K., Ozolins, M., 2012. Longitudinal aging study in india: Vision, design, implementation, and preliminary findings. Aging in Asia: findings from new and emerging data initiatives, 36–74.

Arulampalam, W., Bhalotra, S., 2006. Sibling death clustering in india: state dependence versus unobserved heterogeneity. Journal of the Royal Statistical Society: Series A (Statistics in Society) 169 (4), 829–848.

Arunachalam, R., Logan, T. D., 2006. On the heterogeneity of dowry motives. Tech. rep., National Bureau of Economic Research.

Association, A. D., et al., 2010. Standards of medical care in diabetes2010. Diabetes care 33 (Supplement 1), S11–S61.

Balia, S., 2014. Survival expectations, subjective health and smoking: evidence from share. Empirical Economics 47 (2), 753–780.

Banerjee, K., 1999. Gender stratification and the contemporary marriage market in india. Journal of Family Issues 20 (5), 648–676.

Banks, J., Muriel, A., Smith, J. P., 2010. Disease prevalence, disease incidence, and mortality in the united states and in england. Demography 47 (1), S211–S231.

Basu, S., 1999. She comes to take her rights: Indian women, property, and propriety. SUNY Press.

Becker, G. S., 1991. A treatise on the family (enl. ed.). Cambridge, Mass: Harvard.

Bhalotra, S., Rawlings, S. B., 2011. Intergenerational persistence in health in developing countries: The penalty of gender inequality? Journal of Public Economics 95 (3), 286–299.

Bhalotra, S. R., Cochrane, T., 2010. Where have all the young girls gone? identification of sex selection in india.

Bhandari, S., 2013. The indian gold jewellery market. Alchemist Articles.

Bloom, D., Hu, P., Arokiasamy, P., Risbud, A., Sekher, T., Mohanty, S., Kale, V., OBrien, J., Chien, C., Lee, J., 2014. Longitudinal aging study in india: Biomarker documentation. RAND Work. Pap. WR-1043.

Bloom, D. E., 2005. Education and public health: Mutual challenges worldwide. Comparative Education Review 49 (4), 437–451.

Bloom, D. E., Canning, D., Moore, M., Song, Y., 2006. The effect of subjective survival probabilities on retirement and wealth in the united states. Tech. rep., National Bureau of Economic Research.

Brainerd, E., Menon, N., 2014. Seasonal effects of water quality: The hidden costs of the green revolution to infant and child health in india. Journal of Development Economics 107, 49–64.

Brown, P. H., 2009. Dowry and intrahousehold bargaining evidence from china. Journal of Human Resources 44 (1), 25–46.

Bühler, G., 1964. The Laws of Manu. Translated with Extracts from Seven Commentaries by G. Bühler. Delhi, Motilal Banarsidass [1964].

Burström, B., Fredlund, P., 2001. Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? Journal of epidemiology and community health 55 (11), 836–840.

Campbell, J., Torres, S., Ryan, J., King, C., Campbell, D. W., Stallings, R. Y., Fuchs, S. C., 1999. Physical and nonphysical partner abuse and other risk factors for low birth weight among full term and preterm babies: A multiethnic case-control study 150 (7), 714–726.

Chen, B., Mahal, A., 2010. Measuring the health of the indian elderly: evidence from national sample survey data. Population health metrics 8 (1), 30.

Chin, Y.-M., 2012. Male backlash, bargaining, or exposure reduction?: womens working status and physical spousal violence in india. Journal of Population Economics 25 (1), 175–200.

Clarke, P. S., Windmeijer, F., 2012. Instrumental variable estimators for binary outcomes. Journal of the American Statistical Association 107 (500), 1638–1652.

Cohen, G., Duffy, J. C., 2002. Are nonrespondents to health surveys less healthy than respondents? Journal of Official Statistics 18 (1), 13.

Culleton, B. F., Manns, B. J., Zhang, J., Tonelli, M., Klarenbach, S., Hemmelgarn, B. R., 2006. Impact of anemia on hospitalization and mortality in older adults. Blood 107 (10), 3841–3846.

D O'Broin, S., Gunter, E. W., 1999. Screening of folate status with use of dried blood spots on filter paper. The American journal of clinical nutrition 70 (3), 359–367.

de Bruin, W. B., Fischhoff, B., Millstein, S. G., Halpern-Felsher, B. L., 2000. Verbal and numerical expressions of probability:it's a fifty–fifty chance. Organizational behavior and human decision processes 81 (1), 115–131.

Delavande, A., 2014. Probabilistic expectations in developing countries. Annu. Rev. Econ. 6 (1), 1–20.

Delavande, A., Giné, X., McKenzie, D., 2011. Measuring subjective expectations in developing countries: A critical review and new evidence. Journal of Development Economics 94 (2), 151–163.

Delavande, A., Kohler, H.-P., 2009. Subjective expectations in the context of hiv/aids in malawi. Demographic research 20, 817.

Delavande, A., Kohler, H.-P., 2016. Hiv/aids-related expectations and risky sexual behaviour in malawi. The Review of Economic Studies 83 (1), 118–164.

Delavande, A., Rohwedder, S., 2011. Differential survival in europe and the united states: estimates based on subjective probabilities of survival. Demography 48 (4), 1377–1400.

DESA, U., 2010. United nations, department of economic and social affairs, population division: World urbanization prospects, the 2009 revision: Highlights.

Duflo, E., 2003. Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in south africa. The World Bank Economic Review 17 (1), 1–25.

Dyson, T., Moore, M., 1983. On kinship structure, female autonomy, and demographic behavior in india. Population and development review, 35–60.

Elder, T. E., 2007. Subjective survival probabilities in the health and retirement study: Systematic biases and predictive validity. Michigan Retirement Research Center Research Paper No. WP 159.

Factbook, C., 2010. The world factbook. See also: https://www. cia. gov/library/publications/the-world-factbook.

Farmer, A., Tiefenthaler, J., 1997. An economic analysis of domestic violence. Review of Social Economy 55 (3), 337–358.

Ghosal, S., 2015. Expect a 10 per cent rise in gold demand this wedding season. The Economic Times.

Grubb, M., 2015. Gold demand trends full year 2014. Tech. rep., World Gold Council.

Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G., Woodman, R. C., 2004. Prevalence of anemia in persons 65 years and older in the united states: evidence for a high rate of unexplained anemia. Blood 104 (8), 2263–2268.

Harvard School of Public Health, 2010. Longitudinal ageing study in india. Http://www.hsph.harvard.edu/pgda/lasi/lasi-2/.

Heath, R., Tan, X., 2014. Intrahousehold bargaining, female autonomy, and labor supply: Theory and evidence from india. Tech. rep., mimeo.

Hershman, P., 1981. Punjab kinship and marriage, ed. hilary standing.

Hurd, M. D., 2009. Subjective probabilities in household surveys. Annual Review of Economics 1, 543.

Hurd, M. D., McFadden, D., Merrill, A., 2001. Predictors of mortality among the elderly. In: Themes in the Economics of Aging. University of Chicago Press, pp. 171–198.

Hurd, M. D., McGarry, K., 1995. Evaluation of the subjective probabilities of survival in the health and retirement study. Journal of Human resources, S268–S292.

Hurd, M. D., McGarry, K., 2002. The predictive validity of subjective probabilities of survival*. The Economic Journal 112 (482), 966–985.

Hurd, M. D., Smith, J. P., Zissimopoulos, J. M., 2004. The effects of subjective survival on retirement and social security claiming. Journal of Applied Econometrics 19 (6), 761–775.

Idler, E. L., Benyamini, Y., 1997. Self-rated health and mortality: a review of twenty-seven community studies. Journal of health and social behavior, 21–37.

International Institute for Population Sciences, 2010. Longitudinal ageing study in india. http://iipsindia.org/research_lasi.htm#LASI-pilotdata.

Iyengar, R., 2009. Does the certainty of arrest reduce domestic violence? evidence from mandatory and recommended arrest laws. Journal of Public Economics 93 (1), 85–98.

Jadhav, R., 2015. Fall-off in indian weddings bodes ill for jewellers. Reuters.

Jejeebhoy, S. J., 1998. Associations between wife-beating and fetal and infant death: impressions from a survey in rural india. Studies in family planning, 300–308.

Kishor, S., Johnson, K., 2004. Profiling domestic violence: a multi-country study.

Koenig, M. A., Stephenson, R., Acharya, R., Barrick, L., Ahmed, S., Hindin, M., 2010. Domestic violence and early childhood mortality in rural india: evidence from prospective data. International journal of epidemiology 39 (3), 825–833.

Kypri, K., Samaranayaka, A., Connor, J., Langley, J. D., Maclennan, B., 2011. Non-response bias in a web-based health behaviour survey of new zealand tertiary students. Preventive medicine 53 (4), 274–277.

Lee, D. S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies 76 (3), 1071–1102.

Levin, I. P., Schneider, S. L., Gaeth, G. J., 1998. All frames are not created equal: A typology and critical analysis of framing effects. Organizational behavior and human decision processes 76 (2), 149–188.

Libby, P., Ridker, P. M., Hansson, G. K., 2011. Progress and challenges in translating the biology of atherosclerosis. Nature 473 (7347), 317–325.

Liu, J.-T., Tsou, M.-W., Hammitt, J. K., 2007. Health information and subjective survival probability: evidence from taiwan. Journal of Risk Research 10 (2), 149–175.

Manski, C. F., 2004. Measuring expectations. Econometrica, 1329–1376.

Manski, C. F., Molinari, F., 2010. Rounding probabilistic expectations in surveys. Journal of Business & Economic Statistics 28 (2), 219–231.

Marmot, M., Banks, J., Blundell, R., Erens, B., Lessof, C., Nazroo, J., Huppert, F., 2010. English longitudinal study of ageing: Wave 0 (1998, 1999 and 2001) and waves 1-4 (2002-2009).

Martin, S. L., Tsui, A. O., Maitra, K., Marinshaw, R., 1999. Domestic violence in northern india. American journal of epidemiology 150 (4), 417–426.

McKenzie, D., de Mel, S., Woodruff, C., 2008. Returns to capital: Results from a randomized experiment. Quarterly Journal of Economics 123 (4), 1329–72.

Menon, S., 2015. The impact of a negative health information shock on survival probabilities in england, unpublished.

Miller, P. M., Fagley, N. S., 1991. The effects of framing, problem variations, and providing rationale on choice. Personality and Social Psychology Bulletin 17 (5), 517–522.

Milman, N., Pedersen, A. N., 2009. Blood haemoglobin concentrations are higher in smokers and heavy alcohol consumers than in non-smokers and abstainersshould we adjust the reference range? Annals of hematology 88 (7), 687–694.

Monden, C. W., Smits, J., 2009. Maternal height and child mortality in 42 developing countries. American Journal of Human Biology 21 (3), 305–311.

Morrison, A. R., Orlando, M. B., 1999. Social and economic costs of domestic violence: Chile and nicaragua.

Nasir, K., Hyder, A. A., 2003. Violence against pregnant women in developing countries review of evidence. The European Journal of Public Health 13 (2), 105–107.

Newberger, E. H., Barkan, S. E., Lieberman, E. S., McCormick, M. C., Yllo, K., Gary, L. T., Schechter, S., 1992. Abuse of pregnant women and adverse birth outcome: current knowledge and implications for practice. jama 267 (17), 2370–2372.

Nordenberg, D., Yip, R., Binkin, N. J., 1990. The effect of cigarette smoking on hemoglobin levels and anemia screening. Jama 264 (12), 1556–1559.

O'Donnell, O., Teppa, F., Van Doorslaer, E., 2008. Can subjective survival expectations explain retirement behaviour? De Nederlandsche Bank.

ONS, 2014. Annual mid-year population estimates.

Organisation for Economic Co-operation and Development, 2015a. Life expectancy at 65. Data retrieved from OECD Indicator, `https://data.oecd.org/healthstat/life-expectancy-at-65.htm`.

Organisation for Economic Co-operation and Development, 2015b. Life expectancy at birth. Data retrieved from OECD Indicator, `https://data.oecd.org/healthstat/life-expectancy-at-birth.htm`.

Organization, W. H., et al., 2011. Use of glycated haemoglobin (hba1c) in diagnosis of diabetes mellitus: abbreviated report of a who consultation.

Oster, E., 2009. Proximate sources of population sex imbalance in india. Demography 46 (2), 325–339.

Patel, K. V., Semba, R. D., Ferrucci, L., Newman, A. B., Fried, L. P., Wallace, R. B., Bandinelli, S., Phillips, C. S., Yu, B., Connelly, S., et al., 2009. Red cell distribution width and mortality in older adults: a meta-analysis. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, glp163.

Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon, R. O., Criqui, M., Fadl, Y. Y., Fortmann, S. P., Hong, Y., Myers, G. L., et al., 2003. Markers of inflammation and cardiovascular disease application to clinical and public health practice: a statement for healthcare professionals from the centers for disease control and prevention and the american heart association. Circulation 107 (3), 499–511.

Peedicayil, A., Sadowski, L. S., Jeyaseelan, L., Shankar, V., Jain, D., Suresh, S., Bangdiwala, S. I., 2004. Spousal physical violence against women during pregnancy. BJOG: An International Journal of Obstetrics & Gynaecology 111 (7), 682–687.

Peracchi, F., Perotti, V., 2009. Subjective survival probabilities and life tables: an empirical analysis of cohort effects. Genus 65 (2), 23–57.

Perozek, M., 2008. Using subjective expectations to forecast longevity: Do survey respondents know something we dont know? Demography 45 (1), 95–113.

Petersen, R., Connelly, A., Martin, S. L., Kupper, L. L., 2001. Preventive counseling during prenatal care: Pregnancy risk assessment monitoring system (prams). American Journal of Preventive Medicine 20 (4), 245–250.

Rawlings, S., Siddique, Z., 2014. Domestic abuse and child health.

Rodriguez-Lesmes, P. A., 2014. Using subjective survival expectations for policy analysis: the case of diagnosis of hypertension and smoking, unpublished Manuscript.

Roser, M., 2016. Life Expectancy. Data retrieved from Our World in Data, `http://ourworldindata.org/life-expectancy/`.

Sample Registration System, 2012. Analytical studies - report number-1. Tech. rep.

Sargeant, L. A., Khaw, K.-T., Bingham, S., Day, N. E., Luben, R. N., Oakes, S., Welch, A., Wareham, N. J., 2001. Cigarette smoking and glycaemia: the epic-norfolk study. International journal of epidemiology 30 (3), 547–554.

Schoenbaum, M., 1997. Do smokers understand the mortality effects of smoking? evidence from the health and retirement survey. American Journal of Public Health 87 (5), 755–759.

Sekhri, S., Storeygard, A., 2014. Dowry deaths: Response to weather variability in india. Journal of development economics 111, 212–223.

Siwan, A., Garance, G., 2014. Suicide and property rights in india. NBER Working Papers.

Smith, J. P., 2007. Nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and the socioeconomic status health gradient. Proceedings of the National Academy of Sciences 104 (33), 13225–13231.

Smith, V. K., Taylor Jr, D. H., Sloan, F. A., Johnson, F. R., Desvousges, W. H., 2001. Do smokers respond to health shocks? Review of Economics and Statistics 83 (4), 675–687.

Srinivas, M. N., 1984. Some reflections on dowry. No. 1983. Published for the Centre for Women's Development Studies, New Delhi, by Oxford University Press.

Steckel, R. H., 1979. Slave height profiles from coastwise manifests. Explorations in Economic History 16 (4), 363–380.

Straus, M., Hamby, S., Boney-McCoy, S., Sugarman, D., Finkelhor, D., 1973. Conflict tactics scales (cts).

Strauss, J., Thomas, D., 2007. Health over the life course. Handbook of development economics 4, 3375–3474.

Subramanian, S., Subramanyam, M. A., Selvaraj, S., Kawachi, I., 2009. Are self-reports of health and morbidities in developing countries misleading? evidence from india. Social science & medicine 68 (2), 260–265.

Tauchen, H. V., Witte, A. D., Long, S. K., 1985. Domestic violence: A non-random affair.

The World Bank, 2015. Life expectancy. Data retrieved from World Development Indicators, `http://data.worldbank.org/indicator/SP.DYN.LE00.IN`.

Tolentino, K., Friedman, J. F., 2007. An update on anemia in less developed countries. The American journal of tropical medicine and hygiene 77 (1), 44–51.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.

UNICEF, 2012. Infant and child mortality in india: Levels, trends and determinants.

UNICEF, Organization, W. H., et al., 2001. Iron Deficiency Anaemia: Assessment, Prevention, and Control: a guide for programme managers. WHO.

United Nations Population Division (UN), 2014. World population prospects: The 2015 revision.

United Nations Population Division (UN), 2015. Levals and trends in child mortality.

Urberg, M., Shammas, R., Rajdev, K., 1989. The effects of cigarette smoking on glycosylated hemoglobin in nondiabetic individuals. Journal of Family Practice 28 (5), 529–531.

Vaidyanathan, A., 1999. Consumption of gold in india: Trends and determinants. Economic and Political Weekly, 471–476.

Walby, S., 2004. The cost of domestic violence. Department of Trade and Industry , United Kingdom.

Waters, H., Hyder, A., Rajkotia, Y., Basu, S., Rehwinkel, J. A., 2004. The economic dimensions of interpersonal violence.

Winter, J., 2008. Population aging, intergenerational transfers and the macroeconomy. Ageing and Society 28.

Wood, Z., Wachman, R., 2010. Every indian wedding adds to demand in world's biggest jewellery market.
URL http://www.theguardian.com/business/2010/oct/10/indian-jewellery-market

World Health Organization, 2001. Iron deficiency anaemia: assessment, prevention and control: a guide for programme managers.

World Health Organization, 2011. World health statistics. Tech. rep.

Zhang, T., 2012. An estimation of the economic impact of spousal violence in Canada, 2009. Department of Justice Canada.