Essays on Environmental and Health Economics

Prachi Singh

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

"To my parents, sister and Eva"

Essays on Environmental and Health Economics

Prachi Singh

October 2019

Thesis Supervisor : Dr. Abhiroop Mukhopadhyay

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Acknowledgements

I am deeply grateful to my advisor Dr. Abhiroop Mukhopadhyay for his invaluable guidance and constant support throughout the period of my PhD work. I am grateful to him as he was the first person whom I approached for any query, small or big, and he was always very generous with his time. His expertise on empirical methods was crucial for laying the foundation for all my research papers. His focus on rigour in empirical research has always motivated me to do better and my discussions with him have honed my skills as a researcher.

I would like to thank all the faculty members of the Economics and Planning Unit of Indian Statistical Institute, Delhi Centre for their teaching and support. I am grateful to Dr. E. Somanathan for his comments, helpful advice on my papers and for introducing me to the Environment for Development (EfD) network which has improved my research and helped me grow as a researcher in this field. I am also thankful to Dr. Farzana Afridi for her encouragement and for providing me with valuable feedback on my research. Dr. Tridip Ray has been the kindest soul in the department and the institution was truly a home for many scholars like me because of him. I am also thankful to my co-authors - Dr. Mudit Kapoor, Dr. Ambuj Roy, Dr. Shamika Ravi, Dr. Dinkar Bhasin, Dr. Sagnik Dey, Sourangsu Chowdhury and Kunal Bali for always believing in me when I did not. I am grateful to Dr. Shyamal Chowdhury for his unwavering support and mentorship despite the long distance, and for giving me copious amounts of his time and energy. I am also extremely thankful to the helpful online #econtwitter community which has helped me learned so much more personally and professionally. My work has also benefited from interactions with many scholars and participants at various seminars and conferences. I would like to express my gratitude towards them.

I am thankful to Indian Statistical Institute for the research fellowship and the ideal institutional support it has provided. I am especially grateful to Athisii who has helped me so many times with my queries related to geo-spatial data. The process of research can be gruelling and isolating, I was able to go through these periods of ups and downs because of the support which I got from my family and friends. I would especially like to thank Soham Sahoo and Sabyasachi Das for always extending their help when I needed it. I cherish my friendship with Rohan, Anirban, Pranabesh, Komal, Ahana, Sujaya, Arka, Soumendu, Digvijay, Priyanka and many more. I am eternally thankful to my exceptional friend Monil, he has helped me personally umpteen times.

Above all, I want to thank my parents and sister for their unconditional love, patience, sacrifices and encouragement. My father, P.R. Singh and mother, Suchitra Singh have always inspired me with their simplicity. I am what I am because of them. My sister, Nidhi Singh employed a weird mix of love, humour and sarcasm to encourage me to complete this journey. My niece, Eva, has been the motivating factor for completing this thesis sooner than later. Finally, I am thankful to my partner, Dhritiman for his love and invaluable friendship throughout my PhD journey. His sense of humour, timely discussions, belief in me and constant encouragement have been the source of my strength for years. I can't thank him enough for introducing me to the magical world of books which often provided me with great comfort.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis consists of three empirical essays that investigate issues in environmental and health economics. The main focus of this thesis is to study how environmental degradation or interventions in this domain can affect demographic and health outcomes of the population in developing countries. The first chapter explores the effect of an information campaign about arsenic contamination in Bangladesh on marriage market outcomes. This study finds that providing information about negative effects of arsenic consumption had an unintended consequence in the marriage market. Specifically, information about lower fertility, skin lesions, cancers and higher mortality related to arsenic exposure induced individuals to get married earlier and reduced bride price. The second essay investigates the effect of exposure to outdoor pollution during the in-utero period on child health outcomes: weight-for-age and height-for-age in India. This paper focuses on post survival measures of child health outcomes and solves the endogeniety issues related to pollution exposure by using an instrumental variable strategy. This paper finds that an increase in exposure to pollution during the first trimester reduces height-for-age and weight-for-age for children. Children from Northern India, belonging to poorer households and born to mothers with low level of education are found to be particularly vulnerable to the negative effects of pollution. The third chapter looks into the agricultural practice of biomass burning (crop residue burning) and its association with cardiovascular health for four northern states of India, thus contributing to the emerging literature on the crop residue burning and its health effects. This paper finds that individuals who get exposed to high levels of biomass burning have a greater likelihood of being hypertensive.

The following sections provide a brief description outlining the research questions, empirical strategies and main findings of each chapter of the thesis.

1.1 Information Campaign on Water Quality and Marriage Market: The Case of Arsenic Exposure in Rural Bangladesh

Arsenic contamination of drinking water has caused a major health emergency in Bangladesh owing to multiple health problems associated with it, which range from skin lesions to various types of cancers. It remained largely unknown and became public knowledge only later through a nationwide information campaign. In this chapter, we study the impact of the information campaign on marriage patterns in Bangladesh. Using data on arsenic contamination, we categorize sub-districts of Bangladesh into arsenic and non-arsenic areas and then use a difference-in-difference strategy to evaluate the impact of information campaign on marriage market outcomes. We analyse mainly two marriage market outcomes - the age at marriage and the bride price agreed at the time of marriage. In our empirical specification, we account for sub-district level fixed effects to account for initial differences in outcomes due to location (and differing levels of arsenic contamination) and year of marriage fixed effects to account for secular trends in outcomes. Further we include district level linear trends in our analysis to remove the impact of district level time varying variables. We find that age at marriage and bride price both reduced in arsenic affected areas in comparison to non-arsenic affected areas in response to information campaign. The age at marriage for males reduced by 1 to 3.5 percent and by 0.67 to 0.94 percent for females. Bride price reduced by around 64 percent. The campaign owes its success in terms of generating awareness about water contamination to its unique design. Constant visual reminders, multiple strategies to avoid contaminated water and public forums for information disbursal seem to have worked in Bangladesh. We find evidence that such campaigns which informs people about adverse outcomes, may also lead to undesirable social outcomes; we demonstrate one such consequence in the marriage market.

1.2 Early Life Exposure to Outdoor Air Pollution: Effect on Child Health in India

This chapter examines effect of outdoor air pollution on child health in India. The literature has primarily focused on mortality related outcomes but our paper focuses on post survival measures of child health which includes height-for-age (stunting measure) and weight-for-age (underweight measure). Due to extremely low level of ground based coverage by pollution monitors, we rely on satellite PM2.5 data for our analysis. We combine satellite PM2.5 data with geo-coded Demographic and Health Survey of India(2016). We use an instrumental variable strategy for identification as local pollution levels may be endogenous due to local household behavioural choices like participation in local fuel wood market, burning crop residue etc which are not observed in survey data. Our identification strategy relies on use of wind direction, specifically we use upwind biomass burning events (smoke from these events are flowing towards the sampled clusters) in neighbouring areas to identify the effect of air pollution on child health. We also control for child, mother and household characteristics in our analysis. Further, different clusters can have different levels of development (health infrastructure) which can affect health of a child hence we include cluster level fixed effects in our specification. We also remove any omitted variables that are related to a district in any particular year as well as any seasonality effect specific to a month of a particular year by including district-year and month-year fixed effects.

Our results indicate that one standard deviation increase in exposure to pollution during first trimester lowers Height-for-age (by 7.9 percent) and Weight-for-age (by 6.7 percent). Our study shows that the negative effect of exposure to pollution during the in-utero period doesn't go away with age, as we observe negative effects even for children who are 3 to 5 years old. Children belonging to North India, poorer households and born to mothers with low level of education are found to be particularly vulnerable. A back of the envelope calculation suggests that one standard deviation increase in pollution leads to a 0.18 percent reduction in GDP.

1.3 Impact of Biomass Burning on Blood Pressure: A Study from North India

This chapter investigates whether the agricultural activity of biomass burning affects cardiovascular health of the population. The satellite data on biomass burning is combined with geo-coded Demographic and Health Survey of India(2016) for our analysis. The study focuses on four North Indian states of Punjab, Haryana, Uttar Pradesh and Bihar which exhibit differential biomass burning patterns. In our analysis, an individual is considered to be exposed to high intensity biomass burning if he/she experiences more than 100 fire-events during last 30 days (short term exposure) from the date of interview. Cardiovascular health is captured in the health survey via blood pressure readings, we use these readings to categorize individuals into two groups - hypertensive and non-hypertensive. Our empirical strategy uses a logistic model which explores the relationship between hypertension and high intensity biomass burning. The study accounts for individual and household level risk factors. District, month and year fixed effects are also used to account for any regional or temporal unobserved factors. Our results indicate that short term exposure to high intensity biomass burning is associated with 1 percent increase in likelihood of being hypertensive. The effect is most prominent for older cohorts for both males and females (5.8 and 3.2 percent respective increase in probability of being hypertensive due to high intensity biomass burning). We estimate that elimination of biomass burning would avert 70 to 91 thousand disability-adjusted life years lost per year, valued at USD 520 to 675 Million over 5 years.

Chapter 2

Information Campaign on Water Quality and Marriage Market: The Case of Arsenic Exposure in Rural Bangladesh

2.1 Introduction

Arsenic contamination of drinking water is a major public health problem all over the world. While it is fairly wide-spread, having being detected in atleast 70 countries [1], it is especially severe in South Asia[2], and in particular in Bangladesh. In order to address this problem, while various solutions have been suggested: for example, developing alternative water sources and methods for arsenic testing and removal (Ravenscroft et al (2009)); educating the population through information campaigns still remains one of most important public policy tools. Given the gravity of the problem, Bangladesh was one of the first countries to inform its exposed population about water quality and the various harmful effects of consuming arsenic on human health. In this paper, we seek to estimate the effect of the information campaign on marriage market outcomes, focusing on two outcomes the age at marriage and the bride price (an integral part of marriages in Islamic culture).

Our study is motivated by the recognition by researchers and commentators that while the health effect of arsenicosis (arsenic poisoning) have been extensively studied, there is relatively sparser literature on the socio-demographic (and mental) effects of such poisoning (Brinkel et

[1] Ravenscroft (2007)

[2] Discussed in Mukherjee et al. (2006)

al. (2009)). If such effects are indeed important, then information campaigns increasing awareness may have important demographic consequences, some unintended. For example, while an information campaign telling communities that they are at risk of arsenic exposure may help people take precautions to prevent exposure, it may also change social perceptions towards adults who have already been exposed, irrespective of whether they show any symptoms of poisoning. These changed perceptions would then cause changes in the market marriage equilibrium, some of which may be socially undesirable. Bride price is one such outcome and is a subject of our study. To elaborate further, bride price is the amount which is agreed upon at the time of the marriage, which has to be paid by the grooms family to the brides family in the event of a divorce. Since perceptions of future health, future physical appearance and longevity, informed by information campaigns, may affect the equilibrium price, we focus on this important outcome of the marriage market. Further, for those who live in affected regions, greater awareness that they may be affected by arsenic, may cause behavioural changes: for example, it is plausible that it would make them more risk averse in the marriage market. Such changes in risk perception may significantly affect age at marriage, with more risk averse males marrying sooner than their more risk-loving counterparts (Spivey (2010)) fearing the onset of their symptoms. Hence one may expect changes in the age of adults in the marriage market as a consequence of greater awareness due to information campaigns. Our paper is one of the first to look at the impact of such an information campaign in Bangladesh, on these two important outcomes of the marriage market.

The information campaign was implemented from 1999 to 2005[3] and it was designed to create awareness about ill effects of arsenic in drinking water, visually complementing the information by painting tubewells into red (dangerous) and green (safe) categories and pursuing users to switch to safe sources of drinking water. It also suggested mitigation strategies such as shifting to a safer well in the neighbourhood or well-sharing and finally it informed people about harmful health effects of arsenic exposure via public forums. The information campaign was successful in generating awareness and various studies (Chen et al. (2007); Opar et al. (2007); Bennear et al. (2013); Keskin et. al (2017); Jakariya, M (2007) etc) have

[3] World Bank Completion report 2007 mentions that during the initial 2.5 years the programme didn't take off but gained steam only later that is in 2002

found evidence on reduction of usage of contaminated wells in response to the campaign.

The health outcomes associated with arsenic affect multiple attributes which matter in marriage market. Kalmijn (1998) discusses in his survey that marriages exhibit sorting of prospective matches along many attributes such as age, education, income, race, height, weight, and other physical traits indicative of health status. Buss in his study which spanned 37 cultures found that females value *resource acquisition* in males, while males place high value on *reproductive capacity* in females (Buss (1989)). Other research on marriage market has also established that physical attractiveness and BMI (a rough measure of health) is also valued in spousal match process. Most of these "valued" characteristics are adversely affected by arsenic exposure and thus we explore how individuals looking for a prospective match in marriage market, tend to react to a negative information about health outcomes associated with arsenic exposure.

Information campaigns have been extensively evaluated in economics. Information campaigns, especially the ones which have gain framing in their messaging like "exercising regularly can help you lose weight" are found to be more effective when it comes to inducing behaviour change (Gallagher et al. (2011)). Studies have documented how information campaigns nudge people to change behaviours. For example, Kirby et al. using a randomised control trial find that information about human immunodeficiency virus (HIV) and sexually transmitted disease (STD) had an effect in terms of reducing risky sexual behaviour in males who previously indulged in unprotected sex (Kirby et al. (2004)). Duflo analysis effect of HIV related curriculum in schools in Kenya and finds that girls switch to committed relationships and they are significantly more likely to report faithfulness as a way they protect themselves from HIV (Duflo et al. (2015)). However, these studies focus on the impact of information campaigns on their intended consequences. In this paper we focus on unintended consequences of information campaigns.

To evaluate the impact of the information campaign, the paper uses a Difference in Difference (DID) strategy. As mentioned above, the information campaign was implemented all over Bangladesh in 2002 [4]. However, since it informed people about arsenic contamination,

[4]Keskin et al. (2017) also use the same cutoff for their analysis due to late actual implementation of the programme

the campaign was more relevant for those areas which had high level of arsenic contamination. Given this, we define sub-districts with arsenic levels in excess of 50ug/litre as the treatment group and those below as the control group. These arsenic levels were based on British Geological Survey (BGS) which tested a representative sample of tubewells in 61 districts of Bangladesh. The BGS was conducted in 1999 before the information campaign was implemented. Hence our DID strategy considers the period 1990-2001 as the pre-intervention period and the years 2002-2011 as the post intervention period and compares the difference in the pre-post period outcomes for the treatment group relative to an analogous difference for the control group. In other words, we use the differential impact of information campaign in sub-districts with differential amount of arsenic contamination at the sub-district level. In our specifications, we account for sub-district level fixed effects to account for initial differences in outcomes due to location (and differing levels of arsenic contamination) and year of marriage fixed effects to account for secular trends in outcomes. Further we include district level linear trends in our specifications to remove the impact of district level time varying variables.

Some of the crucial assumptions behind this empirical strategy are important to point out: first, while the information campaign was implemented all over Bangladesh in 2002, we assume that it had greater relevance in sub-districts with high arsenic contamination. If the campaign also had a similar impact in low arsenic contamination sub-districts, this will tend to attenuate the impact of the intervention. Second, since the initial levels of arsenic are different between the treatment and control group (by design), the assumption is that the trends in outcomes, if there was no information campaign, would not vary by these initial arsenic levels. To be more precise, our identifying assumption is that conditional on sub-district and year of marriage fixed effects and district linear trends, our marriage market outcomes do not changing differentially in arsenic versus non-arsenic contaminated sub-districts other than for reasons related to information campaign. To assuage concerns on this front, we conduct an event study to confirm that this identifying assumption does hold true, that is the trends in age at marriage (and bride price) are not statistically different between contaminated and uncontaminated areas in the period before information campaign was executed.

Data for age at marriage analysis is sourced from Census Data of Bangladesh for 2011. We use data on all married individuals captured in the census. For data on bride price we use a primary survey dataset from Palli Karma-Sahayak Foundation (PKSF) Survey conducted by researchers from University of Sydney in December 2010-January 2011. This survey contained modules on marriage, divorce, bride price and dowry. Additionally this survey contains sub-district identifiers which allow us to conduct the DID analysis. Using the sample of individuals who got married between 1990 and 2011[5], we find that both males and females got married earlier and bride price reduced after the information about arsenic in drinking water was made public. In particular, we find that age at marriage for males reduced by 2.6 months (a reduction of 1%) and for females it reduced by 1.4 months (a reduction of 0.67%) in arsenic affected areas in response to information campaign. We also find that bride price reduced by almost 64% in arsenic contaminated areas in response to the information campaign. These results survive a battery of robustness checks.

We also complement our results related to age at marriage by using other demographic surveys such as Demographic Health Surveys for Bangladesh (BDHS) which uses rich spatial information about arsenic contamination of drinking water by relying on GPS coordinates of the interviewed cluster. We provide further support for our results by matching arsenic to non-arsenic sub-districts (using Integrated Public Microdata Series - IPUMS data) based on their demographic and other characteristics and then comparing changes in marriage market outcomes between these two groups. Results using these alternate data-set confirm the earlier results: the effect of information campaign on age at marriage is found to be negative with a even larger magnitude.

Our results are consistent with the hypothesis that individuals who live in arsenic areas learn about their likelihood of developing skin lesions and other health problems due to information campaign and tend to get married earlier in order to avoid discovery of symptoms of diseases related to arsenic exposure like skin lesions, cancer etc. (which in an alternative scenario, where they do get discovered would make their chances of finding a mate quite difficult). Similarly in case of bride price, public knowledge about arsenic contamination in an area generates concerns about beauty and fertility of a prospective female match leading to a lower bride price which gets offered. While our results are consistent with these explanations, a more detailed investigation of these mechanism is outside the scope of this paper due to

[5]Period of analysis for bride price is 1990 to 2010.

limitations of data.

Our paper contributes to the literature of how information campaigns about health cause behavior changes in health and other socio-economic and demographic outcomes. Impact of information campaigns, aimed at health, that cause changes in how people behave, in health seeking behavior and other socio-economic dimensions, have been studied in other contexts. Oster et al. (2013) find that in case of Huntington disease which is a hereditary disease with ramifications on life expectancy, individuals who are not tested for the disease are optimistic about their health and make decisions as if they werent diagnosed with Huntington. However, individuals with confirmed diagnoses (to whom true picture about their health status has been revealed) behave differently. Goldstein et al. (2008) find that HIV-positive mothers who learn their status are more likely to receive medication to prevent transmission to their children. The effect of provision of information on behaviour has also been documented in other studies like Altman and Traxler (2014) (reminders about dental health increase check-up appointments with dentists), Calzo- lari and Nardotto (2014) (reminders induce gym users to increase physical activity), Alaii et al. (2003); Habluetzel et al. (1997); Binka et al. (1996) (information about benefits of using bed-nets to prevent malaria increased usage of bed-nets), Cairncross et al. (2005); Curtis et al. (2001); Luby et al.(2010) (dissemination of information about importance of hand washing to reduce infections increased frequency of hand washing).

In the case of Bangladesh, the literature of the impact of the information campaign we are focusing on is limited to Keskin et al. (2017) that looks at how mothers in Bangladesh increased breastfeeding duration for infants in response to the information about water quality. However none of the studies look at how such information campaigns affect the marriage market. In particular, this paper is the first to examine the causal effect of information about health outcomes related to poor water quality in the neighbourhood on marriage market outcomes. In the process, our paper links information about local disease environment to marriage market outcomes.

Given that we look at the impact of the information campaign on marriage market outcomes, it is important to acknowledge that individuals may already have responded to visible symptoms of arsenic poisoning without knowing the cause. Literature has documented effect of arsenic contamination on longevity for adults, beauty and fertility for females (Hassan et al. (2005); Milton et al. (2005); Rahman et al. (2007); Sohel et al. (2009); Argos et al. (2010)). All of these factors play an important role in the marriage market (Buss (1989)). An information campaign can make these aspects more salient by removing the uncertainty on cause and may therefore still have an effect. However, the pre-existing recognition of some of these effects does imply that the estimated impacts of the information campaign in our paper are likely to be an underestimate.

The paper is organized as follows, the following section provides background on water quality and related policies undertaken in Bangladesh. Section 3 describes various data sources that we use in our analysis. Section 4 describes our empirical strategy and Section 5 presents the corresponding results. Section 6 provides a discussion about the patterns we observe for marriage market and Section 7 concludes.

2.2 Background

2.2.1 Effects of Exposure to Arsenic

Exposure to arsenic causes many health problems which have been explored in bio-medical literature. Skin lesions are among the first few symptoms of arsenic poisoning: a longitudinal study (Ahsan et al., (2006)) has shown that higher dosage of arsenic is positively associated with higher probability of appearance of skin lesions. In addition to skin lesions, other studies have found that risk of all-cause and chronic-disease mortality is higher for people exposed to dangerous level of arsenic (Argos et al. (2010) and Sohel et al. (2009)). Adverse pregnancy outcomes like still birth, spontaneous abortion have also been linked to arsenic exposure (Milton et al. (2005); Rahman et al. (2006)). Various types of cancers (skin, kidney, bladder) and adverse effect on mental health (Chowdhury et al. (2016)) are among other devastating effects of arsenic exposure. The arsenic problem is not just a health problem but a social problem as well. The physical manifestation of early symptoms of arsenicosis like skin lesions is more than just a health effect. Given its visibility on the body, it has larger manifestations on the marriage market, as it is linked to the "beauty" of a person. Due to lack of information and illiteracy skin lesions are often confused with leprosy, which is considered a contagious killer by rural people. The early symptoms of arsenicosis which includes formation of black spots and warts thus leads to ostracism and social isolation (Alam et al. (2002)). Additionally, beyond the social problem, arsenic exposure has also been found to have implications in the labour market in form of reduced labour supply (Carson et al. (2010)).[6]

2.2.2 Arsenic in Water: Bangladesh

During 1970s Bangladesh was struggling with high disease burden due to water borne diseases related to surface water usage. Soon millions of tubewells were dug all across the country. In rural areas the switch to groundwater source was almost universal with almost 95% of households using tubewells (Caldwell et al. (2003)). A nation wide survey conducted by British Geological Survey (BGS) in 1999 found that most of the tubewells being used were of shallow depth and in few areas these tubewells contained dangerous levels of arsenic (shallow tubewells have a greater likelihood of containing arsenic)[7]. The BGS and Department of Public Health Engineering of Bangladesh (DPHE) report (2001) estimated that around 35 Million people were inadvertently exposed to harmful levels of arsenic by sourcing water from tubewells. Smith et al. (2000) also documents that between 35-77 Million people were exposed to arsenic due to contaminated tubewells usage.

2.2.3 Intervention: Information Campaign on Arsenic

The Bangladesh government (Department of Public Health Engineering(DPHE)) along with UNICEF and Non-Profit Organisations implemented a water quality information campaign between 1999 and 2005 called the Bangladesh Arsenic Mitigation and Water Supply Programme (BAMWSP). The campaign didn't pick up any steam during the initial 2.5 years (World Bank (2007) - project completion report) with major roll-out happening during later years especially after 2002. In a paper which looks at the impact of same the intervention on a different outcome (Keskin et al. (2017)) uses 2002 as the cut-off for treatment period. The

[6]Water quality can additionally have an effect on income generating activities as well like fishing. Reference: Baggio, M (2016), Baggio, et al. (2019) and Ranta et al. (1992)

[7]Arsenic was first detected in groundwater in 1985, however no comprehensive study ensued after this discovery until late 1990s, that is when British Geological Survey (BGS) conducted a nation wide survey of tubewells.

campaign had few distinct features: it disseminated information about contamination of tubewells by color coding tubewells into red (unsafe) and green (safe) categories; it also suggested mitigation strategies such as shifting to a safer well in the neighbourhood or well-sharing and finally it informed people about harmful health effects of arsenic exposure via public forums.

The constant visual reminders along with negative information about adverse health effects of arsenic exposure did have an impact on the intended population. The information campaign was quite effective in terms of generating awareness in terms of people reporting that they had heard of arsenic after information campaign and creating awareness about symptoms of arsenicosis after the information campaign (Keskin at al., 2017). Other papers (Chen et al. (2007); Opar et al. (2007); Bennear et al. (2013); Balasubramanya et al. (2014) etc) also find that people did switch to safer sources of water after the information campaign.

2.3 Data

In order to analyse the causal impact of information campaign on marriage market outcomes our analysis combines arsenic contamination information with the demographic data. We describe below our various sources of data:

2.3.1 Contamination data

British Geological Survey (BGS)- 1999

We source data on arsenic contamination, as present in 1999 from British Geological Survey (BGS) conducted in 1998-1999. In this survey, BGS tested a geographically representative sample of 3534 wells all across the country and also recorded the GPS coordinates of the tubewells along with other details such as the depth of the tubewell and the year in which it was constructed. Bangladesh has 64 districts in total but BGS sampled only 61 of them. Hence our study is restricted to these 61 districts. This survey found that more than a quarter of total sampled tubewells contained harmful level of arsenic (Bangladesh Government recognizes >50ug/litre as the dangerous level of arsenic contamination, however WHO prescribes >10ug/litre as the dangerous cutoff).

We categorize sub-districts into arsenic and non-arsenic affected group using mean arsenic contamination for each sub-district. This is calculated by averaging arsenic contamination level for all tubewells located in a sub-district[8]. To decide the threshold level of contamination that may be deemed dangerous, we use the value of $50ug/litre$ as recognized by the Bangladesh government. If the mean arsenic figure for a sub-district was above 50ug/litre then we code it as an arsenic affected sub-district and a sub-district with contamination less than 50ug/litre is coded as non-arsenic affected sub-district. Figure 2.1 shows the geographical variation in arsenic contamination at the sub-district level. The light shaded (green) sub-districts represent non-arsenic affected sub-districts (i.e. arsenic contamination was lower than 50ug/litre) while the dark shaded (red) show arsenic affected sub-districts. As is evident from the figure, contaminated tubewells are more likely to be in the south.

2.3.2 Demographic data

Long Census Survey Data - 2011

The individual and household data are obtained from Long Census Survey Data which was conducted by Bangladesh Bureau of Statistics (BBS) in March 2011. This dataset contains details about place of residence (sub-district), duration of residence in the current district, type of construction of house, education of household members and details related to nuptiality which includes our main variable of interest, that is the age at which an individual first got married. While the 2011 census surveyed 167,293 households across all 64 districts of Bangladesh, for the purpose of analysis in this paper, we focus our attention on the rural sample where the tubewell use was almost universal before the implementation of information campaign (Caldwell et al. (2003)). We combine the census data with arsenic contamination data (BGS data) to sub-districts to their arsenic contamination status. Our final estimation sample comprises of married individuals (75243 males and 86985 females) from 420 sub-districts in Bangladesh. Out of the total 420 sub-districts, 134 (around 32 percent) were found to have arsenic levels above 50ug/litre. It should also be noted that we don't observe the female's original household, that is the household they were born in. We only observe females in the households they got married into, and for this reason we primarily focus on males and suggest interpreting

[8] On an average this survey tested 8 tubewells in each sub-district.

results from the female sample with caution[9].

We provide summary statistics for the male and female estimation sample in Table 2.1 (panels A and B). A greater proportion of both males and females are literate in arsenic areas in comparison to non-arsenic areas [10] In arsenic areas, a greater proportion of individuals (both males and females) belong to households who own assets (land and house). However, this does not mean that they are necessarily richer as only a small proportion of these households have cemented walls. Most of the individuals in both arsenic and non-arsenic regions have their religion as Islam. The difference in characteristics in individuals imply that we cannot compare outcomes across arsenic and non-arsenic regions. Hence as we describe later, we resort to looking at changes over time.

We describe next our outcome variable of interest, age at marriage. When we plot the age at which males and females get married in Figure 2.2(a), we observe that males in Bangladesh tend to get married earlier than the world average with the mean age at first marriage being 24 years with 70 percent of males being married by 25 years of age (70 percent of females were married by the age of 18.3 years, Figure 2.2(b). In Figure 2.3 we plot mean age at marriage overtime for arsenic and non-arsenic areas. The mean age of marriage is higher for both males and females in arsenic affected areas. The age at marriage has been decreasing for males overtime (Figure 2.3(a)) while it has been increasing for females (Figure 2.3(b)). However, the gap (or difference) in age at marriage between arsenic and non-arsenic areas has narrowed especially after the implementation of information campaign (represented by green solid vertical line for year 2002). This relationship is shown in Figures 2.3(c) and 2.3(d) where we plot the difference in age at marriage between arsenic and non-arsenic areas overtime for males and females. We observe that the difference in age at marriage is lower in the post treatment period that is after the implementation of information campaign. Our empirical analysis tests the robustness of this result by controlling for changes in other covariates.

[9]In section 6, we describe how males from arsenic areas get matched to females from arsenic areas in marriage market, which mitigate some concerns when interpreting results for females.

[10]Table 2.1 doesn't show statistical difference between the two groups (arsenic and non-arsenic), however differences are significant for all variables in both male and female sample.

Primary Survey

For our bride price analysis we rely on Palli Karma-Sahayak Foundation (PKSF) Survey. This survey contained modules on marriage, divorce, bride price and dowry. It was conducted by researchers from University of Sydney in December 2010-January 2011. We combine this survey with contamination data (from BGS) to arrive at our final estimation sample for the period of study (1990-2010)[11], with information on 875 marriages from 20 sub-districts. 9 out of these 20 sub-districts were arsenic affected (arsenic contamination greater than 50ug/litre). This dataset contains detailed information about marriage unions which includes year of marriage, the amount of dowry paid, bride price agreed upon, age of bride and groom, education of bride groom and income status of bride and groom's family at the time of marriage. In Table 2.1, Panel C we summarize our data from this survey, education of both bride and groom is more for couples belonging to the arsenic group (the difference is statistically significant when compared to couples belonging to the non-arsenic group). The age at which brides and grooms get married is not statistically different between arsenic and non-arsenic group. However, the age at marriage is particularly low for females, with mean age at marriage being 16.75 years. We also observe that a small proportion of marriages are those where the groom was chosen by a bride rather than by her family. Lastly the mean bride price amount is close to 53300 taka (difference between arsenic and non-arsenic group is not statistically significant). The mean dowry amount is 22700 taka. The mean dowry in non-arsenic areas is around 20000 taka while it is much higher in arsenic areas (dowry amount is 22 percent higher in arsenic affected areas, the difference is statistically significant between the two groups). In Table 2.2 where we provide summary statistics about mean bride price before and after the intervention, we observe that mean bride price increased from 42611 taka in pre-treatment period to 66619 taka in non-arsenic sub-districts (an increase of 56 percent). However the arsenic affected areas saw an increase in bride price by only 30 percent [12].

[11] Since the survey took place from December 2010 to January 2011, so we only observe marriages till year 2010.

[12] We don't provide detailed year by year figures for this dataset as they are imprecisely estimated due to small sample size for each year.

2.4 Estimation Methodology

To estimate the effect of information campaign about arsenic contamination on age at marriage we follow a difference-in-difference approach (DID) using the following specification:

$$Y_{ijdt} = \alpha + \theta Arsenic_j * post_t + \beta X_{ijdt} + \gamma_j + \delta_t + \rho_d t + \varepsilon_{ijdt} \tag{2.1}$$

where Y_{ijdt} is the marriage market outcome - age at first marriage (for males or females) or log of bride price for an individual i who resides in sub-district j, belonging to district d and who belongs to marriage cohort t (i.e. who got married in year t). $Arsenic_j$ is the dummy variable which takes value 1 if the mean arsenic contamination level is greater than 50ug/litre for a sub-district and 0 otherwise, $post_t$ is the dummy variable which takes value 1 for post treatment period (2002-2011) and 0 for pre-treatment period (1990-2001). X_{ijdt} includes controls for individual characteristics which matter in the marriage market. Since they differ for the two outcomes, we describe them in detail below. To control for time-invariant differences between sub-district and location-invariant differences between marriage cohorts we include sub-district fixed effects, γ_j and marriage cohort fixed effects, δ_t. We also control for district level trends, $\rho_d t$ to account for any underlying trends present at the district level. These fixed effects and trends account for any omitted variables at the sub-district and year level and any linearly time varying factors at the district level. The errors have been clustered at the sub-district level for our analysis.

Our coefficient of interest is θ which captures the effect of information campaign. Essentially we compare difference in outcomes between marriage cohorts from arsenic affected sub-districts before and after the information campaign to the difference in outcomes in the same marriage cohorts from non-arsenic sub-districts. The individual level controls (X_{ijdt}) for age at marriage regressions based on Census data include religion being Islam, ownership of land, ownership of a house, literacy status and type of quality of house (that is whether it has cemented walls). For bride price regressions we rely on PKSF survey data and individual level controls here include education level of bride, difference in education level between bride and groom, age of bride, difference in age of bride and groom, a dummy for choosing the marital partner by themselves(rather than being chosen by family), a dummy for brides family being richer than groom' family and a dummy for groom's family being richer than bride' family (omitted category being two families belonging to same income class).

Identification

Our identifying assumption is that trends in age at marriage (and bride price) were not differential across arsenic and non-arsenic sub-districts, other than because of the information campaign. The district specific trends also help in establishing validity of our results, as our estimates are identified off deviations from district trends. The possible threats to our identification thus comes from sub-district level trends which might be correlated with arsenic contamination.

For each of our DID specifications we test whether the parallel trend assumption is satisfied by conducting the following estimation exercise:

$$Y_{ijdt} = \alpha + \Sigma_{2001}^{1990} \lambda_t (Arsenic_j * I_t) + \beta X_{ijdt} + \gamma_j + \delta_t + \rho_d t + \varepsilon_{ijdt} \tag{2.2}$$

where Y_{ijdt} is our outcome of interest: age at first marriage or bride price and j is the sub-district of residence. I_t is an indicator variable for each of the pre-treatment years. In presence of sub-district and year of marriage fixed effect, the interaction terms between year dummies and arsenic dummy reveal whether control and treatment group followed different trends over time. We look for individual significance of all these interaction terms. If these terms are individually insignificant then that reveals that the parallel trend assumption is satisfied as the two groups followed similar trends in the pre-treatment period.

We estimate specification 2 for all our outcome variables for the pre-treatment period and present our results in Figure 2.4 (panels a, b and c). We notice that for almost all our models we satisfy parallel trend assumption as all the interaction terms are insignificant. We also test for joint significance of interaction terms and reject it at 10 percent level for all our results (the only exception is the female sample where parallel trend assumption may not be satisfied as individual and joint significance of interaction terms (is rejected at 5% significance level) is statistically different from zero). In Table 2.3 we test a slightly different version of specification 2 to test whether the trends differed between arsenic and non-arsenic areas. We do this by replacing individual year interacted arsenic variables $(\Sigma_{2001}^{1990}(Arsenic_j * I_t)$ with a single variable for arsenic dummy interacted with a continuous trend variable $Arsenic{*}Trend$. An insignificant coefficient on this new variable reflects that trends were not different between the two groups before the intervention. We find in Table 2.3 (columns 1 to 3) that parallel trend assumption does hold true for all our outcome variables.

2.5 Results

2.5.1 Age at Marriage

In Table 2.4 we present results from specification 1. All columns include sub-district and marriage cohort (year of marriage) fixed effects. Columns 1 and 2 refer to results for the male sample while columns 3 and 4 present results for the female sample. Our main results in columns 2 and 4 control for district level trends. The coefficient of $Arsenic * Post$ shows that both males and females from arsenic affected areas are get married earlier after implementation of information campaign. For the male sample the reduction in age at marriage is 0.22 years (around 2.6 months), while for the female sample the magnitude is slightly lower, around 0.12 years (1.4 months)[13] These results translate into a reduction of age at marriage after information campaign by around 1 percent for males and by 0.67 percent for females. Results on other controls reveal that being literate increases the age at marriage, also religion of an individual being Islam is associated with getting married earlier. Ownership of assets is also associated with higher age at marriage.

2.5.2 Bride Price

Table 2.5 presents our results for bride price. Our dependent variable here is log of bride price, and thus coefficients of interaction variables can be interpreted in terms of the percentage

[13]The results on female's age at marriage should be interpreted with some caution as we don't observe them in their original place of residence, however assuming that in equilibrium marriage matches happen between similar areas, i.e. males from arsenic areas get matched to females from arsenic areas and males from non-arsenic areas get matched to females from non-arsenic areas.

decline in bride price for arsenic areas in the treatment period. Column 1 presents results with no district level trends and we find a 34% decrease in bride price (however this effect is not found to be significant). Column 2 accounts for district level trends and we find that arsenic areas witnessed a 64.4% decrease in bride price post information campaign. We also observe that bride price is positively associated with education level of the bride while it is negatively associated with groom's education level (not significant). Additionally getting married earlier is associated with lower bride price[14].

2.5.3 Robustness Checks

Matched sample analysis

We complement our findings from Census data by using additional demographic surveys. The arsenic affected sub-districts can be systematically different from non-arsenic sub-districts, so we additionally wanted to analyze whether arsenic affected areas which are *similar* to non-arsenic areas exhibit similar marriage patterns after information campaign. We conduct this analysis by matching arsenic affected sub-districts with unaffected sub-districts by using sub-district level characteristics from Integrated Public Microdata Series (IPUMS) Census Data for 2001. The IPUMS Survey is a huge census and for matching purposes we use pre-treatment characteristics which come from 2001 IPUMS survey which covered over 12 million individuals (2.6 million households) residing in all 64 districts of Bangladesh. This dataset was also compiled by Bangladesh Bureau of Statistics (BBS) and has individual and household information available at sub-district level. We use 19 such variables for our matching exercise. In particular, we use details about employment, education, household characteristics, sex-ratios for the unmarried population and sex ratio for children below one year of age, collapsed at sub-district level for our matching purpose[15]. Table A2.1 compares the mean 2001 characteristics of arsenic (treated) and non-arsenic (control) sub-districts. Column 1 and 2 reveal that these two groups are considerably different from each other. To address this concern

[14] Bride price results rely on a small small due to which many estimates have large standard errors because of which we observe insignificant results.

[15] We don't use IPUMS data for our main analysis as it doesn't contain information about our main dependent variable i.e. the age at which individuals got married.

we follow the matching procedure by using sub-district level characteristics from IPUMS data. Our matching exercise results in 79 arsenic affected sub-districts getting matched to unique 79 non-arsenic sub-districts, and column 6 in Table A2.1 shows that the matched treated and matched control sub-districts are statistically similar in terms of various demographic and non-demographic characteristics.[16]

Using these matched sub-districts, we conduct a similar analysis using Census data following specification 1. In Table 2.6 (columns 1 and 2), we observe that for both male and female sub-sample mean age at marriage reduced after information campaign. The magnitude is similar to our original results with age at marriage getting reduced for males and females by 0.94 percent. In columns 3 and 4, we use all sub-districts and additionally control for propensity score which represents an index value for base characteristics of sub-districts before the implementation of information campaign. We find that our estimates are almost identical to our original results.

Local contamination measures

Next, we conduct an analysis for age at first marriage for males using Bangladesh Demographic Health Surveys (BDHS). We turn to BDHS data as it provides rich spatial information about contamination of drinking water[17]. We use four rounds of these surveys conducted in 1999, 2004, 2007 and 2011. These surveys were conducted by National Institute of Population Research and Training (NIPORT) and followed identical questionnaires overtime. For each year around 10000 households were sampled from 350 clusters, except for year 2011 when 17000 households were sampled from 600 clusters. BDHS collected GPS information for all sampled clusters. For our rural sample we focus on the men's questionnaire which contains questions regarding their age at marriage, education and assets. Our final estimation sample has around 5000 men sampled from 1067 clusters. We use this dataset to replicate and complement the results which we find using Census data.

We are able to construct arsenic contamination figures for each individual sampled cluster

[16]The probit regression for generating propensity scores for matching is presented in Appendix Table A2.2 and Figure A2.1 in appendix plots the number of treated and control sub-districts matched over p-score. 410 sub-districts from IPUMs dataset are used for analysis. Caliper used for matching is 0.05.

[17]We don't use female sample here as we don't observe the original village from which she got married, we only observe females in the households they got married to.

in BDHS data. Arsenic contamination for a cluster was calculated by averaging contamination level for all tubewells lying in the 5 mile radius around the cluster location. This is a "local" level of contamination with clusters with mean contamination figure greater than 50ug/litre being classified as arsenic affected and those with less than 50ug/litre being classified as non-arsenic affected cluster. This local measure of contamination in BDHS data provides important information about contamination in the *neighbourhood* of a household which reflects the local visual clues (in form of red and green painted tubewells and people getting affected by arsenicosis symptoms) which people observe in their locality.

We now present our results using BDHS data which uses rich spatial information about arsenic contamination at cluster level. Table 2.7 presents results for males, both column 1 and 2 have cluster and marriage cohort fixed effects. Column 2 additionally controls for district level trends. Analysis using BDHS data shows that the our variable of interest, that is interaction term $Arsenic * Post$ is negative and significant in both the columns[18], with the order of magnitude being 0.84 years (10 months) which is a reduction in age of marriage by as much as 3.5 percent. This is much larger in comparison to the effect seen using census data[19]. We find similar results as before for our control variables as well, with education and religion significantly affecting age at marriage.

Marriage pool before and after information campaign

The results which we observe can potentially arise if the information campaign affected the marriage pool (that is mix of males and females in the marriageable age group). We explicitly test for this by looking at the sex ratio at sub-district level before and after the implementation of information campaign. We use IPUMS Census 2001 data and Census 2011 data to create sub-district level mean sex ratio between males and females in age group 15 to 25 years for years 2001 (before intervention) and 2011 (after intervention)[20]. Table 2.8 presents our results for this analysis, we observe (in columns 1 to 3 which correspond to different specifications) that sex ratio did not change due to information campaign. This essentially cements our

[18] Coefficient of interaction term in column 2 of Table 2.7 is significant at 6% level.

[19] In Appendix Figure A2.2 we provide evidence for parallel trend assumption for models discussed in sections 5.3.1 and 5.3.2.

[20] 410 sub-districts are matched between IPUMS Census 2001 data and Census 2011 data.

original theory that the effects we see in marriage market of information campaign are driven by behavioural factors rather than changes in marriage pool.

Shorter analysis window

Our original analysis was based on marriages which took place between 1990 to 2011, we shorten this analysis window to span from 1996 to 2006 to test whether our original results still hold. In Table 2.9 (columns 1 and 2), we observe that for the male sample estimate remains almost unchanged, while for the female sample although the estimate still retains the negative sign but the magnitude is much smaller and it is no-longer significant.

Placebo tests

Lastly, we conduct two placebo tests to establish validity of our results. First we replicate our analysis for urban sample where we don't expect to see any effect of the information campaign as the urban places don't rely on tubewells as a source of water (which were the focus of information campaign). In Table 2.9 (column 3 and 4), we observe that information campaign had no effect on age at marriage for both males and females which is in line with our expectation.

We also conduct another placebo test by randomly shuffling the arsenic contamination status of sub-districts. This basically randomly assigns sub-districts into new treatment and control groups. We then estimate the coefficient of (Arsenic * Post) variable by following our original specification. We repeat this exercise 1000 times, each time randomly assigning treatment and control status to sub-districts. This gives rise to a distribution of θ based on 1000 simulations. In Figure 2.5, we observe that for both male and female sample true estimate of θ lies beyond the 95% confidence interval which shows that the results we observe are not there by chance, which strengthens the validity of our results.

Other contamination measures and additional data on bride price

In our analysis we have categorized sub-districts into arsenic and non-arsenic sub-districts using $50ug/litre$ as the cut-off for dangerous level of arsenic in drinking water. We now provide additional results where alternate contamination variables are used for analysis. Table 2.10

(age at marriage results) and Table 2.11 (bride price results in columns 1 to 4) presents results based on four alternate measures of contamination. $MeanArsenic$ is a continuous measure of arsenic contamination while $Arsenic10$ is a dummy variable which takes value 1 if mean arsenic contamination in a sub-district is above $10ug/litre$ which is the WHO safety standard for arsenic in water. We use two other measures of contamination which reflect proportion of wells in a sub-district which are above two threshold values - $50ug/litre$ and $75ug/litre$. In Table 2.10 in columns 1 to 4 we present results for the male sample while in columns 5 to 8 we present results for the female sample, Table 2.11 (columns 1 to 4) provides results for bride price. We observe that for both age at marriage and bride price the effect of information campaign based on alternate measures of arsenic contamination is negative. The effect of information campaign using continuous measure of arsenic contamination is significant for both age at marriage (males and females) and bride price. When we use $10ug/litre$ cut-off for defining arsenic contamination then we observe that the point estimate (of $Arsenic10 * Post$) is negative for all outcome variables. The estimate for age at marriage for males is -0.18 which is close of the original estimate of -0.22, while the estimate for females is -0.07 (not significant). The corresponding bride price estimate using $10ug/litre$ cut-off is negative and smaller in magnitude (in comparison to results based on $50ug/lire$ cut-off) but not significant. The results corresponding to two other proportion related contamination measures reveal a similar scenario.

Our bride price analysis uses only PKSF survey till now. Since the sample is small for bride price analysis so we incorporate additional data on more marriages using Bangladesh Rural Urban Linkage (BRUL) Survey. This survey was conducted by International Food Policy Research Institution between December 2004 and January 2005 and it used identical modules on marriage, divorce, bride price and dowry. The combined sample of PKSF and BRULS for marriages between 1990 and 2010 contains 1699 marriages from 80 sub-districts. In Table 2.11 column 5 we replicate our original bride price result using this expanded dataset. We find that bride price reduced by 60.5% in arsenic affected areas after information campaign, which is close to our original estimate of 64.4% reduction in bride price post intervention.

2.5.4 Heterogeneity

The distribution of arsenic contamination is not uniform across all sub-districts in Bangladesh. There were few sub-districts which had very high mean level contamination that is arsenic level greater than $100 ug/litre$. We expect that the intensity information campaign could have been higher for these sub-districts as the arsenic contamination problem was extremely intense for these sub-districts. Hence the behaviour change expected in these sub-districts could also be different. We analyze whether the effect of information campaign was more pronounced in these sub-districts. We do this by comparing individuals from sub-districts with mean arsenic contamination less than $50/ug/litre$ (same as our old control group) to sub-districts with mean arsenic contamination greater than $100/ug/litre$ (new treatment group)[21]. In Table 2.12, we present our results based on this analysis. We find that for both males and females the magnitude of the effect of information campaign on age at marriage is larger than our original results, that is the age at marriage reduces by 1.2% and 1.3% for males and females respectively in response to information campaign.

2.6 Discussion and Conclusion

The patterns that we observe can be explained by the matching mechanism in marriage market. Suppose in the marriage market the matching is based on sorting which is driven by preferences over prospective match's traits like beauty, life-expectancy, fertility, health standard and income generating prospects. Lets assume that all of these traits can be subsumed in a single index value over which the sorting for partners takes place. A higher value of the index represents a more desirable partner. Now we know that for an individual belonging to an arsenic affected area the expected value of this index will take a lower value than the one for an individual who hails from a non-arsenic area. A higher index value is desired more in the market, hence in a stable match the highest ranked woman gets matched to the highest ranked man, the second highest ranked woman gets matched to second highest ranked man and so on. The Gale-Shapley algorithm (Gale and Shapley (1962)) will thus give an equilib-

[21] Parallel trend assumption is satisfied for both male and female sample based on similar analysis as presented in Table 2.3.

rium outcome where high ranked females from non-arsenic areas get matched to high ranked males from non-arsenic areas and low ranked females from arsenic areas get matched to low ranked males from arsenic areas. We call this arsenic-arsenic matching for the purpose of our discussion.

We hypothesize that the information campaign made the contamination information public, people now knew their own likelihood of being affected by arsenic poisoning symptoms. Also people from other areas (who are looking for a match) could take in the visual information in form of red and green painted tubewells and could ascertain probabilistically the chance of a prospective match being affected by arsenicosis symptoms. The information campaign thus complemented the index value in the marriage market. For an individual from an arsenic area given arsenic-arsenic matching and faced with a possibility of low life-expectancy and higher probability of developing skin lesions, the diminished prospects of finding a good match in marriage market leads him to get married earlier before his symptoms get discovered. Following a similar argument (for bride price), given arsenic-arsenic matching, the groom's family is aware about their matching with a female from arsenic affected area. Since the social structure is such that stigma attached to a female with skin lesions (and possible low fertility) is way more than the one attached to a male, hence females disproportionately suffer more than males in the marriage market. Thus concerns about prospective bride's beauty and possible low fertility coupled with asymmetrical costs associated with females developing skin lesions leads to an agreed bride price which is lower [22]. While this mechanism is consistent with our evidence, we are unable to provide evidence for the same in this paper.

To conclude, while the behavioural change induced by an information campaign has mainly been explored in the health domain (for example: increase in breastfeeding time and switch to safer sources of water for concerns related to health), we have gone a step further and shown that the information campaign had a spill-over effect in marriage market as well. Males and

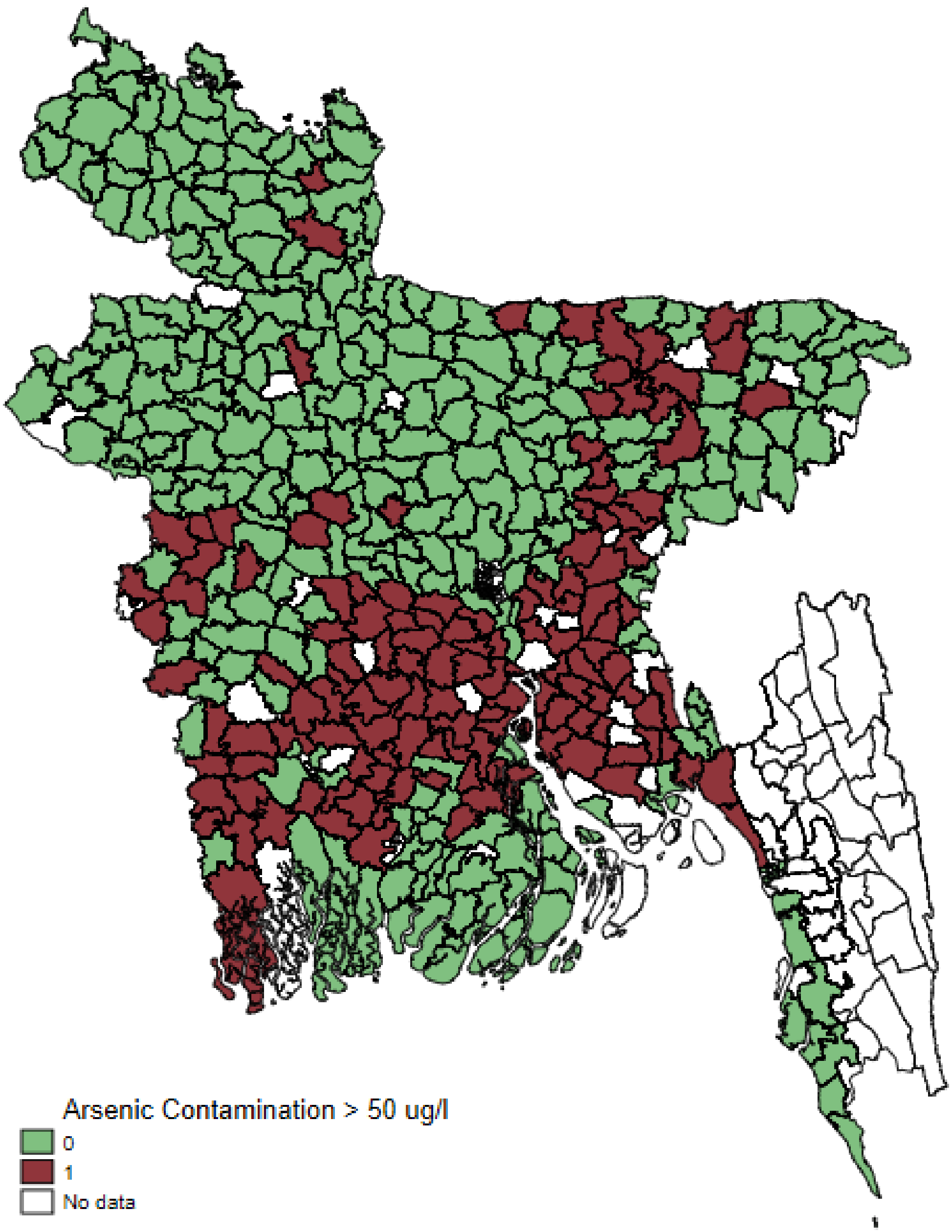
[22] Dowry is also an important aspect of marriages in Bangladesh, however the effect of information campaign on dowry can be quite ambiguous. Dowry unlike bride price is an amount which is exchanged at the time of marriage when the groom may not have started showing symptoms of arsenicosis. Given arsenic-arsenic matching the brides family know that they get matched to a groom from arsenic area with greater probability of being affected with arsenicosis which ideally drives down the dowry amount. However if we look at things from the supply side then the set of possible matches for a bride, is now also smaller since it now excludes non-arsenic grooms, this drives up the dowry amount. Hence we get an ambiguous result (insignificant result) for dowry.

females (and their respective households) in marriage market learned about their likelihood of being affected by arsenicosis symptoms in the future and reacted to this information by reducing their age at marriage. The bride price also suffered a dampening effect due to information campaign owing to information garnered about their adverse fertility and beauty outcomes in the future. The campaign owes its success in terms of generating awareness to its unique design. Constant visual reminders, multiple strategies to avoid contaminated water and public forums for information disbursal seem to have worked in Bangladesh. We find evidence that campaigns with such information, which informs people about adverse outcomes, may also lead to undesirable social outcomes, we demonstrate one such consequence in the marriage market.

Figures and Tables for Chapter 2

Tables and Figures

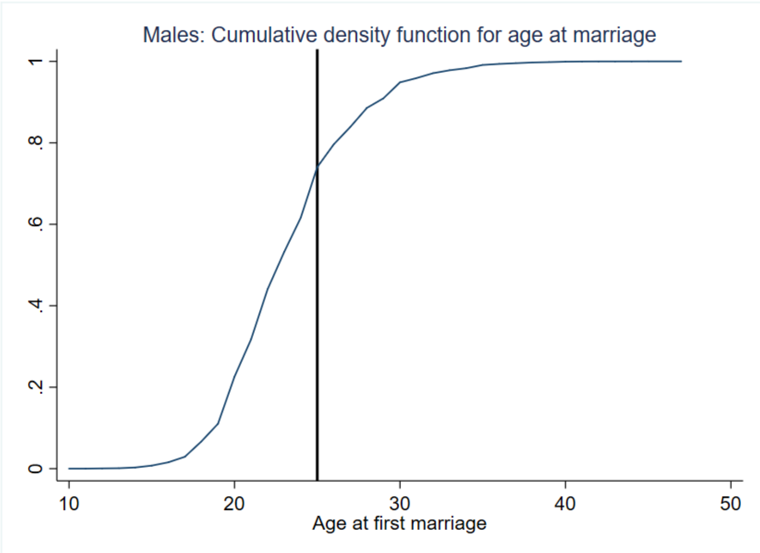
Figure 2.1: Arsenic Contamination Map for Bangladesh



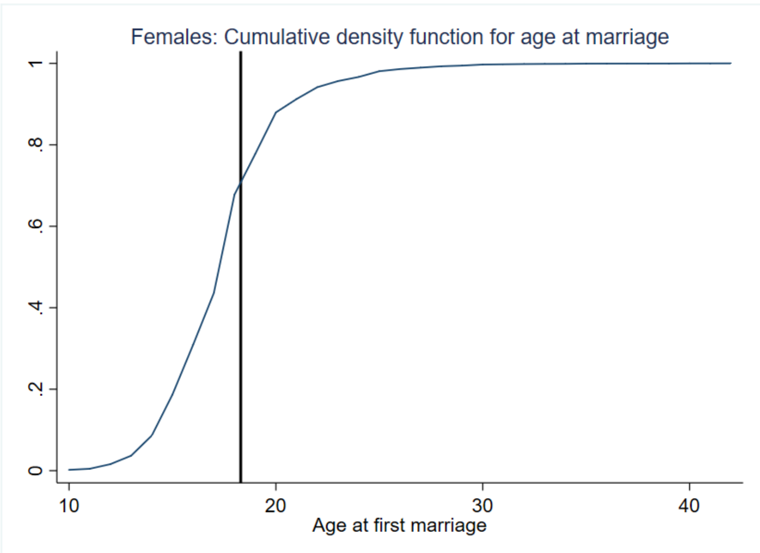
Source: Arsenic contamination data from British Geological Survey 1999

Figure 2.2: Cumulative Density Function for Age at Marriage for Males and Females

(a) Age at marriage for males



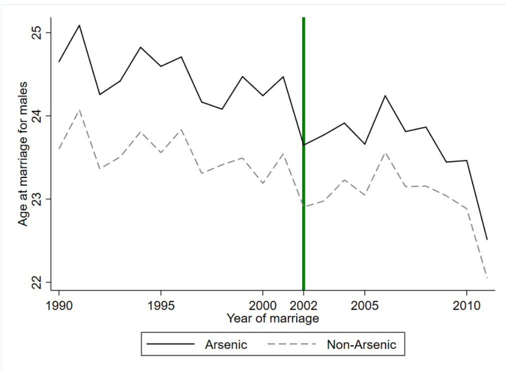
(b) Age at marriage for females



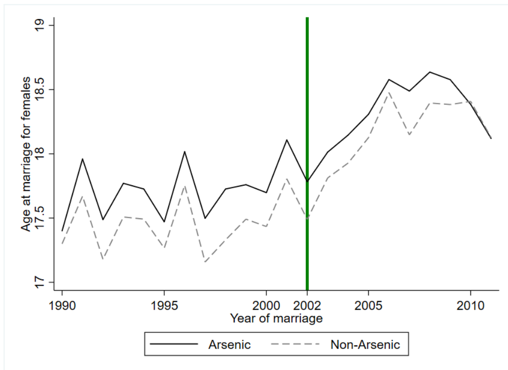
Source: Authors' calculations from Census Data 2011

Figure 2.3: Age at Marriage for Males and Females, Differences in Age of Marriage between Arsenic and Non-arsenic Areas Over time

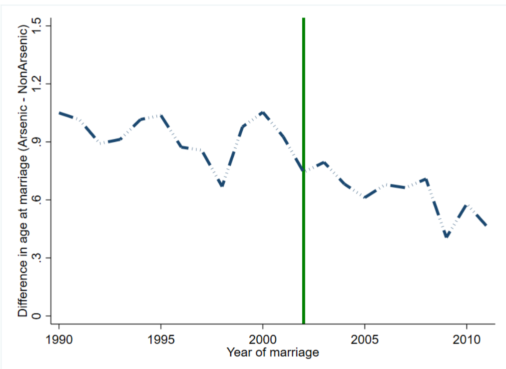
(a) Age at marriage for Males



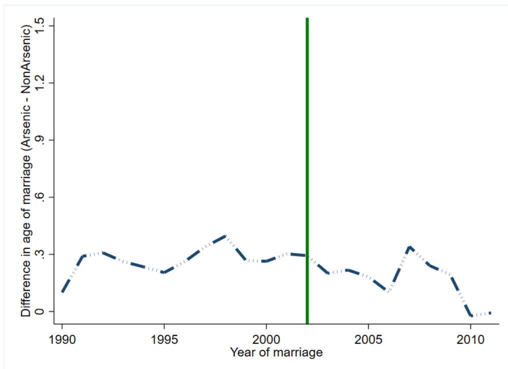
(b) Age at marriage for Females



(c) Difference in Age at marriage for Males



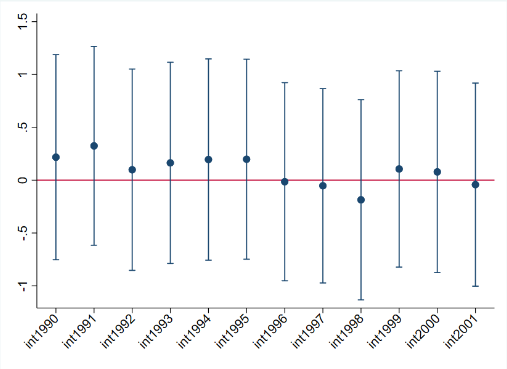
(d) Difference in Age at marriage for Females



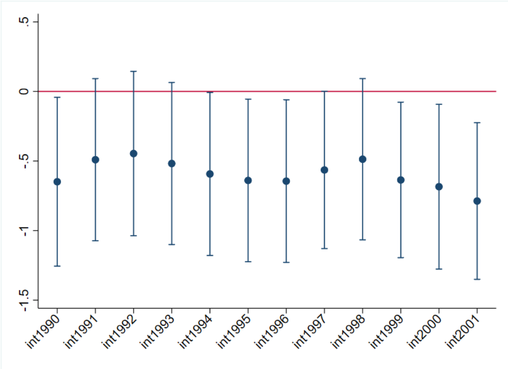
Source: Authors' calculations from Census Data 2011

Figure 2.4: Testing Parallel Trend Assumption

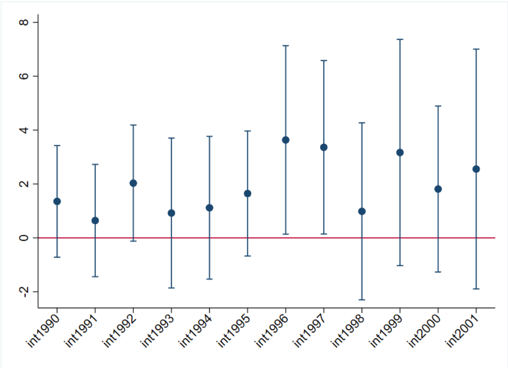
(a) Age at marriage for males



(b) Age at marriage for females



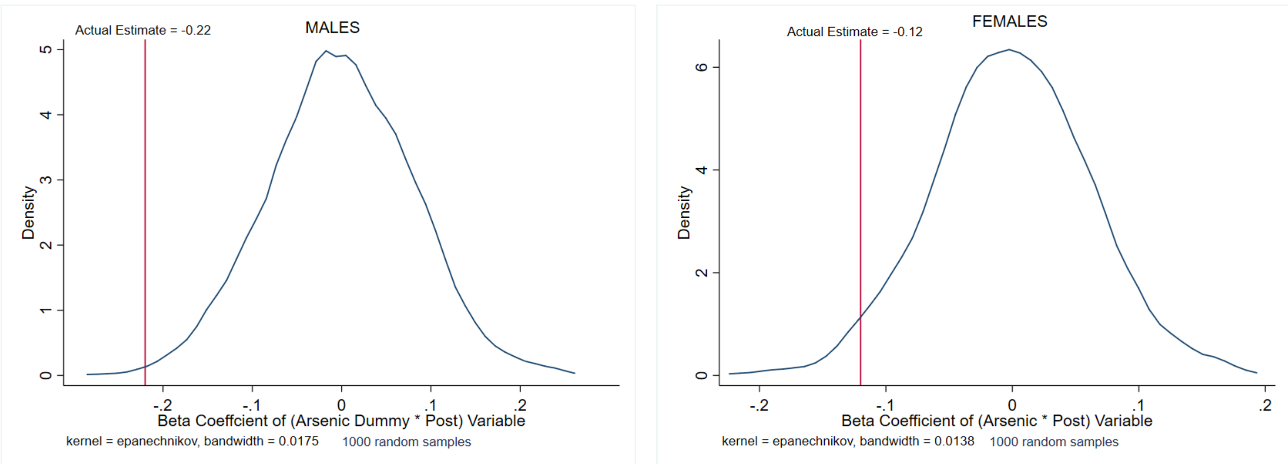
(c) Bride Price



Notes : Testing parallel trends assumption using specification 2: Plotting coefficients of interaction between year dummies and arsenic dummy for all pre-treatment years. Insignificance of coefficients points towards a similar trend being followed by the two groups in the pre-treatment period. The error bars are 95% confidence intervals, clustering at sub-district level.

Figure 2.5: Placebo Test

(a) Males (b) Females



Notes : Distribution of θ (specification 1) using 1000 simulations which randomly assign arsenic contamination status of 1 (treatment) or 0 (control) to sub-districts.

Table 2.1: Summary Statistics

PANEL A. MALES

	All		*Non-Arsenic*		*Arsenic*	
variablename	Mean	SE	Mean	SE	Mean	SE
Literate (%)	0.51	0.002	0.5	0.002	0.54	0.003
Religion is Islam (%)	0.89	0.001	0.9	0.001	0.87	0.002
Owns Land (%)	0.92	0.001	0.91	0.001	0.94	0.001
Owns House (%)	0.89	0.001	0.88	0.001	0.91	0.002
House has cemeneted walls (%)	0.16	0.001	0.16	0.002	0.16	0.002
Age at marriage (years)	23.61	0.015	23.34	0.018	24.14	0.026
Observations	75243		50281		24962	

PANEL B. FEMALES

	All		*Non-Arsenic*		*Arsenic*	
	Mean	SE	Mean	SE	Mean	SE
Literate (%)	0.56	0.002	0.55	0.002	0.6	0.003
Religion is Islam (%)	0.89	0.001	0.9	0.001	0.88	0.002
Owns Land (%)	0.93	0.001	0.91	0.001	0.95	0.001
Owns House (%)	0.89	0.001	0.88	0.001	0.91	0.002
House has cemeneted walls (%)	0.17	0.001	0.17	0.002	0.16	0.002
Age at marriage (years)	17.88	0.01	17.8	0.012	18.03	0.017
Observations	86965		57243		29742	

PANEL C. PKSF Data

	All		*Non-Arsenic*		*Arsenic*	
	Mean	SE	Mean	SE	Mean	SE
Education bride (in years)	5.31	0.13	4.98	0.19	5.56	0.17
Education groom (in years)	5	0.14	4.74	0.21	5.19	0.18
Bride's age at the time of marriage	16.75	0.1	16.88	0.16	16.64	0.12
Groom's age at the time of marriage	23.59	0.15	23.59	0.22	23.6	0.2
Partner chosen by bride	0.08	0.01	0.08	0.01	0.08	0.01
Brides family richer	0.32	0.02	0.33	0.02	0.32	0.02
Groom's family richer	0.25	0.01	0.21	0.02	0.28	0.02
Mehr (in Taka)	53366	2230	53980	3236	52900	3060
Dowry (in Taka)	22709	1034	20168	1291	24642	1528
Observations	875		378		497	

Table 2.2: Average Bride Price

	Pre-2002	**Post-2002**	**Observations**
Arsenic	45952	59600	497
	(4013)	(4569)	
NonArsenic	42611	66619	378
	(4226)	(4801)	

Source: PKSF Data. All figures in Taka.

Table 2.3: Parallel Trend Assumption for Difference-in-Difference

	Males ATM (1)	**Females ATM** (2)	**Bride Price** (3)
Arsenic * Trend	-0.029	-0.019	0.12
	(0.020)	(0.012)	(0.13)
Observations	40288	45855	443
Other HH & Individual Controls	Yes	Yes	Yes
Year of Marriage FE	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes
District trends	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other individual level controls as mentioned in the Table 2.4 (for age at marriage) and Table 2.5 (for bride price). ATM refers to Age at Marriage.

Table 2.4: Effect of Information Campaign on Age at Marriage

	Males		**Females**	
	(1)	(2)	(3)	(4)
Arsenic * Post	-0.25***	-0.22**	-0.073	-0.12*
	(0.086)	(0.085)	(0.065)	(0.066)
Literate	0.30***	0.29***	-0.13***	-0.15***
	(0.041)	(0.041)	(0.027)	(0.027)
Religion is Islam	-1.02***	-1.02***	-0.64***	-0.64***
	(0.10)	(0.10)	(0.063)	(0.063)
Owns land	0.36***	0.36***	0.26***	0.27***
	(0.088)	(0.088)	(0.061)	(0.060)
Owns house	0.22***	0.22***	0.17***	0.17***
	(0.074)	(0.074)	(0.062)	(0.061)
House has cemented walls	0.89***	0.89***	0.42***	0.42***
	(0.054)	(0.053)	(0.035)	(0.034)
Observations	75243	75243	86985	86985
Control Mean Age at Marriage	23.34	23.34	17.79	17.79
Year of Marriage FE	Yes	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes	Yes
District trends	No	Yes	No	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Table 2.5: Effect of Information Campaign on Log of Bride Price

	(1)	(2)
Arsenic * Post	-0.340	-0.644*
	(0.453)	(0.368)
Education bride (in years)	0.108***	0.108***
	(0.0235)	(0.0242)
Education groom (in years)	-0.0485*	-0.0361
	(0.0239)	(0.0215)
Bride's age at the time of marriage	-0.0453	-0.0532
	(0.0425)	(0.0407)
Groom's age at the time of marriage	-0.0116	-0.0139
	(0.0269)	(0.0276)
Partner chosen by bride	0.158	0.106
	(0.305)	(0.310)
Brides family richer	0.123	0.144
	(0.161)	(0.158)
Groom's family richer	0.239	0.266
	(0.167)	(0.158)
Observations	875	875
Control Mean Bride Price (in Taka)	53980	
Year of Marriage FE	Yes	Yes
Sub-District FE	Yes	Yes
District Trends	No	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Table 2.6: Robustness Check 1:
Matched Sample Analysis for Age at Marriage

	Males (1)	**Females** (2)	**Males** (3)	**Females** (4)
Arsenic*Post	-0.22*	-0.17*	-0.22***	-0.13*
	(0.13)	(0.090)	(0.085)	(0.067)
Literate	0.38***	-0.074*	0.30***	-0.14***
	(0.056)	(0.040)	(0.041)	(0.028)
Religion is Islam	-1.38***	-0.68***	-1.02***	-0.65***
	(0.19)	(0.10)	(0.10)	(0.063)
Owns land	0.16	0.27***	0.35***	0.27***
	(0.14)	(0.088)	(0.089)	(0.063)
Owns house	0.31***	0.22**	0.21***	0.15**
	(0.11)	(0.096)	(0.076)	(0.062)
House has cemented walls	0.73***	0.36***	0.88***	0.43***
	(0.080)	(0.044)	(0.054)	(0.034)
Propensity score			-0.66	-5.85*
			(2.99)	(3.29)
Observations	28991	33984	73479	85053
Number of subdistricts	158	158	410	410
Mean ATM in Non-Arsenic subdistricts	23.53	17.92	23.31	17.79
Year of Marriage FE	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes
District trends	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Table 2.7: Robustness Check 2:
Using Alternate Dataset - BDHS (for males only)

	(1)	(2)
Arsenic * Post	-0.80**	-0.84*
	(0.40)	(0.46)
Not Educated	-0.34*	-0.34*
	(0.19)	(0.20)
Religion is Islam	-1.82***	-1.87***
	(0.35)	(0.35)
Electricity connection	0.58**	0.58**
	(0.23)	(0.24)
HH has cemented walls	1.44***	1.42***
	(0.27)	(0.27)
Owns telephone	-0.71	-0.61
	(1.00)	(1.00)
Owns tv	-0.082	-0.072
	(0.24)	(0.24)
Owns bicycle	0.36*	0.36*
	(0.20)	(0.21)
Observations	4947	4947
Number of Clusters	1067	1067
Control Mean Age at Marriage	23.61	23.61
Year of Marriage FE	Yes	Yes
Cluster FE	Yes	Yes
District trends	No	Yes

Note: Standard errors in parentheses are clustered by cluster-year. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Table 2.8: Robustness Check 3:
Sex Ratio Changes between 2001 and 2011

	(1)	(2)	(3)
Treatment*Post	0.031	0.031	0.021
	(0.022)	(0.021)	(0.034)
Post	0.149***	0.149***	0.000
	(0.011)	(0.011)	(.)
Treatment	-0.021		-0.016
	(0.013)		(0.010)
Constant	0.811***	0.804***	0.886***
	(0.006)	(0.005)	(0.007)
Observation	820	820	820
District FE	Yes	No	No
District*Year FE	No	No	Yes
Subdistrict FE	No	Yes	No

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Table 2.9: Robustness Check 4:
Shorter Analysis Window and Placebo Test Using Urban Sample

	Analysis window 1996 to 2006		**Urban sample**	
	Males	Females	Males	Females
	(1)	(2)	(3)	(4)
Arsenic * Post	-0.23**	-0.023	-0.042	-0.049
	(0.11)	(0.091)	(0.22)	(0.14)
Literate	0.31***	-0.15***	0.78***	0.21*
	(0.049)	(0.036)	(0.10)	(0.11)
Religion is Islam	-1.09***	-0.66***	-1.40***	-1.09***
	(0.11)	(0.067)	(0.21)	(0.20)
Owns land	0.47***	0.34***	0.67***	0.40***
	(0.11)	(0.071)	(0.16)	(0.14)
Owns house	0.18*	0.14**	0.075	0.10
	(0.092)	(0.068)	(0.14)	(0.11)
House has cemented walls	0.95***	0.48***	1.49***	0.99***
	(0.067)	(0.045)	(0.18)	(0.093)
Observations	40887	47162	14856	16742
Control Mean Age at Marriage	23.34	17.74	24.85	18.49
Year of Marriage FE	Yes	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes	Yes
District trends	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the Table 2.4.

Table 2.10: Robustness Check 5A: Age at Marriage - Other Contamination Measures

	Males				**Females**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Arsenic * Post	-0.0015**				-0.00077*			
	(0.00060)				(0.00042)			
Arsenic10 * Post		-0.18**				-0.077		
		(0.086)				(0.069)		
Proportion>50 * Post			-0.0039**				-0.0012	
			(0.0016)				(0.0011)	
Proportion>75 * Post				-0.0042**				-0.0028**
				(0.0017)				(0.0012)
Observations	75243	75243	75243	75243	86984	86984	86984	86984
Year of Marriage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the Table 2.4.

Table 2.11: Robustness Check 5B: Bride Price
Other contamination measures and Additional Data

	PKSF Data				**PKSF+BRULS**
	(1)	(2)	(3)	(4)	(5)
Mean Arsenic * Post	-0.007**				
	(0.003)				
Arsenic10 * Post		-0.097			
		(0.266)			
Proportion>50 * Post			-0.009$^{\phi}$		
			(0.006)		
Proportion>75 * Post				-0.016**	
				(0.006)	
Arsenic50 * Post					-0.605*
					(0.312)
Observations	875	875	875	875	1699
Year of Marriage FE	Yes	Yes	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes	Yes	Yes
District trends	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by sub-district. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$, * is $p < 0.1$ & ϕ is significance at 15.5 percent. Regression includes a constant term and other controls as mentioned in the Table 2.5.

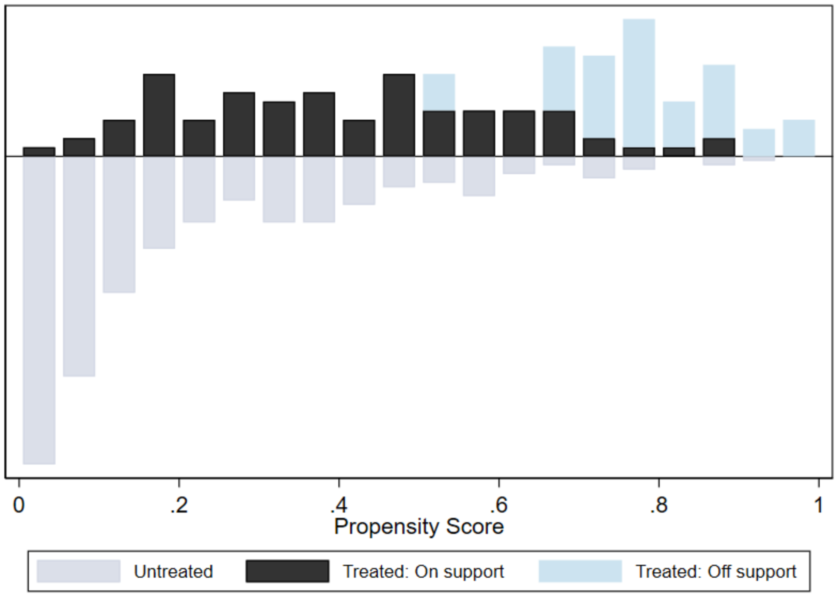
Table 2.12: Heterogeneity

	Males	**Females**
	(1)	(2)
Arsenic100 * Post	-0.28**	-0.23**
	(0.12)	(0.093)
Literate	0.29***	-0.15***
	(0.043)	(0.029)
Religion is Islam	-1.01***	-0.62***
	(0.12)	(0.069)
Owns land	0.35***	0.26***
	(0.097)	(0.066)
Owns house	0.23***	0.14**
	(0.081)	(0.067)
House has cemented walls	0.92***	0.43***
	(0.060)	(0.038)
Observations	63600	73351
Control Mean Age at Marriage	23.34	17.79
Year of Marriage FE	Yes	Yes
Subdistrict FE	Yes	Yes
District trends	Yes	Yes

Note: Treatment Group = Arsenic Contamination > 100; Control Group = Arsenic Contamination < 50. Standard errors in parentheses are clustered by subdistrict. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Appendix

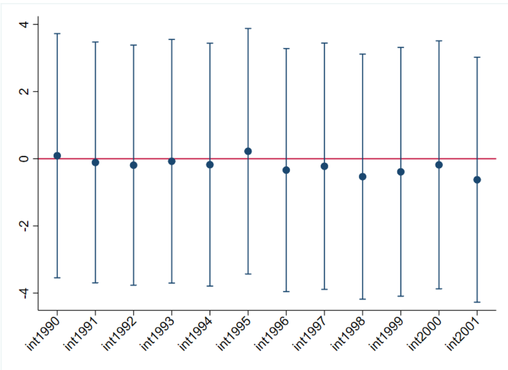
Figure A2.1: Density of Sub-districts Over Propensity Score



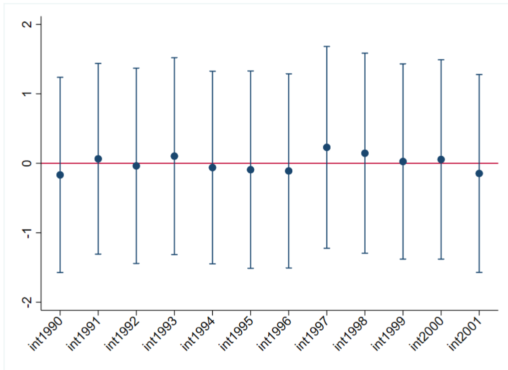
Notes : The matching is done using psmatch2 command in stata, with caliper value 0.05 and unique matching strategy with no replacement.

Figure A2.2: Testing Parallel Trend Assumption for Robustness Checks

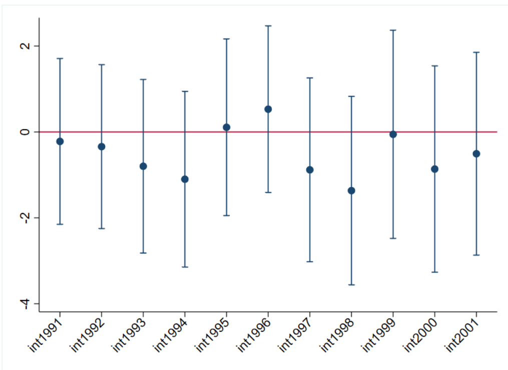
(a) Age at marriage for males (Matched sample)



(b) Age at marriage for females (Matched sample)



(c) Age at marriage for males (BDHS)



Notes : Testing parallel trends assumption using specification 2: Plotting coefficients of interaction between year dummies and arsenic dummy for all pre-treatment years. Insignificance of coefficients points towards a similar trend being followed by the two groups in the pre-treatment period. The error bars are 95% confidence intervals, clustering at sub-district level.

Table A2.1: Matching - Summary of Sub-district Characteristics

	1	2	3	4	5	6
SUBDISTRICT LEVEL MEANS	**Control**	**Treated**	**t-stat ((1) vs (2))**	**Matched Control**	**Matched Treated**	**t-stat ((4) vs (5))**
employment category						
employed in agriculture (%)	0.56	0.63	-5.13	0.58	0.58	-0.01
employed in formal sector (%)	0.16	0.14	2.99	0.16	0.16	0.17
employed in business (%)	0.13	0.11	5.93	0.13	0.12	0.28
employed in others (%)	0.15	0.12	4.22	0.14	0.14	-0.30
education details						
literacy level (%)	0.44	0.40	4.02	0.43	0.43	-0.07
number of years of education	3.30	2.96	4.29	3.19	3.20	-0.08
completed primary education (%)	0.91	0.91	-1.41	0.91	0.91	-0.12
employment status details						
employed (%)	0.42	0.43	-4.50	0.43	0.43	0.84
unemployed (%)	0.02	0.02	3.18	0.02	0.02	-0.65
inactive (%)	0.15	0.14	3.15	0.15	0.15	0.39
involved in housework (%)	0.40	0.40	0.24	0.40	0.41	-1.42
household characteristics						
number of children	1.67	1.61	2.98	1.63	1.63	0.06
number of families	1.45	1.39	5.21	1.44	1.45	-0.36
electricity connection (%)	0.24	0.18	4.55	0.22	0.22	-0.10
ownership of house (%)	0.95	0.94	2.44	0.95	0.95	-0.71
religion is muslim (%)	0.87	0.89	-2.39	0.87	0.89	-1.18
other demographic characteristics						
ratio of males to females for adults	0.49	0.50	-5.44	0.50	0.49	1.65
ratio of males in children with age <1 year	0.52	0.52	-0.84	0.52	0.52	1.12
Number of Sub-districts	277	133		79	79	

*Source - IPUMS data for 2001

Table A2.2: Probit Estimation for Propensity Score Matching

Sub-district level means	**Arsenic Contamination = 1 or 0**
employment category	
employed in agriculture	0.11
	(0.44)
employed in formal sector	0.11
	(0.64)
employed in business	2.05**
	(0.88)
employed in others	-
education details	
literate	-1.88**
	(0.80)
number of years of education	0.39***
	(0.14)
completed primary education	6.55***
	(1.57)
employment status details	
employed	1.40
	(1.45)
unemployed	-2.83
	(5.85)
inactive	0.16
	(0.86)
involved in housework	-
household characteristics	
number of children	0.23**
	(0.11)
number of families	-0.080
	(0.27)
have electricity connection	0.77***
	(0.19)
ownership of house	2.12***
	(0.64)
religion is Islam	-0.93***
	(0.23)
No toilet facility	-0.59***
	(0.16)
other demographic characteristics	
ratio of males to females in adults >15 years	-7.58***
	(1.81)
ratio of males in children with age <1 year	-0.92
	(0.94)
Psuedo R-square	0.3089
Observations (number of subdistricts)	410

Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regression includes a constant term and other controls as mentioned in the table.

Chapter 3

Early Life Exposure to Outdoor Air Pollution: Effect on Child Health in India

3.1 Introduction

Pollution in any form, whether it be air or water, poses an environmental risk to the health of the exposed population. According to WHO global air pollution database, out of the 15 most polluted cities in the world, 14 belong to India. Another recently published report by Health Effects Institute on air pollution in India (2018) reports that air pollution was responsible for 1.1 million deaths in India in 2015. In the absence of effective pollution regulatory policies, air pollution levels have reached alarming levels in various parts of India (Greenpeace, 2017). This warrants a closer look at the air pollution problem from the standpoint of welfare of the younger generation currently being exposed to harmful pollutants with possible long-lasting effect on their health. This article aims at estimating the effect of in-utero exposure to air pollution on child growth indicators, using exogenous changes in biomass burning events which contribute to the air pollution.

Recent studies on India which focus on air pollution and child health focus on pollution and its impact for cities in India (Greenstone and Hanna, 2014). In this paper we conduct a pan-India analysis relying on rich geo-spatial information on air pollution to study its effect on children's growth indicators. In particular, we study the effect of early life exposure to air pollution (as measured by PM 2.5) on children's weight and height measures for children under age five. The rich geo-spatial information on pollution comes from satellite data on aerosol optical depth which has been converted into gridded PM2.5 data (Dey et al., 2012; van Donkelaar et al., 2010). We match the gridded PM2.5 data to GPS locations of sampled clusters in Demographic Health Survey (DHS, 2015-16 round for India) to produce rich geo-spatial information about local (75 km radius) pollution levels in the place and time of conception(residence) of a child.

In an empirical exercise which causally links child health to local pollution levels, household income and behavioural choices are omitted variables which make local pollution levels endogenous. To address this concern, we use temporal changes in pollution causing biomass burning events like crop-burning and forest fires around the sampled clusters. In particular, to alleviate concerns of endogeneity, we only use such fire events that are *upwind* as an instrument for local pollution levels. This strategy critically relies on the assumption that exogenous changes in wind direction are not associated with household's income or behavioural choices. Multiple studies (Rangel and Vogl, 2018; Pullabhotla, 2018) have shown that these wind changes impact local pollution levels. The literature linking air pollution to child health has mostly focused on child mortality. In this paper we show that air pollution affects child's growth indicators even if she survives. To the best of our knowledge, this is the first study for India which addresses the endogeneity issues present while studying the link between children's growth indicators with local pollution levels.

Our analysis shows that air pollution negatively affects children's health. Exposure to air pollution during the first trimester decreases both Height-for-age (stunting measure) and Weight-for-age (underweight measure) for children aged below five years. A standard deviation change in PM2.5 is associated with 7.9% decrease in Height-for-age and a 6.7% decrease Weight-for-age measure. The effect is prominent for poorer households, with Northern states being more vulnerable due to high pollution levels in the area. These results are especially important given the link between stunting and other human capital outcomes. Early life stunting leads to irreversible damage, is associated with shorter adult height, lower cognitive ability, lower educational attainment, reduced adult income, and decreased off spring birth weight (Victora et al., 2008; Mendez and Adair, 1999).

The paper follows the following structure. The next section provides a literature overview of the effect of pollution on child health and highlights the contribution of this paper to the literature. Section 3 describes the various datasets that we use in our analysis. The next section presents the empirical methodology that we follow and is followed by results in Section 5. Section 6 concludes with an estimate of the extent of the problem and discusses current state of policies regarding air pollution in India.

3.2 Previous Literature

Our work is motivated by the "fetal origins" hypothesis (Douglas and Currie, 2011), which states that the *in-utero* period of a child critically determines mortality outcomes, disease prevalence and future health outcomes, abilities and earnings. Fetal growth, if restricted, can negatively affect future outcomes. The biological link between air pollution and fetal growth has not been documented in the literature, but it is mediated by placental growth which determines supply of oxygen and nutrients to the fetus. Exposure to pollution would affect placental function which can be impacted by inflammation caused by maternal infection. Additionally, pollution is known to cause epigenetic changes (interaction between our genes and environment which can cause DNA methylation, which regulates gene expression) which could affect fetal growth as well (Rangel and Vogl, 2018). A recent paper by Chakrabarti et. al (2019) shows how exposure to biomass burning (which causes pollution) affects respiratory health in adults as well as children. Hence this suggests that mothers can possibly be affected during the pregnancy time due to exposure to pollution which can potentially affect fetal growth as well.

In the economics literature, the intrauterine period has been the focus of many studies which have established links between occurrence of early life shocks to multiple outcomes. Early life shocks studied in economics literature include incidence of a) disastrous events (like famines, war, drought); b) nutritional shocks (like introduction of iodised salt, pregnancy during Ramadan) and c) pollution (air or water). Currie and Vogl (2013) provide a review of these early life shocks (a and b) on various outcomes; broadly summarised, these shocks negatively affect adult cognition, years of schooling, literacy status, adult height and stunting measures; and increase the likelihood of presence of birth defects, prevalence of heart disease and obesity. Maitra et al. (2018) study on China explores the link between early life (season of birth) conditions and its effect on child height, cognitive and non-cognitive ability of children.

The focus of our study is in-utero exposure to air pollution and Currie et. al (2014) reviews landmark studies which have been conducted in this area. Most of these studies are from developed nations with a few exceptions. Similar to previous studies, a major part of the literature focuses on learning outcomes (test-scores) and earnings which are negatively affected due to in-utero exposure to pollution (Bharadwaj et al., 2013; Isen et al., 2013 & Sanders, 2012). We extend this literature by looking at the link between early life exposure to pollution and stunting and underweight measures.

The strand of literature which is most relevant for our study has mainly looked at the effect of in-utero or early life exposure to air-pollution on infant mortality and birth weight. Some of these papers use natural experiments to causally identify the effect of air pollution on infant survival, for example, Chay and Greenstone (2003a and 2003b) use introduction of environmental regulations under Clean Air Act, 1970 and recession in 1981-82 in United States to show that reduction in pollution levels led to reduction in infant mortality. Currie and Walker (2011) shows that introduction of congestion-reducing automated toll payment systems in United States (which reduced number of idle vehicles emitting harmful pollutants) reduced pre-mature and low birth-weight births. Currie and Neidell (2005) use spatial and temporal variation in CO levels to analyse the effect of CO levels on infant mortality. Most of the studies in this domain are from developed nations where availability of high resolution pollution data is not a constraint. We focus on a developing nation which has much higher pollution levels in comparison to developed nations. Lack of data on pollution for developing nations has been a major limitation in the past but with availability of rich spatial information on pollution from satellite data, we link local exposure to air pollution with child's growth factors.

In a developing country context, the paper by Greenstone and Hanna (2014) analyses the effect of water and air pollution regulation policies on infant mortality in India. Foster et al. (2009) uses Mexico's clean industry certification program to study its effect on pollution (we use a similar measure of pollution i.e. satellite data on Aerosol Optical Depth to infer PM2.5 levels) and resulting respiratory related infant deaths. Wildfires and their negative health effects (like increase in infant mortality, reported asthma cases, pre-term births etc) have also been studied in context of Indonesian wildfire of 1997 (Jayachandran, 2009; Rukumnuaykit, 2003; Kunii et al., 2002; Frankenberg et al., 2005 and Barber and James, 2000), California wildfires (Holstius et al.,2012) and Australian wildfires (O'Donnell and Behie, 2015). A few recent papers assess the effect of in-utero exposure to biomass burning events and pollution: Rangel and Vogl (2018), Pullabhotla (2018) and Soo and Pattnayak (2019) study impacts on birth weight, infant mortality and long-term health outcomes like adult height respectively. These papers come closest to our paper as we also explore the link between in-utero exposure to pollution and child health. However our focus is somewhat different from most of the afore mentioned papers[1], we focus on solving the endogeneity problem in our paper rather than focusing on reduced form effect of biomass burning events on child health and we look at post-natal growth instead of survival. Another recent study on Bangladesh (Goyal and Canning, 2017) provides evidence for in-utero exposure to air pollution and increased risk of stunting, underweight and wasting but it doesn't address the endogeneity issues related to local pollution levels.

Our paper also adds to the growing literature of the effect of pollution on child health in India. These effects have been demonstrated by two recent papers on effects of water pollution in India. Brainerd and Menon (2014) have focused on use of fertilisers in India during crop sowing season which increases concentration of harmful chemicals in water. They find that exposure to these pollutants during the month of conception increases infant mortality and reduces Height-for-age and Weight-for-age for children. Do et al. (2018) have shown that regulations targeting industrial pollution in the Ganga River led to reduction in water pollution levels and infant death.

3.3 Data

3.3.1 Demographic Data

The demographic dataset used in this paper is sourced from the Demographic and Health Survey (DHS Round-IV for 2015-16) for India. DHS-IV contains detailed information about

[1]Soo and Pattnayak (2019)'s paper is closest to us in terms of its causal focus.

birth history of each woman who was interviewed. This survey sampled 601,509 households and interviewed 0.7 million eligible [2] women in the age group 15-49. Further, anthropometric measures of health were collected for 0.22 million children of ages five years and below. The DHS sample is a stratified two-stage sample and the primary sampling units (PSUs or clusters) correspond to villages in rural areas and blocks in urban areas. The DHS-IV comprises of around 28526 clusters with GIS information on almost all clusters[3]. To hide the identity of the village (block in urban areas), all clusters were displaced by five kilometres (two kilometres for urban clusters), with one percent of the clusters being displaced by as much as 10 kilometres. We account for this displacement when we discuss our identification strategy in the next section.

Our focus is on in-utero exposure to pollution for which we need the location and time of conception. To measure in-utero exposure to pollution we use the birth history of every child ever born to a woman. We use the location of the cluster, birth date and pregnancy duration of a child to impute exposure to pollution during the first trimester[4]. We make an important assumption that the place of stay of the mother when the child was in-utero is the same as the current residence of a child[5].

We measure impact of air pollution on child health by using anthropometric measures: Height-for-age and Weight-for-age (WHO standard z-scores) for children aged five and below. In addition all other demographic and household level variables which are used in our analysis are sourced from the DHS. We provide summary statistics for the sample we use for our analysis in Table 3.1. 52 percent of children in our estimation sample [6] are males with mean age around 29 months (2.5 years old) and the mean birth order of children is 2.2. The average age at which mothers have children is 24.5 years. Mothers had on an average 6.2 years of education. Three-fourth of our sample consists of rural households and a similar proportion of households report their religion to be Hindu. Marginalised groups (which include schedule

[2]Eligible women - married or unmarried women of reproductive ages.

[3]131 clusters have no GIS information.

[4]We also construct separate measures of exposures to pollution for other trimesters and first three months after birth.

[5]This assumption is a standard one which is employed by many papers which have used DHS data for analysis (for example, Brainerd and Menon, 2014). In our sample the mean number of years for which the interviewed family has stayed at the place of residence is around 15 years.

[6]Details about estimation sample are discussed in the section on pollution in section 3.2.

caste and schedule tribes) account for 37% of our sample. The mean household size for our estimation sample is 6.5. 88 percent of the households are headed by a male member and the average age of household head is 44.5 years. 85 percent of the households have an electricity connection, but only 23 percent of the households use piped water as their source of drinking water; 28 percent of our sample uses clean source of cooking fuel like LPG or bio-gas and the mean open defecation rate in a cluster is 43 percent.

The mean Height-for-age (HFA) and Weight-for-age (WFA) z-score for the sample of children used in this study is -1.46 and -1.52 respectively (mean weight-for-height is -0.97 for our sample). Height-for-age is a measure of stunting and it represents the effect of early life shocks that a child receives. Stunting generally occurs before age two and its effects are largely irreversible. It is associated with an underdeveloped brain, with long-lasting harmful consequences, including diminished mental ability and learning capacity, poor school performance in childhood, reduced earnings and increased risks of nutrition-related chronic diseases such as diabetes, hypertension, and obesity in future. Weight-for-age (underweight measure) reflects body mass relative to chronological age. It is influenced by both the height of the child (height-for-age) and his or her weight (weight-for-height). Deaton and Dreze (2009) advocate the use of Weight-for-age as the health status indicator for children as its a comprehensive measure which captures both stunting and wasting.

3.3.2 Pollution Data

In India, ground-based pollution measurement started post 2009 under the National Ambient Air Quality monitoring program maintained by Central Pollution Control Board. The network has slowly expanded to around 90 sites across 35 cities over the years, which leaves majority of India unmonitored[7]. Amongst these cities, only Delhi has greater than 20 monitoring sites while most other cities have a single monitoring site. Furthermore, most of the sites do not have continuous temporal data. We use PM2.5 as our measure of pollution. However we interpret our results more broadly as the impact of pollution since PM2.5 is a correlate of other pollutants (like NO2, SO2, CO) which are not captured in our analysis. To address the

[7] India has around 600 ground based monitors to cover the entire country with only 148 monitors which capture PM2.5 for the entire country.

paucity in ground-based pollution data in India, we estimate PM2.5 exposure using satellite data (van Donkelaar et al., 2010). We convert Aerosol Optical Data (AOD) retrieved at 0.5 x 0.5 degree resolution from Multiangle Imaging SpectroRadiometer (MISR) to PM2.5 data (Liu et al, 2004; Kahn and Gaitley 2015; Dey et al. 2010) using a spatially and temporally heterogeneous conversion factor (Dey et al., 2012). The PM2.5 data is further statistically downscaled at 0.1 x 0.1 degree resolution using spline interpolation. The PM2.5 thus obtained is available at monthly frequency at 0.1 * 0.1 degree resolution (10km*10km grid).

As background, we explore the spatial variation in PM2.5 by plotting a heat map in Figure 3.1. We plot mean annual PM2.5 (averaged over monthly data for years 2010 to 2016) for each district of India. As the figure shows, the Northern region of the country is severely impacted by high and dangerous levels of pollution, especially the states which lie in the Indo-Gangetic plains (Punjab, Haryana, Uttar Pradesh, Bihar) have the highest levels of pollution. On the other hand, the Southern part of the country has much lower levels of pollution as shown by the lighter shades in heat map. The WHO guideline for maintaining safe standards of pollution recommends a threshold of mean annual pollution levels of $10ug/m^3$. Other standards include WHO-IT1 which is $35ug/m^3$, WHO-IT2 which is $25ug/m^3$ and WHO-IT3 which is $15ug/m^3$. The Indian National Ambient Air Quality Standards (NAAQS) sets the threshold at $40ug/m^3$. For our estimation sample, the mean level of PM2.5 is $54ug/m^3$ during the first trimester (Table 3.1).

We use cluster location from DHS data and calculate weighted mean PM2.5 in the 75km radius around it, for each month since the time of conception. This is done by weighting the PM2.5 in each grid belonging to the circle by inverse of the distance of each grid point from the cluster location (center of the circle). We use these monthly pollution measures to construct trimester level pollution exposure by calculating mean PM2.5 for three month periods. Our estimation sample is constrained by the availability of PM2.5 data as we only keep those children in our estimation sample for whom the pollution measure for each month in the first trimester is present. The missing PM2.5 are due to missing satellite retrievals due to cloud covers. In the appendix Table A3.1, we show that our outcome variables along with our control variables are very similar between the estimation sample and out-sample (with missing PM2.5 information).[8]

We now link exposure to pollution during the first trimester with anthropometric measure (Height-for-age z-score) for children in Figure 3.2. The descriptive graph is a bin-scatter plot which shows a negative relationship between Height-for-age and exposure to pollution during first trimester. We convert pollution exposure figures to z-scores for ease of interpretation. We transform the mean PM2.5 in first trimester into z-scores based on the average and standard deviation of first trimester mean PM2.5 in the estimation sample. However this correlation may be driven by other factors. We explore this relationship empirically in greater detail in Section 4. We also split our sample into subgroups (5 quintiles) based on exposure to mean PM2.5 during first trimester. Figure 3.3 shows the relationship between Height-for-age (and weight-for-age) and child's age for two groups - the first group comprises of children who belong to 1st quintile (lowest mean PM2.5 exposure during first trimester) while the second group comprises of children who had exposure to high levels of PM2.5 (5th quintile) during their 1st trimester. The graph essentially shows the dose effect of exposure to pollution, children in low exposure group have higher Height-for-age (and weight-for-age) in comparison to children belonging to high exposure group. This relationship is shown in greater detail in Table 3.1, where child growth indicators (height-for-age and weight-for-age) progressively become worse as exposure to pollution during first trimester increases. The mean PM2.5 in first quintile is $18ug/m^3$ while it is $106ug/m^3$ for the fifth quintile. It is however also important to note that other characteristics also differ across the quintiles. Hence we conduct a multivariate analysis.

3.4 Empirical Model

As pointed out above, we seek to investigate whether early life exposure to outdoor pollution during first trimester has an impact on future child health, measured by Height-for-age (z-scores) and Weight-for age (z-scores). Formally, we estimate the following empirical model:

$$H_{icdmt} = \theta_1 PM_{cdmt} + \beta X_{icdmt} + \gamma_c + \delta_{dt} + \rho_{mt} + \varepsilon_{icdmt} \tag{3.1}$$

[8]To show that there is no difference, we use as controls the same set of covariates and trends as in our main specification. This is done to ensure that there are no systematic differences, after partialling out the covariates used in our analysis.

Our main outcomes of interest (H_{icdmt}) are z-score for Height-for-age (stunting measure) and Weight-for-age (underweight measure) for child i who was conceived in cluster c belonging to district d in month m and year t. The main variable of interest is PM_{cdmt} which captures the standardized PM2.5 (i.e. z-scores) in the 75km radius during first trimester for a child. We control for confounding factors in the vector X_{icdmt} which includes gender, birth order and age of child, mother's and father's educational status, mother's age at birth, age and gender of household head, dummy variables for whether the household has piped water, has clean cooking source, whether household practices open defecation and the fraction of households who practice open defecation in the cluster (excluding self). Since all children in our sample are aged five or below, we use the assumption that these controls have not changed a lot over time (i.e. from the time of conception to the time when they were surveyed).

Different clusters (villages or blocks) can have different levels of development (health infrastructure) which can affect health of a child hence we include cluster fixed effects in our specification. We also remove any omitted variables that are related to a district in any particular year as well as any seasonality effect specific to a month of a particular year by including a district year specific fixed effect, δ_{dt} and a month year specific fixed effect, ρ_{mt}. The inclusion of these fixed effects means that the variation that remains is the spatial variation in pollution within clusters and temporal variation within a year for a district. We cluster standard errors at the sampling-cluster level.

While our estimation exercise removes systematic variation using various fixed effects, endogeneity concerns still remain. These endogeneity concerns arise as the *local* residential area for a household corresponds to the region of economic activity that a household depends on and in turn affects based on its behavioural decisions. The economic activity of a household determines key inputs (like income) which feed into the production function of health of a child. An example of this can be dependence of a household on nearby forest resources for fuel-wood consumption or for livelihood (if it sells these resources in a market). In this case the choice of use of fuel-wood by household affects the local pollution level in the region. Additionally the forest cover is affected by the demand for forest resources (like fuel-wood) in the market, which in turn affects the pollution level in the area where they are finally consumed. A similar logic holds true for crop residue burning as well, it is a conscious decision taken by a household which impacts local pollution levels and at the same time affects a farmer's income which is a determinant of child health. Thus, local pollution level is endogenous in the region of economic activity of the household.

3.4.1 Identification

As pointed out above, the household behavioural choice of collecting fuel-wood or crop-burning and household income are omitted variables in our specification hence the local pollution variable is endogenous. To solve this endogeneity problem, we use a standardized measure of number of upwind fire-events (more on this below) which take place in the 75 to 100 km radius of the sampled cluster as an instrument. Fire-events are sourced from satellite image that are divided into pixels. Number of fire-events refers to the number of pixels where atleast one fire-event is located within the pixel. These are biomass burning events that include crop residue burning and forest fires. Further, when fire incidents are recorded then each of them has a confidence value attached (interpreted as probability) which depicts the quality of the observation and therefore, using this we construct a confidence weighted count of fire-events around a cluster. We use only upwind fire-events, that is fire-events from which wind is blowing towards the cluster [9]. Further, following Rangel and Vogl, (2018) for ease of interpretation of results, we standardize the events by calculating z-scores for these fire-events occurring in each cluster.

We use such fire-events only in the radius between 75 and 100 km (we refer to this area as a non-local area) as they impact local mean PM2.5 levels (within 75km radius of a cluster) but are not affected by household behavioural choices. To elaborate further, these fire events belong to a region which is not a part of economic activity area of a household. This essentially removes the effect of dependence on crop-burning or nearby forest resources (or farmlands) for livelihood or fuel-consumption. Thus by capturing fire-events in this non-local area, we ensure that we only capture the part which contributes to the local pollution levels but is not correlated with household behavioural choices. Further exogenous changes in wind direction is unlikely to be correlated with local behavioural choices or economic activity in the area and thus non-local *upwind* fire-events serves as an ideal instrumental variable. Zheng et. al

[9] Downwind fire-events refer to events with wind blowing away from the cluster.

(2019) use a similar instrumental variable (IV) in their paper where they study the impact of air pollution on happiness levels using pollution levels of neighbouring areas as an IV for local pollution levels.

The IV that we use has been explained diagrammatically in Figure 3.4, where the light grey center denotes the cluster location, the white circle forms the 75 km radius around the cluster and the grey ring represents the area between 75 and 100 km radii around the cluster. Our endogenous variable is the mean PM2.5 variable which is calculated for the white circle (within 75 km) and the probability weighted number of upwind fire-events in the grey ring forms the IV (between 75 and 100 km).

Fire-events and Wind Data

Our source of biomass burning events (fire incidents) is NASA's Fire Information for Resource Management System (FIRMS) data which captures real-time active fire locations across the globe. The FIRMS data that we use is called MODIS (shortform for MODerate Resolution Imaging Spectro radiometer) data and it records fire incidents at pixel level where each pixel is identified by a latitude and longitude reading. Each latitude (and longitude) is the centroid of a one kilometre pixel (1 km X 1 km in size). This data records not just the location of a fire but also the brightness (temperature) of fire (in Kelvin units) and date and time when the incident was picked by the Terra satellite. An observation for a fire incident in MODIS data for a latitude and longitude does not necessarily mean that the size of the fire is one square kilometre, but it means that atleast one fire is located within this fire pixel (under good conditions the satellite can detect fires as small as $100m^2$). The MODIS data is available on a daily basis since November 2000 and NASA reports that the fires captured by this dataset are mostly vegetation fires. NASA data on fire incidents also provides a variable "confidence", which depicts the quality of the observations and it ranges from 0-100[10]. Following Rangel and Vogl (2018), we use this variable to construct a probability weighted count of fire-events around a cluster (between 75 and 100 km)[11].

We use the cluster GIS information from DHS data and calculate the probability weighted

[10] We convert confidence figures into probability figures by dividing them by 100

[11] NASA's FIRMs data can also have some missing values attributable to satellite sensor outage. However major incidents reported for sensor outage happened in years 2001-2003 which precedes our analysis period.

count of fire events which took place between 75 and 100 km radii (non-local exposure) during the first trimester of a child. To ensure respondent confidentiality, all clusters in the DHS data are displaced from their true location. The displacement is done by displacing an urban cluster by two kilometre and a rural cluster by five kilometre with one percent of the rural clusters being displaced by as much as 10 kilometres. The displacement can take place in any direction but the cluster remains within the country boundary, within the same state and district. We take the radius for our analysis to be 75 kilometre which is large enough so that the true location of the cluster and sphere of economic activity of a household is contained within the 75 kilometre radius circle.

Meteorological variables such as wind speed and wind direction are expected to play an important role in modulating the outflow of fire burning residues emitted from a fire event. To account for this, we tag each fire event with wind direction. We use ERA-Interim data of u (zonal wind) and v (meridional wind) at 10m from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim dataset at 0.125*0.125 degree resolution.

The wind direction was estimated as in equation (2) (Chowdhury et al., 2017):

$$winddirection = [atan(u/v) * (180/pi] + 180 \tag{3.2}$$

The wind direction is coded in degrees, such that 0 corresponds to wind from due North, and 180 corresponds to wind from due South. For our analysis, using the wind direction of a fire-event, we construct a 45 degree modal octant[12] around this wind direction which captures the pollution dispersion from polluting source that is fire-events in our case. If a cluster falls in this modal octant then this fire-event is tagged as an upwind fire-event for that cluster (fires occurring in opposite octant are tagged as downwind fire-events). Upwind fire-events are likely to affect local pollution levels.

Fire-events and Pollution

India has a substantial amount of land under cultivation(60%) and under forest cover(25%), with majority biomass burning events taking place in these areas. Over the past few decades,

[12]Following Rangel and Vogl, (2018)

Indian agriculture has been marked with expansion of irrigation facilities, adoption of high yield variety seeds and increased mechanisation (like use of combine harvester). A combination of these factors led to adoption of multi-cropping system by farmers which leaves little time in between the harvest of one crop and sowing of another. In this scenario, crop residue burning thus emerged as the quickest and cheapest way to get the farm ready for the next crop. Cereals are the prime contributor to crop burning activity in India, with rice and wheat crop residue burning forming the major chunk of residue burning process (Jain et al, 2014). Two major residue burning seasons are thus related to crop harvest seasons: kharif crop harvest (rice stubble burning) which takes place in the months of October and November; and rabi crop harvest (wheat straw burning) which happens in the months of March to May.

Biomass burning in India is not limited to just crop residue burning, it covers forest fires as well. Forest fires or wildfires are caused by various factors acting in conjunction with each other. These factors include availability of biomass (dry vegetation) and appropriate climatic conditions(high temperature, low pressure, windy conditions). Forest Survey of India lists vulnerable months for each state when forest fires are most likely to happen, which mainly span the high temperature months from March to June. Wildfires happen due to both intentional and unintentional human activity. In North Eastern states and in states along the Eastern Ghats, slash and burn activity is rampant wherein vegetation in forests is cut (slashed) and then burned to clear the piece of land for human use. In a lot of cases unintentional human activities like leaving active cigarette butts behind in open forests lead to forest fires. Other natural factors which cause forest fires include lightening which produces a spark to start a fire in dry vegetation.

Figure 3.5 provides a linear fit plot between local PM2.5 levels and non-local fire-events (all fire-events - left panel and just upwind fire-events - right panel). A strong positive relationship between the two is evident from this graph and forms the basis for using non-local fire-events as an IV for local pollution levels. In our empirical work though, we use the within cluster and temporal variation of these variables.

In Figure 3.6, we plot the temporal variation in PM2.5 and non-local fire-events. The figure plots mean levels across all sampled clusters in the latest DHS round for India. We look at mean PM2.5, mean count of total fire-events which take place in non-local areas and mean count of total upwind fire-events which take place in non-local areas. As shown in the graph (solid blue line) the winter months (from October to January) have highest pollution levels in comparison to summer months (March to June), with lowest pollution levels recorded in monsoon period (August-September). Corresponding to two harvest seasons we see two peaks in fire-events plots (both all fire-events and upwind fire-events in dashed lines).

In western countries forest fires are mainly responsible for the carbon content release due to biomass burning; however, in case of India (and other South Asian countries) crop residue burning contributes the most to total carbon release. In South Asia, India stands out both in terms of total area burned (4.5 million hectares burned in 2015) and in terms of total carbon content (1.5 million metric tonnes) released due to biomass burning. A raw count of biomass burning events in India shows that roughly both crop residue burning and forest fires contribute equally. However, if we weigh these events based on the population density [13] of the area in which these events occur then crop burning events contribute more to the total biomass burning events (65 percent). This mainly happens because residue burning activities happen in more populated areas as against forest fires which happen in low density areas. Appendix Figure A3.1 provides the population weighted split between forest fires and crop residue burning in few selected states in India. As can be seen in this graph, with an exception of Punjab, almost all other states are affected by both forest fires and residue burning.

Biomass burning is a major source of pollution as it releases harmful pollutants like Carbon Dioxide(CO_2), Carbon Monoxide (CO), Sulphur Oxides and particulate matter (PM) in the atmosphere. The release of harmful pollutants in the atmosphere is captured by aerosol loading [14] in the region. To summarise, fine particulate matter released during biomass burning incidents have long range travel properties and affect not just the local areas but far away regions as well.

Arguments above provide some suggestive evidence about the fact that non-local fire events are associated with local pollution levels. Further evidence on this will be provided when we discuss the first stage of 2SLS regression. In addition what we also require for our IV strategy

[13] Geo-coded fire events have been projected onto land mask cover for India to categorise each fire event as an event which happens in a forest area vs cropped area. This data is then projected onto density map of India, to get the density of the population in which these events take place.

[14] Aerosol loading is the suspensions of solids and/or liquid particles in the air that we breathe. Dust, smoke, haze are also part of aerosol loading.

to work is that our IV should be uncorrelated with other factors which are related to child health. We provide evidence in next section that it is likely to be true.

3.5 Results

3.5.1 Pollution and Child Health

We begin by presenting OLS results on effect of mean outdoor pollution in the first trimester on child health outcomes in Table 3.2. Column 1 and 2 in Table 3.2 show that pollution exposure during first trimester is negatively correlated to weight-for-age (WFA-Z) and height-for-age (HFA-Z). The OLS point estimates are small: a one standard deviation change in local PM2.5 reduces WFA-Z by 0.012 and HFA-Z by 0.011 standard deviation units (the latter estimate is not significant). A possible reason behind small coefficients could be the fact that local pollution exposure subsumes the effect of both income and physiological effect of PM2.5 on child health. Since these two effects can affect child health in opposite ways so the OLS estimate we notice is smaller. Also, as described in the previous section, in equation (1) local pollution exposure variable is riddled with endogeneity problem, hence the OLS estimates are inconsistent.

To address the endogeniety problem, we use an Instrumental Variable strategy where upwind fire events in the non-local areas are used as an instrument for local pollution levels. These upwind fire-events are assumed to be orthogonal to the income levels, so we are able to capture the pure effect of PM2.5 disentangling it from the income effect. We present the first stage of 2SLS results in Table 3.3. As hypothesized, we find that the relationship between the endogenous variable - local PM2.5 in 75km radius and non-local upwind fire-events is positive and highly significant. A one standard unit change in number of upwind fire-events leads to a 0.105 standard deviation unit increase in local pollution levels. This is in line with our hypothesis that particulate matter from fire-events far away affect local pollution levels. The first stage rk-LM statistic is 1005 and is much above the Stock & Yoko bias cut off. Hence our instrument is strong. To summarize, local PM2.5 variation is robustly affected by the seasonality present in biomass burning events happening in non-local adjacent areas.

We test whether our IV meets exclusion restriction by providing some suggestive evidence in Table 3.4. We regress various characteristics of a household (and its members) on our main IV - upwind fire-events, essentially an insignificant result shows that there is no systematic relationship between our IV and household (and its member's) characteristics. We do this by regressing variables which affect child health on our IV, columns 1 to 10 in Table 3.4 shows that the controls in our main specification are not systematically related to the IV. Columns 11 to 16 provides results for other variables (these include wealth class, asset ownership, religion, dummy for minority group, household size and vaccination) which can potentially affect child health and we find that they are also not related to fire-intensity in non-local areas.

Next, we move to the second stage results obtained using 2SLS strategy. We find that both WFA-Z and HFA-Z are negatively affected by outdoor pollution experienced in-utero during the first trimester. Columns 3 and 4 in Table 3.2 present our 2SLS results using upwind fire events in non-local area as an IV. We find that a standard deviation unit change in mean PM2.5 during first trimester leads to a decrease in WFA-Z by -0.103 standard deviation units and a decrease in HFA-Z by -0.116 standard deviation units which translates into a 6.7 percent decrease in WFA-Z and 7.9 percent decrease in HFA-Z.

Moving to other covariates (results reported in Appendix Table A3.2), we find that being a male child, or being born later (higher birth order) is associated with lower HFA-Z and WFA-Z. Similar to previous findings in the literature, we find that child growth indicators are positively associated with mother's education level and also age at which mother gives birth. Source of water being pipedwater seems to have no affect on child health while use of clean cooking fuel is associated with better child health outcomes. Household's open defecation practice is negatively associated with stunting and underweight measures. Finally, an older household head perhaps contributes to better child care and hence is associated positively with child health outcomes, while gender of the household head being male only affects stunting measure.

We now focus on other time windows of critical development, that is second, third trimester and the post-natal period of first three months after birth. Table 3.5 summarises our results, we find that in-utero exposure to outdoor pollution which is experienced by the mother (and her foetus) for second, third trimester and post-natal period[15] has no impact on Height-for-age, but there is a negative effect for Weight-for-age corresponding to exposure in second trimester. To summarize, the effects of pollution have a robust negative effect on health of a child when the exposure is in the first trimester but less so in other trimesters.

3.5.2 Robustness Checks

Extended Controls

a) Local weather conditions

In this section we provide multiple robustness checks for our results. Local weather condition like rainfall can play an important role as rainfall makes the ash and other pollutant particles settle on the ground thereby reducing pollution levels. Temperature also plays an important role in pollution dynamics. We control for both local temperature and rainfall in columns 1 and 2 in Table 3.6. The number of observations is slightly smaller than before due to missing rainfall and temperature information for some clusters. Our original results still hold and the point estimate is slightly larger after accounting for weather controls.

b) Gestational period, Household Size, Caste

Next, we add more control variables to our regression. The gestational period or pregnancy duration is also an important determinant of intrauterine growth of a foetus which affects future child health. We additionally control for household size and minority status of a household (being schedule caste or schedule tribe) to see if extended controls affect our original results. We find in column 3 and 4 (in Table 3.6) that our estimates remain unchanged. Duration of gestational period is positively associated with child growth indicators while children belonging to minority group have worse health outcomes. Household size seems to have no effect on child growth indicators.

[15]The number of observations differ depending upon availability of pollution data for all months for the window of analysis.

Sensitivity Analysis

The analysis so far used upwind fire-events happening in 75 to 100 km radius as the IV for local mean PM2.5 in the 75 km radius around the cluster location. We now provide results for alternate radii specifications to test the sensitivity of our model. In Table 3.7, columns 1 and 2, the IV being used is the probability weighted total number of upwind fire-events in 50 to 100 km radius (compressing the white inner circle in Figure 3.4). In columns 3 and 4, the IV being used is the probability weighted total number of upwind fire-events in 50 to 75 km radius for local mean PM2.5 in the 50 km radius (compressing the donut in Figure 3.4). Reducing the local pollution radius to 50 kms leads to a significant drop in total number of observations as PM2.5 information is missing for a lot of observations. However we still find that our results are of similar magnitude (they are slightly smaller for HFA-Z analysis) and still remain significant. The HFA-Z result in Table 3.7 column 2 is significant at 10 percent level while in column 4 it is marginally significant at 10 percent level (p-value = 0.109). Lastly in columns 5 and 6, we drop the observations corresponding to the state of Punjab. This has been done to ensure that our results are not driven in any way by the state of Punjab which is affected by high levels of pollution corresponding to highest level of recorded fire-events in India[16]. Our point estimates become larger in magnitude and are more significant after dropping the state of Punjab.

Falsification Tests

In Table 3.8 (column 1 and 2) instead of using upwind fires as an IV we use downwind fires (which lie in opposite octant from that of upwind fires with wind blowing away from cluster location). We find that using downwind fire-events as an IV makes our results insignificant and in case of HFA-Z the insignificant point estimate has the opposite sign. In columns 3 and 4, we provide results on the effect of pollution on child health where the location of a child has been randomly shuffled. This random assignment of location leads to counter-intuitive positive (and insignificant) effect of early life exposure to pollution during first trimester on child health which strengthens our hypothesis that location does matter when it comes to

[16] Almost 25% of total fire-events in India take place in Punjab.

pollution exposure (and in turn affects child health).

Avoidance Behaviour

Do mothers plan conception?

An important threat to our identification can be avoidance behaviour by mothers, that is if mothers purposely avoid particular months for conception due to their concern about future child health related to seasonal biomass burning activities. We test this by looking at birth history of mothers for the estimation period i.e. years 2010 to 2016. We do this by creating a mother-month-year panel. We create a dummy variable which takes value 1 if a mother successfully conceives in a particular month of an year. We estimate a linear probability model to test whether mother's conception behaviour is systematically linked to non-local fires. We control for mother's education, characteristics of household head along with other household characteristics like source of water, toilet facility, choice of cooking fuel. We introduce the same fixed effects which are present in our initial specification to control for regional and seasonal factors. We present these results in Table 3.9. In column 1, we present results where we try and see whether there is correlation between three month exposure to non-local upwind fire-events and conception. In column 2, we assess whether exposure to fire-events in the month of conception has any effect on conception behaviour. We find that in both the cases non-local upwind fire-events increase the probability of conception. This suggests that we have some positive selection, as probability of mothers conceiving is more during the time when incidence of fire-events is high. A one standard deviation in non-local fire events in a month increases the probability of conception by 0.015% (corresponding figure for 3 month exposure is 0.03%. Our analysis provides some evidence that mothers do not practice avoidance behaviour.

3.5.3 Heterogeneity

i) By Background Characteristics

We now provide disaggregated regressions for Height-for-age. We split our estimation into Poor (wealth index lower than 2) and Rich sample (wealth index greater than equal to 3). In Table 3.10 (Column 1 and 2), we find that the negative effect of pollution is present only for poor households. This can possibly be due to the fact that children in poor households have less access to health care to abate negative effect of pollution on health.

Next, we compare children who are born to mothers with different educational attainment. In column 5 and 6 of Table 3.10, we find the negative effect of pollution on child health is mainly present for mothers who have less (till primary level) or no education. There is negative effect present for educated mothers (secondary or above) as well but it is not significant.

These results point out that while there are negative effects of exposure, there can be offsetting practices correlated to education and wealth that alleviate the negative effect.

Finally, we focus on the location of the residence of a child. While mean PM2.5 is above $56ug/m^3$ during all critical windows of development for children in North India, the corresponding figure for South India is as low as $35ug/m^3$. We do sub-sample analysis on observations from Northern and Southern States in columns 3 and 4 and find that most of the effects that we see are limited to North India which have alarmingly high levels of pollution throughout the year.

ii) By Child's Age

In Table 3.11 we test whether the effect of in-utero exposure to pollution persists overtime. We observe that the effect is negative for all age-groups (zero to one year, greater than one but less two years and greater than three years old). Studies have suggested that stunting is irreversible after age of two years, we do find a significant negative persistent effect of early exposure to pollution on stunting outcome for one to two year old children. Although the effect is smaller but it continues to be present for children older than 3 years as well.

3.6 Conclusion

Outdoor pollution in India breaches safe standards in many areas. We link outdoor pollution to biomass burning which is a significant source of carbonaceous aerosols, it plays a vital role in atmospheric chemistry, air quality, ecosystems, and human health. Our analysis shows that outdoor pollution is affected by neighbouring biomass burning events; this is used to causally infer the effect of outdoor pollution (as measured by PM2.5) on child growth indicators.

We find that a z-score increase in PM2.5 levels during first trimester leads to a reduction in Height-for-age (HFA-Z, stunting measure) and Weight-for-age (WFA-Z, underweight measure) by 0.115 and 0.102 standard deviation units respectively. Figure 3.7 summarises our results graphically, exposure to outdoor pollution during different critical windows of growth of a child is associated with worse child health outcomes. Almost all the estimates are negative with significant effect present for exposure to pollution during first trimester and second trimester (only WFA-Z measure).

The above results establish that exposure to pollution is linked to stunting measure (HFA) in childhood. What impact does this have on the economy? We provide a back-of-an-envelope calculation based on the Galasso et al. (2016) study. This study does a literature review of the effect of stunting on GDP. Stunting affects GDP of a nation via three channels: lower returns to lower education, lower returns to lower height and lower returns to lower cognition. For India, where 66 percent of the workforce was stunted in childhood, this study estimates that a complete elimination of stunting would have increased GDP by 10 percent [17]. We use a point estimate of probability of being stunted due to outdoor pollution, and find that one standard deviation increase in outdoor pollution leads to a 0.18 percent reduction in GDP.

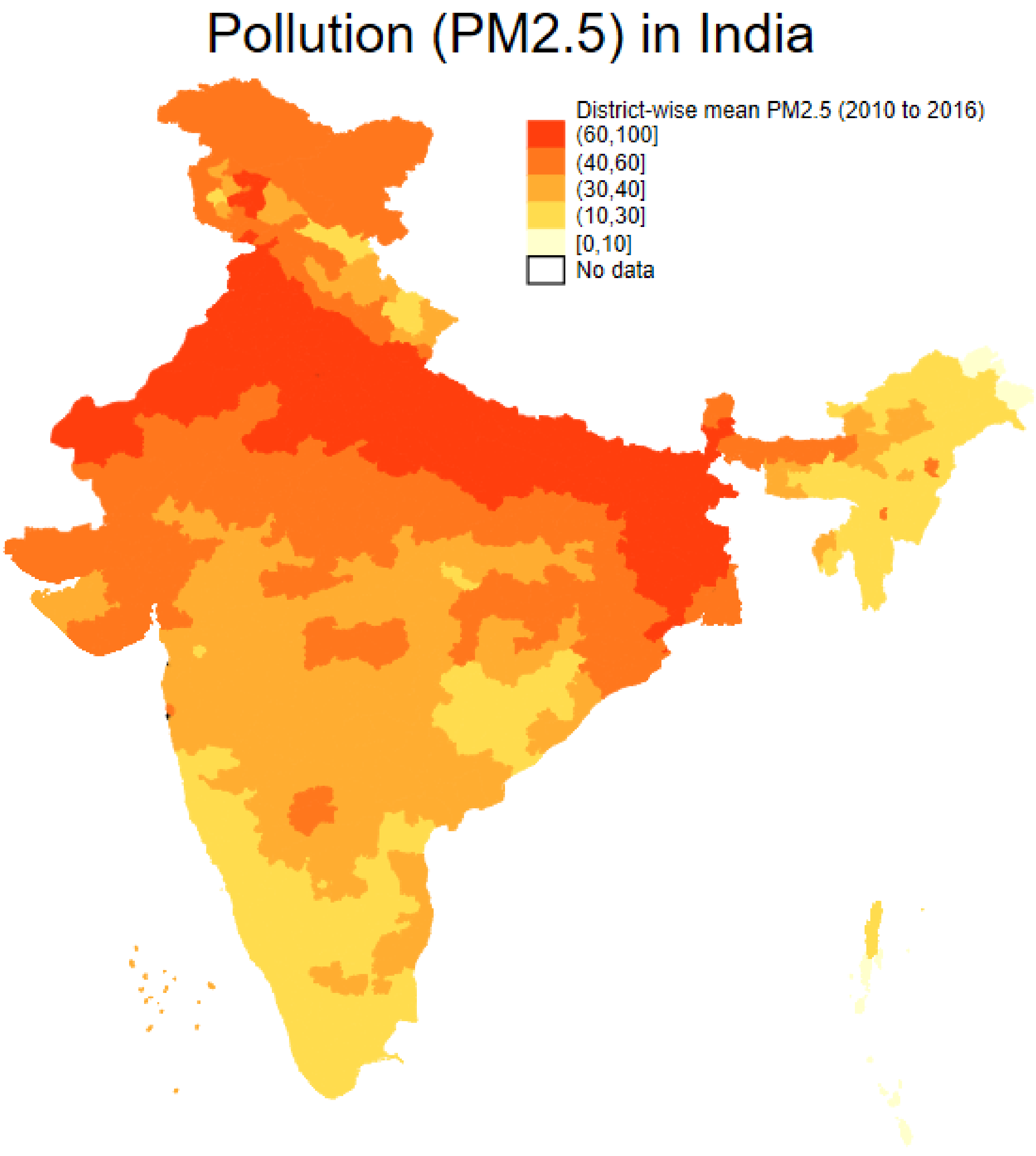
India needs effective policies regarding regulation and management of outdoor pollution, since the current policies are ineffective. Cross-border policies are needed to tackle the problem of pollution. To curb air pollution, effective management of forest fires is needed; however, the budget allocation for this purpose is really small and remains unused in every financial year. Similarly the government has committed itself to subsidising the use of happy-seeder technology (this is an alternative to combine harvester, it leaves rice residue in form of a mulch on farm which doesn't hamper wheat crop sowing and hence doesn't require burning), however the uptake of this policy remains quite low due to high initial investment in the machine (Gupta, 2014). The National Clean Air Program (2018) is a welcome step in this domain as it plans to extend air quality monitoring network, conduct intensive awareness and monitoring campaigns, create city-specific action plans, among many other initiatives.

[17] This is an average figure for South Asia.

Figures and Tables for Chapter 3

Tables and Figures

Figure 3.1: Spatial Variation in Pollution



Mean PM2.5 in districts of India (2010 to 2016)

Figure 3.2: Binscatter Plot between Height-for-age, Weight-for-age and Mean PM2.5 in First Trimester

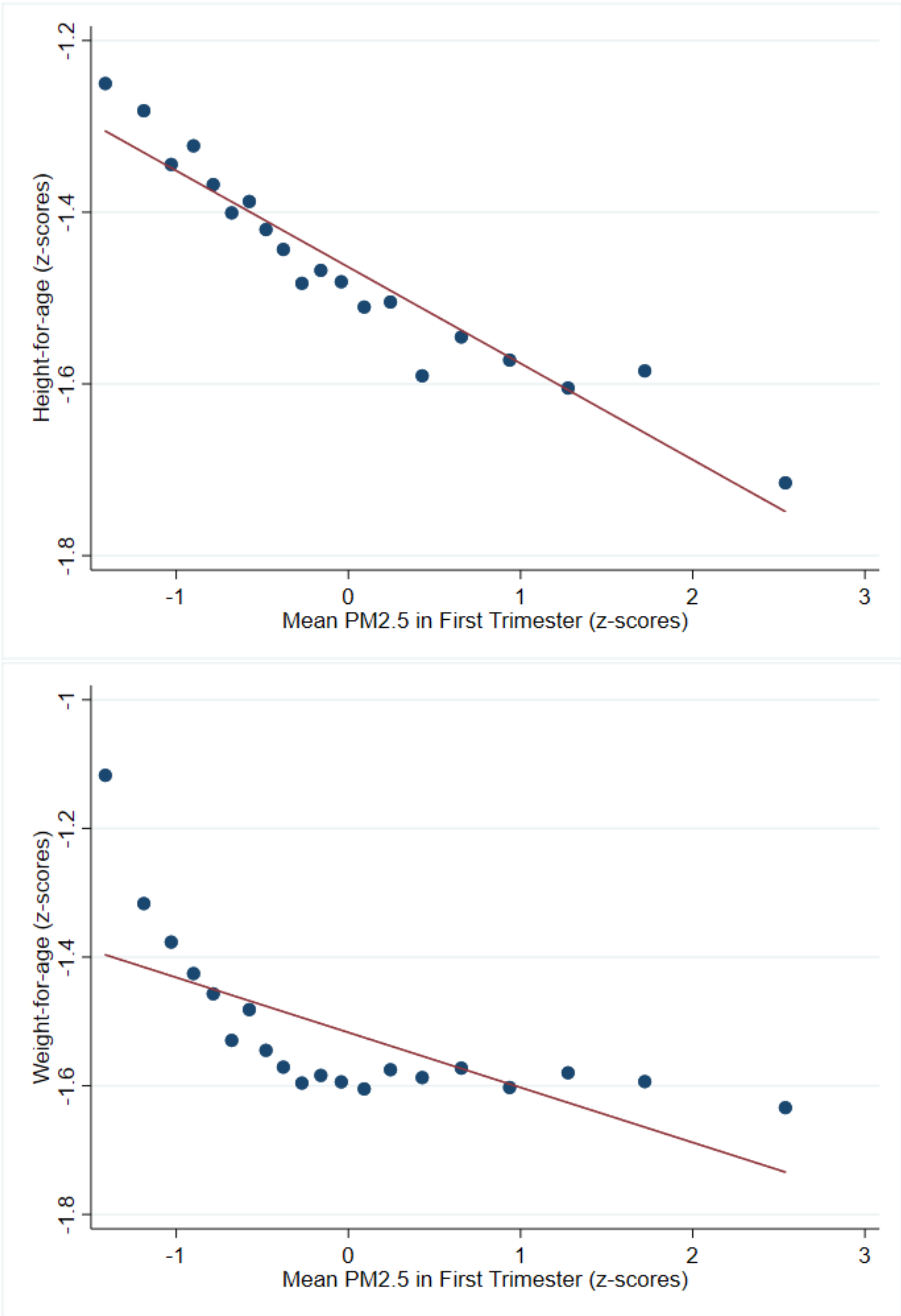
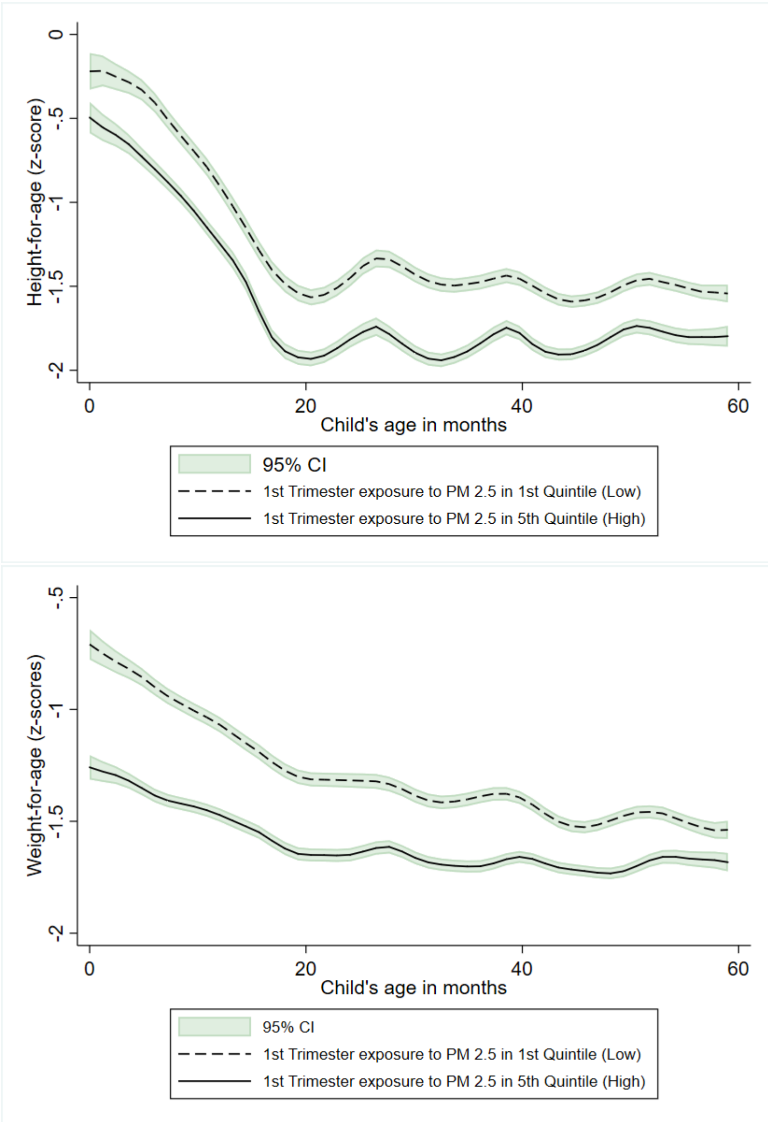
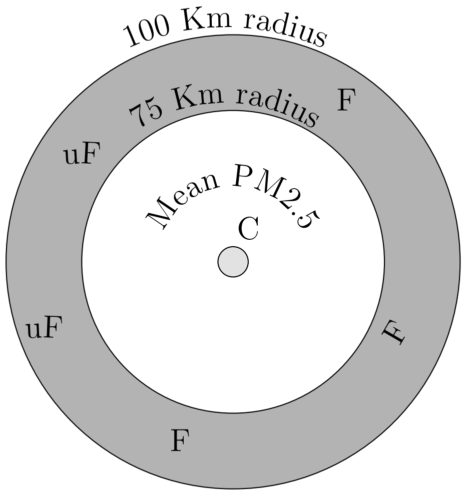


Figure 3.3: Polynomial Fit Plot between Height-for-age, Weight-for-age and Child's Age by Exposure to Pollution



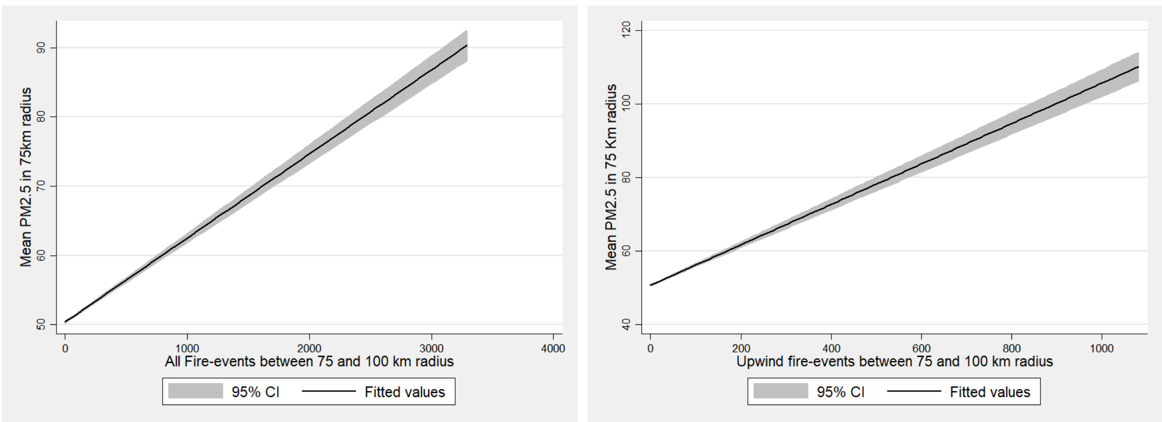
Polynomial fit plot between height-for-age (and weight-for-age) and child's age in months for children who had low level of exposure to pollution (1st Quintile) versus those who had high level of exposure (5th Quintile) to pollution during their first trimester. Shaded area is 95% confidence interval.

Figure 3.4: Identification Strategy



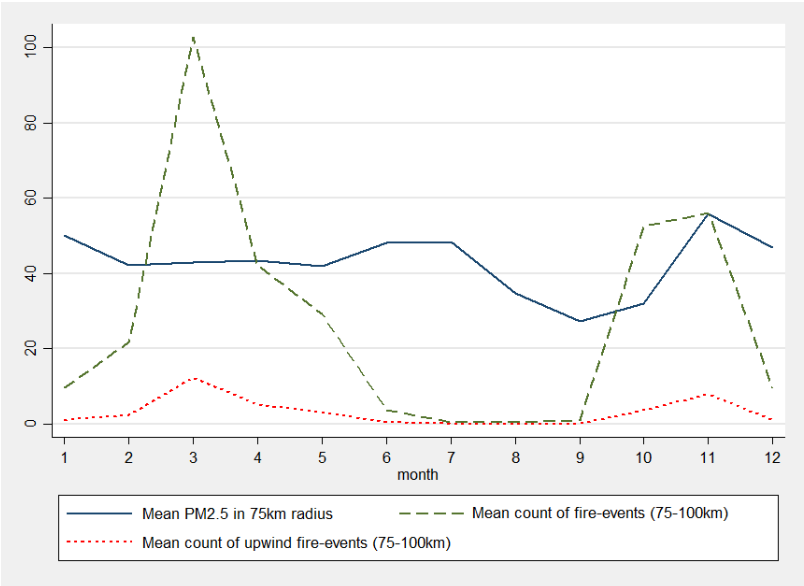
Center (smallest grey circle) represents the cluster location, White circle corresponds to 75km radius circle around the cluster location, grey ring area (donut shape) corresponds to area between two circles (75 and 100 Km radii circles) with cluster location as the center. Mean pollution level is calculated for the white circle, we call this *local* pollution level for cluster C. *Local* pollution level is instrumented using *upwind non-local* biomass burning events which take place in the grey ring area (only uF). Probability weighted counts used everywhere.

Figure 3.5: Relationship between Pollution and Fire-events



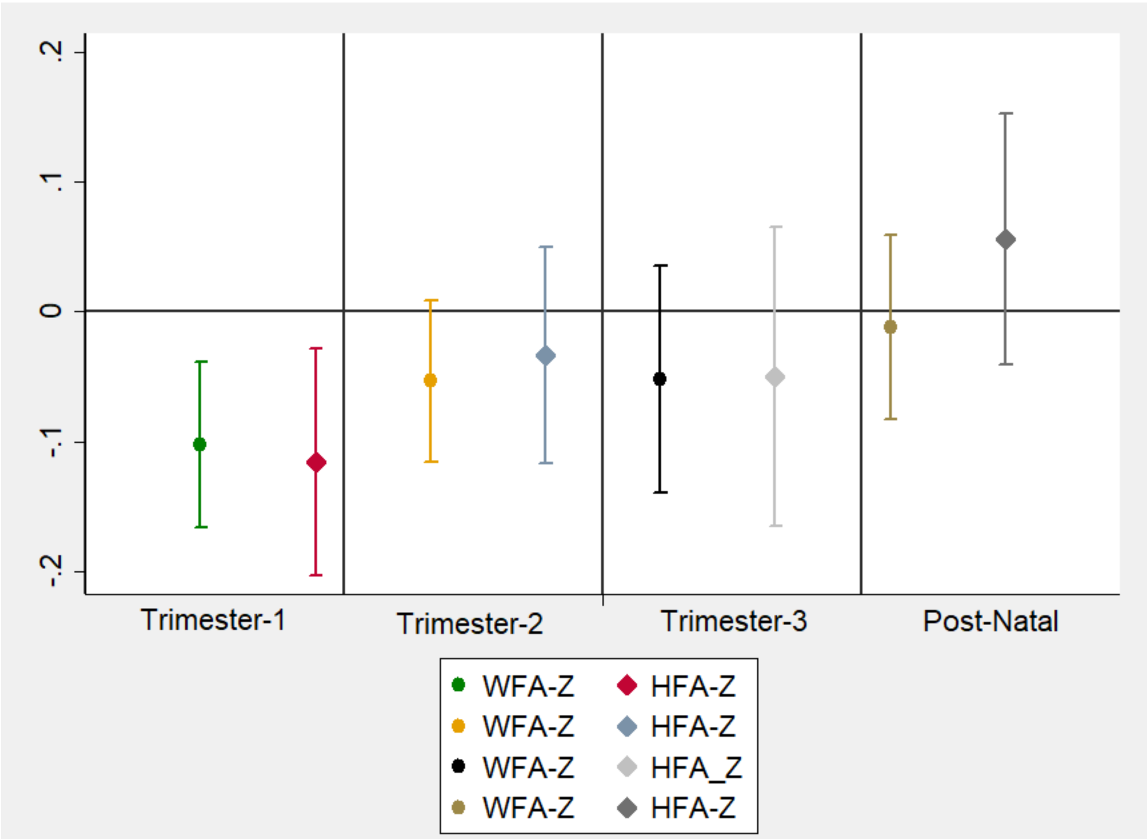
Linear fit plot between Mean PM2.5 (in 75 km radius), Total number of fire-events between 75 and 100 km radius(Non-local fire-events) & Total number of upwind fire-events between 75 and 100 km radius (Non-local upwind fire-events). Unit of observation is a child, shaded area is 95% confidence interval.

Figure 3.6: Temporal Changes in Pollution and Fire-events



Mean PM2.5, Mean count of all fire-events (in 75-100 km radius) & Mean count of upwind fire-events (in 75-100 km radius) for each month in every year from 2010 to 2016. Figure represents mean over all sampled clusters belonging to all states of India.

Figure 3.7: Trimester Wise Effect of Outdoor Pollution on Health Outcomes



Coefficient of 2SLS regression of outcomes(HFA-Z and WFA-Z) on outdoor air pollution for different critical windows of development of a child. Vertical lines represent 95 percent confidence intervals.

Table 3.1: Summary Statistics

		Mean PM2.5 (1st Trimester)				
		Quintiles				
variablename	All-India	1	2	3	4	5
Height-for-age z score	-1.46	-1.30	-1.39	-1.47	-1.54	-1.62
Weight-for-age z score	-1.52	-1.31	-1.50	-1.59	-1.58	-1.60
Weight-for-height z score	-0.97	-0.82	-1.01	-1.05	-1.00	-0.97
Child characteristics						
Dummy for male child	0.52	0.51	0.52	0.52	0.52	0.53
Birth-order	2.26	2.16	2.16	2.23	2.35	2.41
Childage in months	29.00	30.67	27.78	28.35	29.21	29.00
Pregnancy duration	9.02	9.04	9.05	9.03	9.00	8.99
Parents characteristics						
Mother's age at birth	24.50	24.98	24.32	24.27	24.48	24.45
Mother's number of education years	6.26	7.16	6.53	6.16	5.79	5.68
Dummy for literate father	0.97	0.85	0.85	0.82	0.81	0.79
Household characteristics						
Rural	0.76	0.75	0.76	0.76	0.77	0.75
Dummy for head of household being a male	0.88	0.88	0.89	0.88	0.87	0.87
Age of head of household	44.59	44.28	44.51	44.84	44.64	44.69
Dummy for source of water: Pipedwater	0.23	0.27	0.25	0.23	0.21	0.21
Dummy for using clean cooking fuel	0.28	0.34	0.29	0.26	0.25	0.28
Dummy for household practicing OD	0.42	0.30	0.44	0.47	0.46	0.43
Fraction of HHs practicing OD in a village	0.43	0.32	0.45	0.48	0.47	0.44
Caste = SC or ST	0.37	0.50	0.40	0.36	0.32	0.29
Has electricity connection	0.85	0.91	0.88	0.85	0.82	0.80
Religion = Hindu	0.73	0.60	0.76	0.78	0.77	0.75
Household size	6.57	6.01	6.41	6.68	6.82	6.95
Mean PM2.5 in 75 km radius (1st Trimester)	54.06	17.99	34.01	47.30	65.39	105.62
Observations	181361	36273	36273	36273	36273	36273

Note: The estimation sample has been divided into 5 quintile groups, each containing 20 percent of the observations. Quintile 1 refers to the group with lowest exposure to PM2.5 during first trimester while Quintile 5 refers to the group with highest exposure to PM2.5 during first trimester.

Table 3.2: Effect of Exposure to Pollution During 1st Trimester on Child Health - OLS and IV Estimates

	OLS		**IV: Upwind fire-events between 75 to 100 kms radius**	
	(1)	(2)	(3)	(4)
	WFA-Z	**HFA-Z**	**WFA-Z**	**HFA-Z**
Trimester-1: Mean PM2.5 in 75km radius (z-score)	-0.0120**	-0.0116	-0.103***	-0.116***
	(0.005)	(0.007)	(0.032)	(0.044)
Mean of Dependent Variable	-1.52	-1.46	-1.52	-1.46
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes
Observations	179816	179816	179816	179816

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual OLS or 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester (z-score).Regressions include other controls - gender, birth order and age of child, mother's years of education, mother's age at birth and its square, age and gender of household head, dummy for whether household has pipedwater, has clean cooking source and whether household practices open defecation.

Table 3.3: First-stage Estimates for IV Regression

	Mean PM2.5 in 75km radius (z-score)
IV: Number of upwind fire events between 75 and 100km radius (z-score)	0.105***
	(0.003)
First Stage F-stat	863
rk LM statistic	1005
Anderson Rubin wald statistic (p-value)	0.0016
Stock & Yoko critical values:	
10 %	16.38
25 %	5.53
Observations	179816
Includes other controls from 2nd stage	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes

Note: Each coefficient corresponds to an individual FIRST stage 2SLS regression of HFA-Z or WFA-Z on variables mentioned in the first column. Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. PM2.5 & Fire-events variables have all been converted into z-scores. Regressions include controls which are same as those mentioned in Table 3.2 notes.

Table 3.4: IV Validity

	(1) **Male Child**	(2) **Birth Order**	(3) **Child Age**	(4) **Mother's Education**
Upwind fire events between 75 and 100km radius in 1st Trimester (Z)	-0.0020 (0.0017)	0.0028 (0.0032)	-0.0019 (0.0014)	0.016 (0.013)
Observations	179816	179816	179816	179816
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes
	(5) **HH Head Age**	(6) **HH Head is Male**	(7) **Mother's Age at Birth**	(8) **Source of Water is Pipedwater**
Upwind fire events between 75 and 100km radius in 1st Trimester (Z)	-0.011 (0.047)	0.0012 (0.00083)	-0.0014 (0.013)	-0.00027 (0.0012)
Observations	179816	179816	179816	179816
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes
	(9) **Uses Clean Cooking Fuel**	(10) **HH OD**	(11) **Poor**	(12) **Vaccination**
Upwind fire events between 75 and 100km radius in 1st Trimester (Z)	0.00046 (0.0012)	-0.00045 (0.00073)	0.0012 (0.00089)	0.00075 (0.0011)
Observations	179816	179816	179816	179816
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes
	(13) **Religion is Hindu**	(14) **Caste is SC or ST**	(15) **Household Size**	(16) **Asset Ownership**
Upwind fire events between 75 and 100km radius in 1st Trimester (Z)	0.00012 (0.00095)	0.00094 (0.0011)	0.010 (0.0065)	-0.0023 (0.0029)
Observations	179816	179816	179816	179816
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes

Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Fire-events have been converted to z-scores.

Table 3.5: Effect of Exposure to Pollution on Child Health: IV Estimates for 2nd Trimester to Post-natal Period

	(1) **WFA-Z**	(2) **HFA-Z**
Trimester-2: Mean PM2.5 in 75km radius (Z)	-0.05*	-0.03
	(0.03)	(0.04)
Observations	184183	184183
Trimester-3: Mean PM2.5 in 75km radius (Z)	-0.05	-0.05
	(0.04)	(0.05)
Observations	172917	172917
Post-natal: Mean PM2.5 in 75km radius (Z)	-0.01	0.05
	(0.03)	(0.04)
Observations	190717	190717
Includes Child, Mother & HH characteristics	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 a particular trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

Table 3.6: Robustness Checks: Extended Controls

	IV: Upwind fire-events b/w 75 to 100 km radius			
	(1)	(2)	(3)	(4)
	WFA-Z	HFA-Z	WFA-Z	HFA-Z
Trimester-1: Mean PM2.5 in 75km radius (Z)	-0.111***	-0.125***	-0.103***	-0.116***
	(0.0328)	(0.0453)	(0.0323)	(0.0448)
Mean rainfall in 75km radius	-0.0463**	-0.0492		
	(0.0217)	(0.0300)		
Mean temperature in 75 km radius	0.00336**	0.00348		
	(0.00151)	(0.00213)		
Gestational period			0.141***	0.172***
			(0.0252)	(0.0347)
Household Size			0.00549***	0.00304
			(0.00143)	(0.00194)
Caste is SC or ST			-0.131***	-0.153***
			(0.00877)	(0.0118)
Observations	179459	179459	178718	178718
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

Table 3.7: Sensitivity Analysis

	IV: Upwind Fire events b/w 50 and 100 Km radius		IV: Upwind Fire events b/w 50 and 75 Km radius		IV: Upwind Fire events b/w 75 and 100 Km radius Dropping Punjab	
	(1) **WFA-Z**	(2) **HFA-Z**	(3) **WFA-Z**	(4) **HFA-Z**	(5) **WFA-Z**	(6) **HFA-Z**
Trimester-1: Mean PM2.5 in 50km radius (Z)	-0.101***	-0.0867*	-0.111***	-0.0770		
	(0.0331)	(0.0461)	(0.0349)	(0.0481)		
Trimester-1: Mean PM2.5 in 75km radius (Z)					-0.0910***	-0.121***
					(0.0285)	(0.0413)
Observations	164462	164462	164462	164462	175141	175141
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

Table 3.8: Falsification Tests

	IV: Downwind Fire-events in 75 to 100km radius		IV: Upwind Fire-events in 75 to 100 km radius (shuffled location)	
	(1) **WFA-Z**	(2) **HFA-Z**	(3) **WFA-Z**	(4) **HFA-Z**
Trimester-1: Mean PM2.5 in 75km radius (Z)	-0.008	0.003	0.033	0.054
	(0.03)	(0.04)	(0.033)	(0.048)
Observations	179816	179816	179539	179539
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

Table 3.9: Mother's Conception Behaviour

	(1)	(2)
	Dummy for successful conception	
3 month exposure to upwind fire-events in 75-100kms	0.000307*** (0.000049)	
Exposure to upwind fire-events in 75-100kms in the month of conception		0.000151*** (0.000042)
Number of Unique Mothers	144833	144833
Includes Child, Mother & HH characteristics	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual OLS regression of log number of conceptions on controls mentioned in the table. Regressions include other controls for mother's education, father's literacy level, characteristics of household head and wealth index of the household.

Table 3.10: Heteregeneity: By Background Characteristics

	IV regression: Height-for-age Z score					
					Mother's Education	
	Poor	Rich	North	South	Primary or less	Secondary or higher
	(1)	(2)	(3)	(4)	(5)	(6)
Trimester-1: Mean PM2.5 in 75km radius (Z)	-0.139**	-0.0824	-0.120***	-0.0226	-0.176**	-0.0775
	(0.059)	(0.063)	(0.043)	(1.813)	(0.081)	(0.056)
Observations	84458	88932	139656	40159	76965	95642
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z on weighted mean PM2.5 in first trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

North Indian states: Arunachal Pradesh, Assam, Bihar, Chandigarh, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Jharkhand, Madhya Pradesh, Manipur, Meghalaya, Mizoram Nagaland, Delhi, Punjab, Rajasthan, Sikkim, Tripura, Uttar Pradesh and Uttarakhand.

South Indian states: Andhra Pradesh, Karnataka, Kerala, Maharashtra, Chhattisgarh, Odisha, Telangana, West Bengal, Lakshwadeep Islands, Andaman and Nicobar Islands, Dadar and Nagar Haveli, Daman and Diu, Puducherry and Goa.

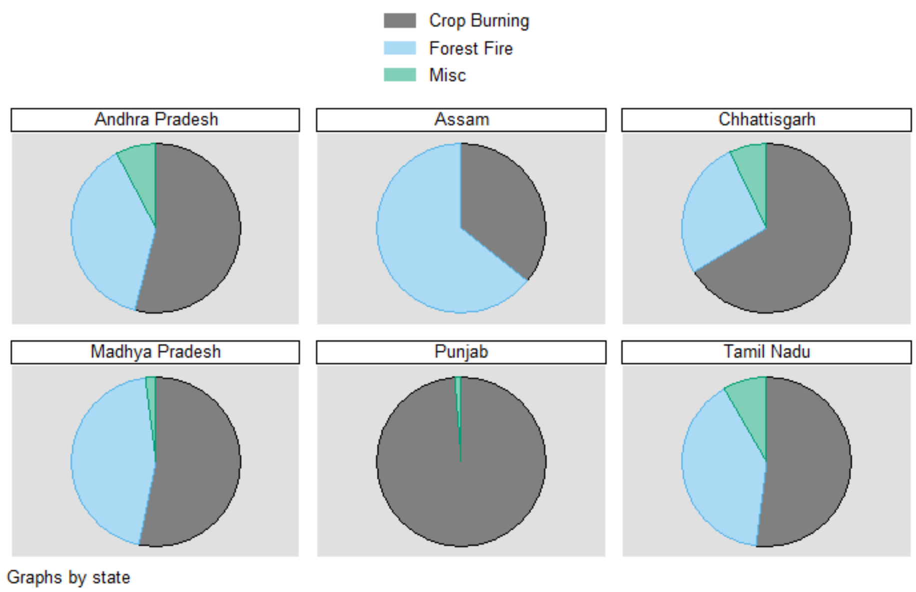
Table 3.11: Heteregeneity: By Age

	IV regression: Height-for-age Z score		
	Age of child		
	0 to 1 years (1)	1 to 2 years (2)	3+ years (3)
Trimester-1: Mean PM2.5 in 75km radius (Z)	-0.0564 (0.349)	-0.719** (0.338)	-0.110* (0.0588)
Observations	30456	28669	100701
Includes Child, Mother & HH characteristics	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester (z-score). Regressions include controls which are same as those mentioned in Table 3.2 notes.

Appendix

Figure A3.1: Biomass Burning Events by Type for Select States



Population weighted split of all biomass burning events which took place from 2010-2016 for select states.

Table A3.1: Missing PM2.5 Analysis

	(1) **HFA-Z**	(2) **Male Child**	(3) **Child's Age**	(4) **Mother's Education**	(5) **Source of water: Pipedwater**	(6) **Uses clean cooking fuel**	(7) **HH Openly Defecates**	(8) **Poor**
Dummy for missing PM2.5 information	-0.003	-0.003	-0.004	-0.017	0.001	-0.002	-0.002	0.002
	(0.014)	(0.004)	(0.004)	(0.032)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	223150	223150	223150	223150	223150	223150	223150	223150
Includes Child, Mother & HH characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster.
Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$.

Table A3.2: Full Model : IV Regression

	(1) WFA-Z	(2) HFA-Z
Trimester-1: Mean PM2.5 in 75km radius (z-score)	-0.103***	-0.116***
	(0.032)	(0.045)
Child is Male	-0.036***	-0.099***
	(0.006)	(0.008)
Birth Order	-0.046***	-0.062***
	(0.003)	(0.004)
Child's Age in Months	-0.082***	-0.106***
	(0.008)	(0.010)
Mother's Education (in years)	0.027***	0.031***
	(0.001)	(0.001)
Household Head's Age	0.002***	0.002***
	(0.000)	(0.000)
Houehold Head is Male	0.011	0.031**
	(0.010)	(0.014)
Mother's Age at Birth	0.048***	0.058***
	(0.005)	(0.007)
Mother's Age at Birth square	-0.001***	-0.001***
	(0.000)	(0.000)
Source of water is Pipedwater	0.016	0.005
	(0.011)	(0.015)
Uses clean cooking fuel	0.134***	0.140***
	(0.010)	(0.013)
Household Defecates in Open (OD)	-0.136***	-0.166***
	(0.009)	(0.013)
Observations	179816	179816
Includes FEs for Cluster, Month*Year & District*Year	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster.
Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$.

Chapter 4

Impact of Biomass Burning on Blood Pressure: A Study from North India

4.1 Introduction

Controlled burning in agriculture is an activity which is practised globally. It comprises of burning crop residues and clearing forests for agricultural land. This biomass burning activity is associated with increase in pollution levels and has serious health effects on both child (Rangel and Vogl (2018); Jayachandran (2009)) and adult health (Tan-Soo and Pattanayak, (2019); Frankenberg et al. (2005)). Although biomass burning is a global phenomenon, India stands out. Only China and United States America have higher carbon emissions associated with crop residue burning (FAO, 2010-2016). The pollution causing biomass burning can potentially have an adverse effect on cardiovascular health of the population. Currently India faces a growing burden of Non Communicable Diseases (NCDs) which poses a threat to sustainable development. Particularly, within NCDs cardiovascular diseases take a center stage, the global disease burden (GDB) study for India finds that cardiovascular diseases contributed to 28.1 percent of total deaths and 14.1 percent of total disability-adjusted life year (DALYs) in 2016.[1]

Epidemiological and clinical literature has shown that air pollution affects not just the respiratory health but cardiovascular health as well. Both short term (days) and long term exposure (years) to air pollution affect cardiovascular health which includes ischaemic heart

[1] Dandona et al., 2018

diseases, heart failure, stroke, cardiac arrhythmia, blood pressure (BP) among many other illnesses. Exposure to pollution can cause an elevation in arterial BP, and a rapid increase in BP could conspire to provoke instability or even rupture the atherosclerotic plaque (build up of fat inside arteries which causes narrowing of blood vessels) and therefore promote an acute cardiac event. This event could manifest itself in a heart attack or stroke. While such studies have shown pathways between pollution and cardiovascular health, our paper is the first to show the association between agricultural activity of crop burning and its potential harmful effect on cardiovascular health of the exposed population.

Biomass burning is a common agricultural activity in Northern India. Punjab is known as the "food basket" of India as it is one of the most agriculturally productive regions. It produces both rice and wheat, with rice crop being harvested in the months of October and November while wheat crop is harvested in April and May. The state of Haryana (adjacent to Punjab) practices similar agricultural routine. With the spread of mechanisation and irrigation facilities, farmers moved away from manual harvest to mechanised harvest (use of combine harvesters). A combination of these factors led to adoption of multi-cropping system by farmers which left little time in between the harvest of one crop and sowing of another. In this scenario, crop residue burning thus emerged as the quickest and cheapest way to get the farm ready for the next crop.

Studies have found that aerosols released in Punjab due to residue burning spread to the western and central Indo-Gangetic Plains (IGP) (Kaskaoutis et al. (2014)). The pollutants thus released affect the health and well being of millions of people who live in this region. The resulting smog has been linked to myriad negative health effects which includes infant mortality (Pullabhotla (2018), respiratory infections (Chakrabarti et al. (2019)) and even increased risk of lung cancer(Goss et al. (2014)). Additionally studies from developed nations have also looked at wildfires occurrence and its effect on cardiovascular health. These studies have mostly focused on big events like California wildfires or Australian bush fires and hospital admissions related to adverse respiratory or cardiac events (Haikerwal et. al (2015), Delfino et. al (2009), Finlay et. al (2012)). We contribute to this growing literature by looking at the effect of biomass burning on cardiovascular health in India. Agricultural burning in Indo-Gangetic region has increased over time which warrants serious attention as it directly poses a health risk to children, adults and elderly population.

To analyse the association between cardiac health and biomass burning activity (an activity which deteriorates air quality) in Northern India we focus on four states in Indo-Gangetic plains (Punjab, Haryana, Uttar Pradesh and Bihar). These four states form a geographically contiguous region with a natural variation in biomass burning activity. While many parts of Punjab and Haryana witness high levels of biomass burning, the states of Uttar Pradesh and Bihar experience biomass burning of relatively lesser intensity. We focus on blood pressure readings of individuals which have been collected as a part of Demographic and Health Survey (DHS, round 4) for India to classify individuals into hypertensive and non-hypertensive categories. We do this exercise by combining satellite data on biomass burning events with geo-coded cluster location information in DHS-4. By merging these two datasets together we construct a measure of exposure to biomass burning in the neighbourhood of an individual. We find that exposure to extreme or high intensity biomass burning in last 30 days is associated with an increase in likelihood of being hypertensive by 1 percent. The effect is even more pronounced for older cohorts (aged 40 and above) for both males and females, with probability of being hypertensive due to exposure to high intensity biomass burning being 5.8 and 3.2 percent respectively. India has a growing burden of non-communicable diseases (NCDs) and cardiac diseases are a major contributor to the growing pool of disability adjusted life years associated with NCDs. While many diseases under NCDs have been associated with lifestyle but some proportion of these diseases are related to the environmental quality as well. The aim of this study is to bring focus on biomass burning which causes air pollution and elimination of which can accrue many benefits in form of reducing the burden of NCDs which translates into a monetary saving to the tune of USD 520 to 675 Million over a course of five years.

The paper has been organised as follows - section 2 gives a background of biomass burning in India and prevalence of cardiovascular diseases in four focus states, section 3 describes various datasets which have been used in this study, section 4 explains the empirical methodology that has been adopted to conduct the analysis, sections 5 and 6 discuss the results and related economic benefit associated with elimination of biomass burning. Section 7 concludes.

4.2 Background

4.2.1 Biomass Burning in India

India has a substantial amount of land under cultivation(60%) and under forest cover(25%), with majority biomass burning events taking place in these areas. Over the past few decades, Indian agriculture has been marked with expansion of irrigation facilities, adoption of high yield variety seeds and increased mechanisation (like use of combine harvester). A combination of these factors led to adoption of multi-cropping system by farmers which leaves little time in between the harvest of one crop and sowing of another. In this scenario, crop residue burning has emerged as the quickest and cheapest way to get the farm ready for the next crop. Cereals are the prime contributor to crop burning activity in India, with rice and wheat crop residue burning forming the major chunk of residue burning process (Jain et al (2014)). Two major residue burning seasons are thus related to crop harvest seasons: kharif crop harvest (rice stubble burning) which takes place in the months of October and November; and rabi crop harvest (wheat straw burning) which happens in the months of March to May. This activity is quite prevalent in Northern states of India especially in Punjab and Haryana.

Biomass burning in India is not limited to just crop residue burning, it covers forest fires as well. Forest fires or wildfires are caused by various factors acting in conjunction with each other. These factors include availability of biomass (dry vegetation) and appropriate climatic conditions (high temperature, low pressure, windy conditions). Forest Survey of India[2] lists vulnerable months for each state when forest fires are most likely to happen, which mainly span the high temperature months from March to June. Wildfires happen due to both intentional and unintentional human activity. In many states of India, the practice of slash and burn is rampant wherein vegetation in forests is cut and then burned to clear the piece of land for human use. In a lot of cases unintentional human activities like leaving active cigarette butts behind in open forests lead to forest fires. Other natural factors which cause forest fires include lightening which produces a spark to start a fire in dry vegetation.

In developed countries, forest fires are mainly responsible for the carbon content release due

[2] See Vulnerability of Forest Fire (2017), technical report by Forest Survey of India.

to biomass burning; however, in case of India (and other South Asian countries) crop residue burning contributes the most to total carbon release. In South Asia, India stands out both in terms of total area burned (4.5 million hectares burned in 2015) and in terms of total carbon content (1.5 million metric tonnes) released due to biomass burning. Biomass burning is a major source of pollution as it releases harmful pollutants like Carbon Dioxide(CO_2), Carbon Monoxide (CO), Sulphur Oxides and particulate matter (PM) in the atmosphere which can have harmful effect on human health. The release of harmful pollutants in the atmosphere is captured by aerosol loading[3] in the region. Studies have found that aerosol loading increases in the regions which lie in the upwind path (smoke blowing away from source towards a particular region) and in the vertical direction as well. Kaskaoutis et. al (2014) find that crop burning in Punjab has an effect on aerosol properties of the Indo-Gangetic Plains, also particulate matter (PM2.5) concentrations increase near ground surface and the concentration of pollutants fall as we move from west to east India. Fine particulate matter released during biomass burning incidents have long range travel properties and affect not just the local areas but far away regions as well.

In Figure 4.1 we plot split of biomass burning events (for years 2015 and 2016) into crop burning and forest fire activity for four focus states - Punjab, Haryana, Uttar Pradesh and Bihar. Geo-coded biomass burning events have been projected onto land mask cover for India to categorise each fire event as an event which happens in a forest area vs cropped area. The state of Punjab witnesses highest number of fire-events(35086), followed by Uttar Pradesh(8298, Haryana(7037) and Bihar(1797). In case of Punjab and Haryana almost all biomass burning activity is related to crop fires (98.5%), while in case of Uttar Pradesh and Bihar forest fires also contribute to the pool of vegetation fires (20.4% and 12.4% respectively).

4.2.2 Cardiovascular Health in four focus states

The global disease burden (GBD) study for India (Dandona et. al (2017)) finds that Punjab and Haryana belong to the high and higher-middle epidemiological transition level (ETL) group. In other words, between 2000 and 2016, the burden of diseases in these states shifted

[3]Aerosol loading is the suspensions of solids and/or liquid particles in the air that we breathe. Dust, smoke, haze are also part of aerosol loading

away from communicable to non-communicable diseases substantially. Within non-communicable diseases, when we look at cardiovascular diseases (Ischaemic heart disease and Cerebrovascular disease), then we find that both Punjab and Haryana are among the top states with highest number of Disease Adjusted Life Years (DALYs) being associated with Ischaemic heart disease.[4] We include two more states in our analysis which lie in Northern India. These states are in the contiguous region of Indo-Gangetic plains and belong to low ETL group, in these states the burden of disease is still concentrated more towards communicable diseases rather than non-communicable diseases. The GBD study shows that the DALYs associated with cardiovascular diseases in Uttar Pradesh and Bihar are lower than the national average.

4.3 Data

4.3.1 Demographic Data

We use data from the latest round of the Demographic and Health Survey (Round-4 for 2015-16) for India. The DHS-4 collected data on prevalence of hypertension by measuring blood pressure for women and men in the age group 15-49 years and 15-54 years respectively. It provides consistent and reliable estimates of blood pressure indicators (Systolic and diastolic measurements) at the national, state and district level. Omron Blood Pressure Monitor was used to take three readings for each individual with a five-minute interval in between each reading. We select four North Indian states which lie in Indo-Gangetic (IG) plains for our analysis which exhibit differential biomass burning patterns. In our sample states, Punjab and Haryana have extremely high levels of biomass burning, followed by Uttar Pradesh and Bihar.

The selection of the survey sample followed a multi-stage sample design. First, the sampling frames were developed on the basis of non-overlapping units of geography, which were identified as the primary sampling units (PSUs), by states and urban and rural areas within each state, and then a fixed proportion of households were selected using systematic sampling within each PSU (interchangeably referred to as cluster). The NFHS round 4 also provides

[4]DALYs associated with Cerebrovascular diseases for Punjab and Haryana are 979 and 1311 per 100000.

GPS coordinates of the sampled PSUs which we use in our analysis. To ensure respondent confidentiality the GPS locations of the PSUs have been displaced randomly by 5 kilometres for rural clusters and by 2 kilometres for urban clusters.[5]

The survey includes data on 211,152 individuals from the four states of Punjab, Haryana, Uttar Pradesh and Bihar (Figure 4.2). For our study we select only those individuals for whom all three readings were available. We remove individuals from our estimation who already report taking a medicine for hypertension, females who are pregnant, individuals who have missing covariates (like BMI) information and individuals with error prone reading for systolic (greater than 80 mmHg) or diastolic (less than 50 mmHg).

We followed European Society of Hypertension (ESH) and European Society of Cardiology (ECH) guidelines to categorize blood pressure readings into hypertensive and non-hypertensive categories. An individual is classified as hypertensive if he/she has a systolic blood pressure (SBP) level greater than or equal to 140 mmHg, or a diastolic blood pressure (DBP) greater than or equal to 90 mmHg[6]. While three readings are recorded for each individual in the survey, we use the average of last two readings.[7] .

The survey also captures risk factors which are associated with hypertension: age group and gender of an individual, consumption of alcohol, smoking behaviour, body mass index (BMI), whether an individual is educated (primary education or above), use of clean cooking fuel used by the household (LPG or induction), wealth index of the household that an individual belongs to and type of residence of the household (rural or urban).

4.3.2 Fire-events Data

The GPS location of the clusters is combined with NASA's Fire Information for Resource Management System (FIRMS) data which captures real-time active fire locations across the globe. The FIRMS data that we use is called MODIS (MODerate Resolution Imaging Spectro radiometer) data and it records fire incidents at pixel level where each pixel is identified by a latitude and longitude reading. Each latitude (and longitude) is the center of a 1 km fire pixel (1 km X 1 km in size). This data records not just the location of a fire but also the

[5]One percent of the clusters are displaced by as much as 10 kilometres.
[6]Following Williams et al., 2018
[7]Following Jose, Arun Pulikkottil, et al. (2019).

brightness (temperature) of fire (in Kelvin units) and date and time when the incident was picked by the Terra satellite. The MODIS data is available on a daily basis since November 2000 and NASA reports that the fires captured by this dataset are mostly vegetation fires. NASA data on fire incidents also provides a variable confidence, which depicts the quality of the observations and it ranges from 0-100. We use this variable to construct a probability weighted count of fire-events around a cluster (100 km radius) for the last 30 days from the date of interview of the individual. The fire-count variable is constructed in the following way: total number of probability weighted fire-events are calculated in the 100 km radius around the cluster location, from this probability weighted total count of fire-events in 10 km radius is subtracted. This is done to take into account uncertainty of true cluster location due to displacement of clusters. Figure 4.3 shows a district wise distribution of mean number of fire-events for our estimation sample. The districts in the Bihar have much lower incidence of fire-events (less than 30) as depicted by the lighter shade. As we move westwards towards Punjab and Haryana we notice much higher incidence of these events (with many districts having more than 100 fire-events) as depicted by darker shades.

Our level of analysis is an individual sampled in a cluster during the survey. In Figure 4.4 the histogram shows the distribution of exposure to fire-events during last 30 days from the date of survey. The mean number of fire-events for the estimation sample is 47.11 and around 10 percent of the surveyed individuals are exposed to high intensity biomass burning, experiencing more than 100 fire-events during last 30 days from the date of interview. Our key exposure variable is high intensity biomass burning (HIB) which is a dummy variable which takes value one if number of fire-events in the neighbourhood (described above) during last 30 days for an individual was greater than 100 and zero otherwise. These are cases which lie in the extreme right tail in the distribution of fire-events for our estimation sample.[8] Combining demographic data with data on fire-events and weather factors like temperature and rainfall results in a final sample of 188,190 individuals from 6809 PSUs with complete information (Figure 4.2).

[8]We provide results for other cut-off levels in our robustness checks.

Estimation sample summary

In Table 4.1 we provide a description of our estimation sample. Individuals exposed to high intensity biomass burning (HIB = 1) have a higher prevalence of hypertension (9.9 percent) in comparison to the group which was exposed (fire events less than 100 OR HIB = 0) to low intensity biomass burning (8.1 percent). The mean systolic and diastolic blood pressure are also higher for individuals exposed to high intensity biomass burning. The age distribution is almost identical in the two groups, with almost 20 percent of the sample in both the groups belonging to individuals in the age group 40 to 54 years. The proportion of younger cohort (age 20 to 39 years) is slightly higher in the group which was exposed to extreme biomass burning, while the proportion of youngest cohort (15 to 19 years) is higher in the other group which didn't have exposure to intense biomass burning.

In terms of other demographic and risk factors, the high intensity biomass burning group has more males in the sample and the individuals in this group also have a higher BMI. Close to 5 percent of the individuals in the HIB group report consuming alcohol as against 3.5 percent in the other group (that is HIB = 0); the self-reported smoking behavior is however of greater magnitude in the low intensity exposure group than in the high intensity exposure group. A greater proportion of individuals in the high intensity exposure group are educated (studied till primary level) and use clean cooking fuels (use of dirty cooking fuels is a known risk-factor). Close to 75 percent of the sample in low intensity exposure group lives in rural areas while the corresponding figure for the exposed group is smaller at 64 percent. Lastly the wealth distribution is different between the two groups, with a greater proportion of individuals in high intensity exposure group belonging to middle, richer and richest category. The low intensity exposure group is more evenly split across all wealth classes. To account for these differences between the high intensity exposure group (HIB = 1) and low intensity exposure group (HIB = 0) group we control for all demographic and risk-factors in our statistical analysis.

In Figure 4.5, we plot a polynomial relationship between probability of being hypertensive and age of an individual. We do this for two groups (high intensity and low intensity exposure group) which have differential exposure to fire-activity in last 30 days. As documented in medical literature the probability of being hypertensive increases with age, this is reflected in the upward sloping curves. However we find that individuals who have low intensity exposure to fire-activity (HIB = 0) have a lower probability of being hypertensive than individuals who were exposed to high intensity fire-activity (fire events greater than 100, HIB = 1) across all ages.

4.3.3 Other data

We also use rich spatial data on pollution (PM2.5), which has been estimated using satellite data (van Donkelaar et al., 2012). We convert aerosol data retrieved from MISR (Multiangle Imaging SpectroRadiometer) to PM2.5 using a conversion factor (Dey et al., 2012 and van Donkelaar et al., 2010). The PM2.5 data after processing is available at monthly frequency with a 0.1 * 0.1 degree resolution (10km*10km grid). To account for weather conditions, we also use rainfall and temperature data from MODIS-Terra LST dataset.

4.4 Empirical Methodology

In this section, a multivariate logistic regression model is used to examine the association between hypertension and occurrence of high intensity biomass burning (HIB) in the neighbourhood. Our binary outcome of interest is Y, which indicates whether the blood pressure of an individual belongs to hypertensive category ($Y = 1$) or not ($Y = 0$).

$$\Pr(\mathrm{Y} = 1 \mid \mathrm{HIB},\ \mathrm{X}) = G(\alpha + \beta HIB_{icdmt} + \theta X_{icdmt} + \gamma_d + \rho_m + \lambda_t) \tag{4.1}$$

where G is the cumulative density function of logistic distribution.

We look at individual i, from cluster c, belonging to district d, surveyed in month m and year t. HIB_{icdmt} is the dummy variable which represents exposure to high intensity of biomass burning during last 30 days from the survey date. We control for multiple risk factor in X_{icdmt} like age group, gender, BMI, smoking and alcohol consumption behaviour and education of an individual. Additionally we also control for wealth index, type of cooking fuel and rural-urban status at the household level. The marginal effect of interest has the same sign as β which depicts the effect of short-term exposure to HIB on probability of being hypertensive.

Since our analysis is descriptive in nature, our estimation strategy accounts for the sampling design where sample weights were used in the estimation of the coefficients. Errors were clustered at the cluster level to account for potential correlations between observations within the same cluster. In addition, we also include district fixed effects to account for unobserved regional differences. We also account for month and year fixed effects to account for any seasonality in the data. The errors have been clustered at the cluster level throughout our analysis.

The mechanism behind our hypothesis lies in the exposure to pollution due to biomass burning or fire-events in the local neighbourhood. We use cluster-month-year panel for time-period 2015-16 to show that the month level changes in pollution as measured by PM2.5 are affected by changes in fire-events after controlling for weather conditions which play an important role in determining local pollution levels.

4.5 Results

4.5.1 Biomass Burning and Hypertension

We now explore the association between hypertension and high intensity biomass burning using logistic regression as outlined in equation 1 in previous section. We observe in Table 4.2, column 1 that marginal effect is positive and significant, that is being exposed to more than 100 fire-events in last 30 days (high intensity biomass burning) increases the probability of being hypertensive by around one percent.[9] In line with the literature, we find (results presented in appendix Table A4.1) that as age increases the probability of being hypertensive also increases, individuals in older age groups (40 years and above) are 10 percent more likely to be hypertensive than younger individuals (age 15-19 years). Males are also more likely to be hypertensive than females. Consumption of alcohol also increases the chances of being hypertensive. Furthermore, Body mass index is a strong predictor of hypertension, as an increase in BMI increases the likelihood of being hypertensive. The relationship between hypertension and wealth is found to be insignificant. Smokers generally have low body weight

[9]We also alternatively use a continuous variable for exposure to biomass burning and find that it is positively associated with probability of being hypertensive.

due to which on an average we observe that there is lesser incidence of hypertension in smokers in comparison to the non-smoking group. Educated individuals have lesser probability of being hypertensive while rural households and households which use clean cooking fuel are found to have slightly higher chance of being hypertensive.

4.5.2 Biomass Burning and Pollution

The mechanism through which biomass burning or fire-events have a physiological effect is pollution. In Table 4.3 we show how local pollution levels as measured by PM2.5 is related to fire activity in the region. We create a panel of all clusters (6809 in total) over a 24-month period (January 2015 to December 2016) and notice that areas which had high intensity of fire-activity (HIB = 1) vs areas with low fire-activity (HIB = 0) had a significant difference in the PM2.5 levels (PM2.5 is higher by $11.58ug/m^3$ in areas which experienced HIB).

4.5.3 Dose Effect

We now analyse the dose effect relationship between hypertension and exposure to fire-events. We split the population into multiple groups with varying levels of exposure to fire-events. The group which no exposure to fire-events is the base category while other groups have increasing levels of exposure to fire-activity. We present our results in Table 4.4 where we find that the coefficient for all groups with a positive exposure to fire-activity (fire-events > 0) is positive which indicates higher risk of being hypertensive as exposure to fire-events increases. However these coefficients are insignificant and we see an increase in magnitude (and a significant effect) for the group which had exposure to high intensity biomass burning, that is fire-events greater than 100. The estimate for this group is 1.4 percent, that is the probability of being hypertensive is 1.4 percentage points higher for this group in comparison to the base group (group with no exposure to fire-activity).

4.5.4 Age Interacted Effect

Next, we focus on differential impact of exposure to high intensity biomass burning on hypertension based on age of individuals. We conduct this analysis by interacting HIB with a continuous age variable. In Figure 4.6 we plot the marginal effect of HIB on probability of being hypertensive by age groups (5 year bins, for example 15-19, 20-24 to 49 to 54). We observe that effect becomes significant for older cohorts, that is individuals who are aged 40 years and above. we use this 40 year cut-off for further analysis in heterogeneity section.

4.6 Sensitivity Analysis and Robustness Checks

As a sensitivity check we use multiple cut-offs to define high intensity biomass burning, that is in addition to the original 100 cut-off we also use 200 and 300 as cut-offs to see if our results our sensitive to choosing a particular value for describing high intensity fire activity areas. In Table 4.2, column 2 and 3 we show that using these alternate higher cut-offs for HIB, that is being exposed to higher number of biomass burning events (200 or 300) in last 30 days is associated with an increase in likelihood of being hypertensive (1.8 % and 1.4% respectively).

Next we vary the radius of our analysis to alternate outside radii of 75 kilometer, to establish that our results are not driven by the choice of a particular radius for analysis. We find that changing our radius of analysis from 100 kilometers to 75 kilometers doesn't change our results (Table 4.5 column 1), we still find that individuals who live in areas which had higher incidence of fire-events in last 30 days are more likely to have blood pressure readings corresponding to hypertensive category.

We also expand our dataset to include individuals who have already been diagnosed as being hypertensive and who take medicines for this ailment. These individuals are treated as hypertensive individuals in our analysis even if their blood pressure readings fall in the normal range. In Table 4.5 (column 2) we find that inclusion of these individuals does not change our original findings, that is we still find that being exposed to high intensity biomass burning is associated with a one percent increase in likelihood of being hypertensive.

4.6.1 Heterogeneity

We now divide our sample by age and gender (and place of residence) to conduct sub-sample analysis in Table 4.6. We limit our sample to younger (age below 40 years) and older cohort (age above 40 years) for females and males separately and observe that for both female and male sub-population its the older cohort which drives the results. Females aged above 40 years are 3.2 percent more likely to be hypertensive if they are exposed to HIB in last 30 days (for males the corresponding figure is 5.8 percent). Its the older population which is most vulnerable to adverse health effect due to high intensity biomass burning. We notice a similar pattern for rural versus urban areas as well where again its the older cohort which has a greater probability of being hypertensive due to high biomass burning in the neighbourhood.

4.7 Economic Benefit from Elimination of Biomass Burning

In this section we estimate the value of economic benefit if high intensity biomass burning is eliminated. We do this by focusing on three states from our sample Punjab, Haryana and Uttar Pradesh which experience high intensity biomass burning (No individuals in Bihar had exposure to high intensity biomass burning). We follow Chakrabarti et. al (2019) in Table 4.7 to estimate the economic benefits attributable to elimination of biomass burning. We begin by documenting disability-adjusted life year (DALY) rates for each state for cardiovascular diseases (cerebrovascular diseases and Ishemic heart disease) which are obtained from India's State Level Disease Burden study. The GBD study (Dandona et. al (2017)) provides 95% lower and upper confidence interval values for DALY's for each state of India. These DALY rates when multiplied by the total state population reflect the total burden in terms of DALYs lost due of cardiovascular diseases in each of these states. We also calculated the proportion of hypertension cases which would be averted in the population if everything else remained same but high intensity biomass burning was eliminated. When this estimate is multiplied with total burden of cardiovascular diseases then we obtain the number of DALYs which will be saved if high intensity biomass burning is eliminated (Table 4.7 row 4). Lastly, these DALYs saved are converted into monetary terms by multiplying by per capita state GDP. Using a discount factor of three percent we compound economic benefit over a 5-year period. Based on our estimates, we find that eliminating biomass burning would avert 70 to 91 disability-adjusted life years lost per year leading to a saving of USD 520 to 675 Million over five years for states of Haryana, Punjab and Uttar Pradesh[10].

4.8 Conclusion

Our analysis shows that high intensity biomass burning is associated with hypertension. We find that short term exposure to extreme biomass burning events increases the likelihood of being hypertensive by one percent. The most vulnerable population is that of older cohort (aged 40 and above), where both females and males have higher probability of being hypertensive (3.2% and 5.8% respectively) if they were exposed to extreme biomass burning during last 30 days. The underlying mechanism behind these results is pollution and we show that high intensity biomass burning (that is occurrence of more than 100 fire-events in last 30 days) increases PM2.5 by a substantial amount.

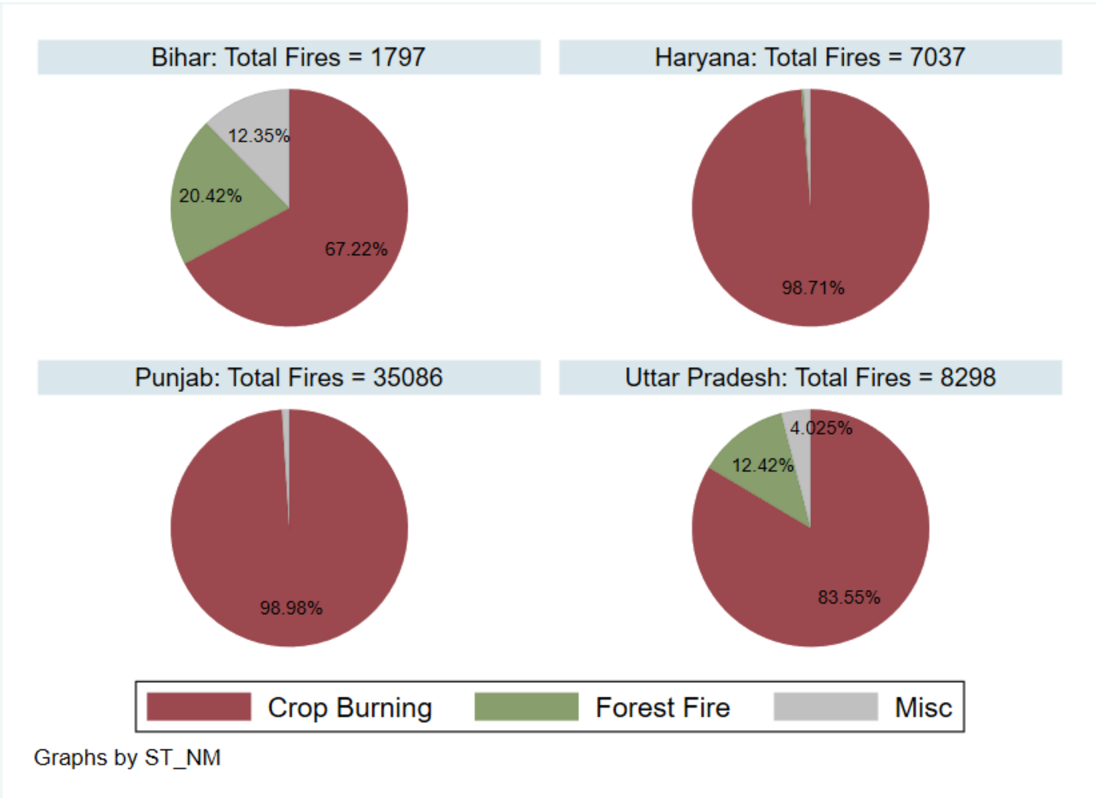
India needs effective policies regarding regulation and management of biomass burning. The government has committed itself to subsidising the use of happy-seeder technology (this is an alternative to combine harvester, it leaves rice residue in form of a mulch on farm which doesn't hamper wheat crop sowing and hence doesn't require burning), however the uptake of this policy remains quite low due to high initial investment in the machine (Gupta (2011)). Under a new revised scheme government has allocated 1150 crore INR (125 Million USD) for curbing crop burning in 4 North Indian states (Punjab, Haryana, Uttar Pradesh and National Capital Region) by adopting a three pronged approach - providing machines at a subsidized rate, higher monitoring effort on part of government officials and creating greater awareness about harmful effects of crop burning by information campaigns. Such multi-pronged policies are needed urgently to eliminate both short term and long term negative effects of biomass burning. Additionally, effective management of forest fires is needed; however, the budget allocation for this purpose is really small and remains unused in every financial year. Our analysis finds that elimination of biomass burning can lead to a saving of 520 to 675 Million USD over a period of 5 years.

[10] Adding columns 1, 3 and 5 to obtain the lower estimate and adding columns 2, 4 and 6 to obtain the higher estimate.

Figures and Tables for Chapter 4

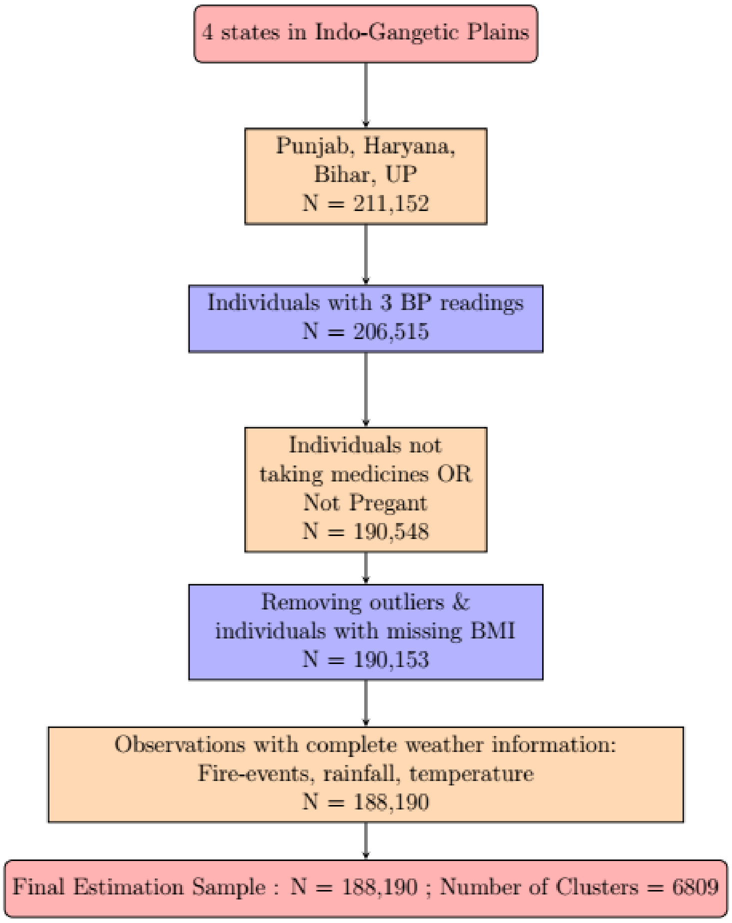
Tables and Figures

Figure 4.1: Biomass Burning by Type of Land



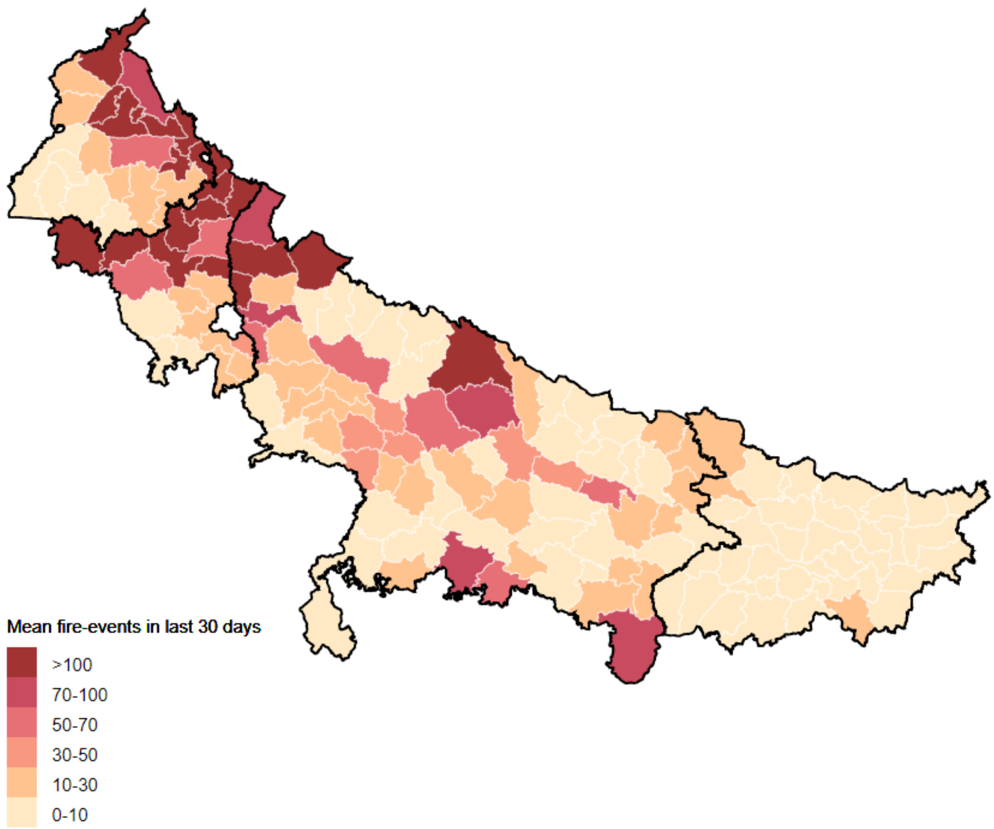
Biomass burning events categorized into type of fire on the basis of type of land in which they occur for four states of India - Punjab, Haryana, Uttar Pradesh and Bihar. Data for biomass burning events (raw count of fire-events) for years 2015 and 2016.

Figure 4.2: Estimation Sample



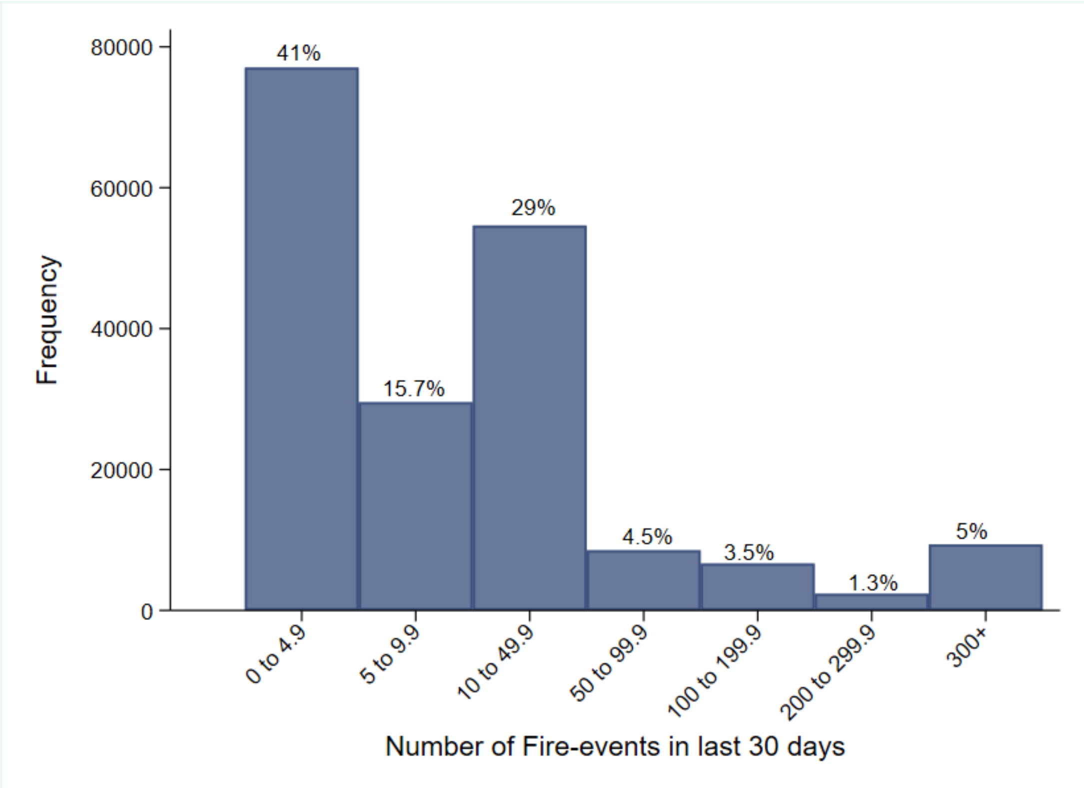
Our estimation sample consists of individuals beloging to 4 states which lie in the Indo-gangetic region. We remove individuals who report taking a medicine for hypertension or those who are pregnant. The final sample consists of 188,190 males and females belonging to 6809 clusters who have complete information on blood pressure, BMI and local weather conditions.

Figure 4.3: District Level Fire-activity in Last 30 Days



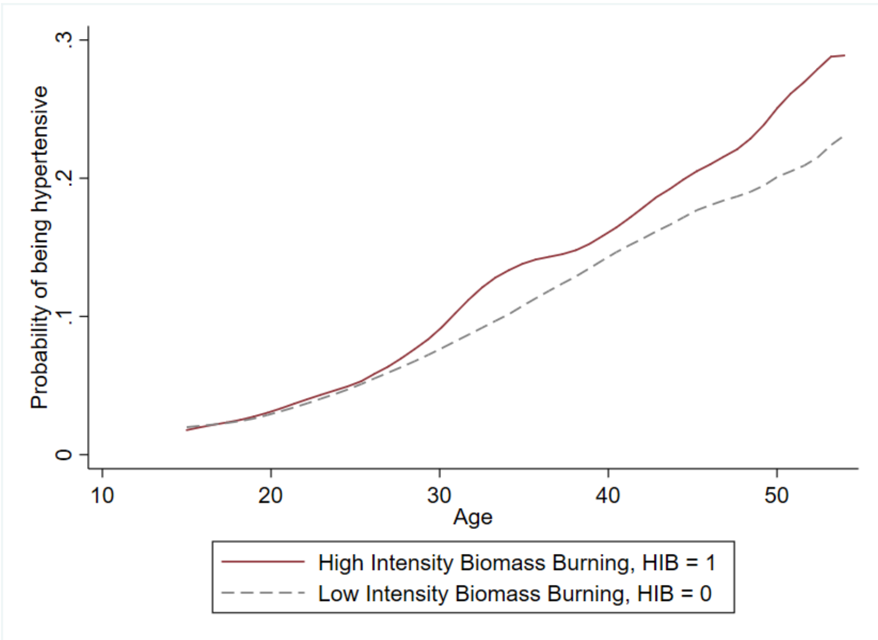
Mean number of fire-events are calculated for all 150 districts in our sample which belong to 4 states in the Indo-Gangetic region (Haryana, Punjab, Uttar Pradesh and Bihar). The mean is calculated by using individual level information about number of fire-events occurring in 100 km radius around the cluster location for last 30 days. Cluster is the village or the block that an individual resides in.

Figure 4.4: Individual Level Fire-activity in Last 30 Days



For each individual total number of fire-events in 100 km radius around a cluster location for last 30 days are calculated. Around 10 percent of individuals in our sample live in areas which have high intensity of biomass burning (number of fire-events >= 100).

Figure 4.5: Hypertension and Age Relationship by Exposure to Fire-activity



High intensity biomass burning refers to exposure to more than 100 fire-events during last 30 days while Low intensity biomass burning refers to exposure to less than 100 fire-events during last 30 days.

Figure 4.6: Conditional Marginal Effects of Exposure to HIB on Hypertension for Different Ages.

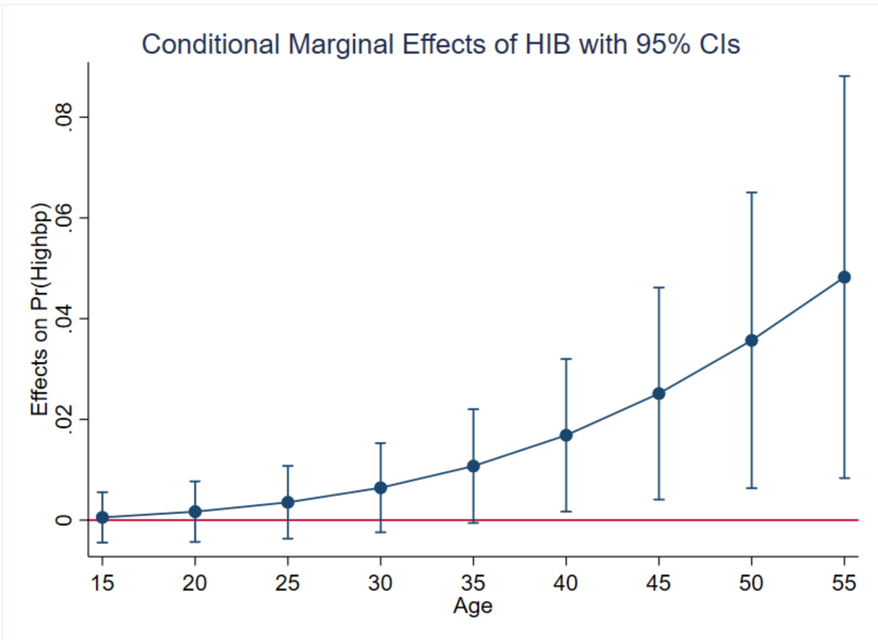


Table 4.1: Summary Statistics

	High Intensity Exposure HIB = 0		Low Intensity Exposure HIB = 1		
variable	Mean	SD	Mean	SD	p-value
Hypertenion (%)	8.1	27.3	9.9	29.8	<0.01
Systolic BP	114.3	12.9	116.0	13.2	<0.01
Diastolic BP	76.1	9.6	76.9	9.6	<0.01
Age-Group (%)					
50-54 years	0.8	9.1	1.0	9.8	<0.10
45-49 years	9.6	29.4	9.6	29.4	<0.99
40-44 years	10.4	30.5	10.9	31.2	<0.05
35-39 years	12.2	32.7	12.9	33.5	<0.01
30-34 years	13.1	33.7	14.3	35.0	<0.01
25-29 years	15.2	35.9	16.8	37.4	<0.01
20-24 years	17.0	37.6	17.9	38.3	<0.01
15-19 years	21.8	41.3	16.7	37.3	<0.01
Demographic Characteristics					
Gender = Male (%)	12.9	33.5	17.9	38.4	<0.01
Consumes alcohol (%)	3.4	18.2	4.9	21.6	<0.01
Smokes (%)	10.7	30.9	9.0	28.7	<0.01
BMI	21.4	4.0	22.5	4.4	<0.01
Educated (%)	66.6	47.2	79.7	40.2	<0.01
Uses clean cooking fuel (%)	31.0	46.2	51.8	50.0	<0.01
Rural (%)	74.9	43.4	64.0	48.0	<0.01
Wealth Index (%)					
Richest	19.7	39.8	48.7	50.0	<0.01
Richer	16.4	37.1	21.6	41.1	<0.01
Middle	17.1	37.6	13.9	34.6	<0.01
Poorer	20.5	40.4	7.6	26.5	<0.01
Poorest	26.3	44.0	8.2	27.5	<0.01
Observations	169796		18394		

Source: National Family Health Survey, 2016

Table 4.2: Hypertension and Fire-activity: Marginal Effects

	Fire events >100 (1)	Fire events >200 (2)	Fire events >300 (3)
High Intensity Biomass Burning	0.0099**	0.018***	0.014**
	(0.0049)	(0.0058)	(0.0066)
Other Ind and HH Level Controls	Yes	Yes	Yes
District, Month and Year Fixed Effects	Yes	Yes	Yes
Temperature & Rainfall controls	Yes	Yes	Yes
Observations	188190	188190	188190

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Other controls not shown in the table as same as those in Table A4.1. Sample weights have been used in all regressions.

Table 4.3: Pollution and Fire-activity

	Mean PM2.5 in 100km radius
High Intensity Biomass Burning	11.6***
	(0.24)
Mean temperature	-2.08***
	(0.040)
Mean Rainfall	-0.10***
	(0.0017)
Observations	158592
Unique Clusters	6809
Cluster Fixed Effects	Yes
Month Fixed Effects	Yes
Year Fixed Effects	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$.

Table 4.4: Dose Effect: Marginal Effects

	Hypertension
Fire-events equal to 0	Base Category
Fire-events greater than 0 but less than 10	0.0039
	(0.0044)
Fire-events between 11 and 50	0.0043
	(0.0052)
Fire-events between 51 and 75	0.0065
	(0.0076)
Fire-events between 76 and 100	0.00067
	(0.0090)
Fire-events greater than 100	0.014*
	(0.0074)
Other Ind and HH level Controls	Yes
District, Month and Year Fixed Effects	Yes
Temperature & Rainfall controls	Yes
Observations	188190

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Other controls not shown in the table as same as those in Table A4.1. Sample weights have been used in all regressions.

Table 4.5: Robustness Check: Alternate Analysis Radii and Expanded Estimation Sample (Marginal Effects)

	Hypertension	
	(1)	(2)
	Analysis radius 75km	Includes individuals taking medicines for HBP
High Intensity Biomass Burning	0.015***	0.010*
	(0.0056)	(0.0055)
Observations	188190	193413
Individual & HH level controls	Yes	Yes
Rainfall & Temperature control	Yes	Yes
District, Month & Year fixed effect	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Other controls not shown in the table as same as those in Table A4.1. Sample weights have been used in all regressions.

Table 4.6: Heterogeneity: Gender and Place of residence (Marginal Effects)

	Females			**Males**		
	All Ages	Age <40	Age >= 40	All Ages	Age <40	Age >= 40
	(1)	(2)	(3)	(4)	(5)	(6)
High Intensity Biomass Burning	0.0071	0.0017	0.032**	0.011	-0.0062	0.058*
	(0.0054)	(0.0050)	(0.015)	(0.014)	(0.013)	(0.033)
Individual & HH level controls	Yes	Yes	Yes	Yes	Yes	Yes
District, Month and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Temperature & Rainfall controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	163043	130162	32881	25147	18808	6339

	Rural			**Urban**		
	All Ages	Age <40	Age >= 40	All Ages	Age <40	Age >= 40
	(1)	(2)	(3)	(4)	(5)	(6)
High Intensity Biomass Burning	0.0051	0.0014	0.031*	0.014	0.0052	0.043
	(0.0061)	(0.0055)	(0.017)	(0.0096)	(0.0089)	(0.028)
Individual & HH level controls	Yes	Yes	Yes	Yes	Yes	Yes
District, Month and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Temperature & Rainfall controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138894	110046	28848	49296	38835	10328

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Other controls not shown in the table as same as those in Table A4.1. Sample weights have been used in all regressions.

Table 4.7: Economic Benefit from Elimination of HIB

	Haryana		**Punjab**		**UP**	
	(1) Haryana (LCI)	(2) Haryana (HCI)	(3) Punjab (LCI)	(4) Punjab (HCI)	(5) UP (LCI)	(6) UP (HCI)
1. DALY rates for Hypertension Diseases per person [a]	0.025	0.032	0.041	0.051	0.015	0.019
2. State population (in Millions) [b]	25.35	25.35	27.74	27.74	199.8	199.8
3. Proportion of Hypertension cases attributed to biomass burning [c]	0.030	0.030	0.020	0.020	0.010	0.010
4. DALYs saved (Million years) [d]	0.019	0.0250	0.022	0.028	0.028	0.038
5. Per capita GDP ($/per person) [e]	2527	2527	1860	1860	722	722
6. Economic Value to DALYs saved per year ($ Million/year) [f]	48.3	63.1	41.6	52.6	20.4	27.3
7. Economic Value to DALYs saved over 5 years ($ Millions) [g]	228	297	196	248	96	129
Total ($ Millions)			[520 , 675]			

LCI, Lower 95% Confidence Interval Value; HCI, Higher 95% Confidence Interval Value
DALY, disability-adjusted life years; GDP, gross domestic product.
a From The India State-Level Disease Burden Initiative, 2017. DALYs for Hypertension.
b From Indian Population Census 2011 (Office of the Registrar General & Census Commissioner 2011).
c From PUNAF after estimating equation 1
d Row 1*Row 2*Row 3.
e From RBI (For year 2015-16).
f Row 4*Row 5
g From Row 6 for 5 years discounted at 3% per year.

Appendix

Table A4.1: Full Model Results - Hypertension and Fire-activity : Marginal Effects

	Fire events >100 (1)	Fire events >200 (2)	Fire events >300 (3)
High Intensity Biomass Burning	0.0099**	0.018***	0.014**
	(0.0049)	(0.0058)	(0.0066)
15-19 years	(Base category)	(Base category)	(Base category)
20-24 years	0.027***	0.027***	0.027***
	(0.0036)	(0.0036)	(0.0036)
25-29 years	0.053***	0.053***	0.053***
	(0.0035)	(0.0035)	(0.0035)
30-34 years	0.079***	0.079***	0.079***
	(0.0034)	(0.0034)	(0.0034)
35-39 years	0.100***	0.100***	0.100***
	(0.0034)	(0.0034)	(0.0034)
40-44 years	0.12***	0.12***	0.12***
	(0.0034)	(0.0034)	(0.0034)
45-49 years	0.14***	0.14***	0.14***
	(0.0035)	(0.0035)	(0.0035)
50-54 years	0.13***	0.13***	0.13***
	(0.0061)	(0.0061)	(0.0061)
Female	(Base category)	(Base category)	(Base category)
Male	0.040***	0.040***	0.040***
	(0.0027)	(0.0027)	(0.0027)
Consumes alcohol	0.018***	0.018***	0.018***
	(0.0035)	(0.0035)	(0.0035)
Smokes	-0.0062**	-0.0061**	-0.0062**
	(0.0025)	(0.0025)	(0.0025)
BMI	0.0077***	0.0077***	0.0077***
	(0.00019)	(0.00019)	(0.00019)
Uses clean cooking fuel	0.0077***	0.0077***	0.0078***
	(0.0023)	(0.0023)	(0.0023)
Educated	-0.0045**	-0.0045**	-0.0045**
	(0.0018)	(0.0018)	(0.0018)
Rural	0.0047**	0.0046**	0.0047**
	(0.0023)	(0.0023)	(0.0023)
Poorest	(Base category)	(Base category)	(Base category)
Poor	-0.0017	-0.0016	-0.0016
	(0.0024)	(0.0024)	(0.0024)
Middle	-0.00021	-0.00019	-0.00016
	(0.0027)	(0.0026)	(0.0027)
Rich	0.0013	0.0013	0.0013
	(0.0030)	(0.0030)	(0.0030)
Richest	-0.0022	-0.0021	-0.0022
	(0.0036)	(0.0036)	(0.0036)
District, Month and Year Fixed Effects	Yes	Yes	Yes
Temperature & Rainfall controls	Yes	Yes	Yes
Observations	188190	188190	188190
Unique Clusters	6809	6809	6809

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Sample weights have been used in all regressions.

Bibliography

[1] Assessment of Air Pollution in Indian Cities (2017), Greenpeace Report. Technical report.

[2] Food and agriculture organization (FAO) (2017), Technical report.

[3] Health Effects Institute, State of Global Air (2018), Technical report.

[4] Vulnerability of Forest Fire (2017), Forest Survey of India. Technical report.

[5] Acharya, P., Sreekesh, S., and Kulshrestha, U. (2016), "Ghg and Aerosol Emission from Fire Pixel during Crop Residue Burning Under Rice and Wheat Cropping Systems in North-West India.", The International Archives of Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 41:753.

[6] Ahmad, J., Goldar, B. N., Misra, S., and Jakariya, M. (2003), "Willingness to pay for arsenic-free, safe drinking water in Bangladesh.", World Bank Water and Sanitation Program-South Asia.

[7] Ahmad, S. A., Sayed, M., Barua, S., Khan, M. H., Faruquee, M., Jalil, A., Hadi, S. A., and Talukder, H. K. (2001), "Arsenic in drinking water and pregnancy outcomes.", *Environmental health perspectives*, 109(6):629-631.

[8] Ahsan, H., Chen, Y., Parvez, F., Zablotska, L., Argos, M., Hussain, I., Momotaj, H., Levy, D., Cheng, Z., Slavkovich, V., et al. (2006), "Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the health effects of arsenic longitudinal study.", *American Journal of Epidemiology*, 163(12):1138-1148.

[9] Ahsan, H., Perrin, M., Rahman, A., Parvez, F., Stute, M., Zheng, Y., Milton, A. H., Brandt- Rauf, P., Van Geen, A., and Graziano, J. (2000), "Associations between drinking

water and urinary arsenic levels and skin lesions in bangladesh.", *Journal of occupational and environmental medicine*, 42(12):1195-1201.

[10] Alaii, J. A., Van Den Borne, H., Kachur, S. P., Mwenesi, H., Vulule, J. M., Hawley, W. A., Meltzer, M. I., Nahlen, B. L., and Phillips-Howard, P. A. (2003), "Perceptions of bed nets and malaria prevention before and after a randomized controlled trial of permethrintreated bed nets in western Kenya.", *The American journal of tropical medicine and hygiene*, 68(4):142-148.

[11] Alam, M., Allinson, G., Stagnitti, F., Tanaka, A., and Westbrooke, M. (2002), "Arsenic contamination in bangladesh groundwater: a major environmental and social disaster.", *International Journal of Environmental Health Research*, 12(3):235-253.

[12] Alexander, D. and Schwandt, H. (2019), "The Impact of Car Pollution on Infant and Child Health: Evidence from Emissions Cheating.", Chicago Working Paper No. WP-2019-4

[13] Almond, D. and Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. Journal of Economic Perspectives, 25(3):153-72.

[14] Altmann, S. and Traxler, C. (2014), "Nudges at the dentist.", *European Economic Review*, 72:19-38

[15] Argos, M., Kalra, T., Rathouz, P. J., Chen, Y., Pierce, B., Parvez, F., Islam, T., Ahmed, A., Rakibuz-Zaman, M., Hasan, R., et al. (2010), "Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in bangladesh (heals): a prospective cohort study.", *The Lancet*, 376(9737):252-258.

[16] Baggio, M. (2016), "Short-run Economic Benefits from Water Quality Improvements: an Application to the American Lobster (Homarus Americanus) Fishery of Long Island Sound.", *Estuaries and coasts*, 39(6), 1816-1826.

[17] Baggio, M., Chavas, JP, Falco, S., Hertig, A., Pomati, F. (2019), "The effect of anthropogenic and environmental factors in coupled humannatural systems: Evidence from Lake Zrich", *Natural Resource Modelling*.

[18] Balasubramanya, S., Pfaff, A., Bennear, L., Tarozzi, A., Ahmed, K. M., Schoenfeld, A., and van Geen, A. (2014), "Evolution of households' responses to the groundwater arsenic crisis in bangladesh: information on environmental health risks can have increasing behavioral impact over time.", *Environment and development economics*, 19(5):631-647.

[19] Barber, C. and Schweithelm, J. (2000), "Trial by Fire: Forest Fires and Forestry Policy in Indonesias Era of Crisis and Reform.", Technical report, World Resources Institute, Washington, DC (EUA). Forest Frontiers Initiative.

[20] Bennear, L., Tarozzi, A., Pfaff, A., Balasubramanya, S., Ahmed, K. M., and Van Geen, A. (2013), "Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in bangladesh.", *Journal of environmental economics and management*, 65(2):225-240.

[21] Bharadwaj, P., Gibson, M., Zivin, J. G., and Neilson, C. (2017), "Gray Matters: Fetal Pollution Exposure and Human Capital Formation.", *Journal of the Association of Environmental and Resource Economists*, 4(2):505-542.

[22] Binka, F. N., Kubaje, A., Adjuik, M., Williams, L. A., Lengeler, C., Maude, G., Armah, G., Kajihara, B., Adiamah, J., and Smith, P. G. (1996), "Impact of permethrin impregnated bednets on child mortality in kassena-nankana district, ghana: a randomized controlled trial.", *Tropical Medicine & International Health*, 1(2):147-154.

[23] Bjerk, D. (2009), "Beauty vs. earnings: Gender differences in earnings and priorities over spousal characteristics in a matching model.", *Journal of Economic Behavior & Organization*, 69(3):248-259.

[24] Brainerd, E. and Menon, N. (2014), "Seasonal E ects of Water Quality: The Hidden Costs of the Green Revolution to Infant and Child Health in India.", *Journal of Development Economics*, 107:49-64.

[25] Brinkel, J., Khan, M., and Kraemer, A. (2009), "A systematic review of arsenic exposure and its social and mental health effects with special reference to bangladesh.", *International journal of environmental research and public health*, 6(5):1609-1619.29

[26] Buchmann, N., Field, E. M., Glennerster, R., and Hussam, R. N. (2019), "Throwing the baby out with the drinking water: Unintended consequences of arsenic mitigation efforts in bangladesh.", National Bureau of Economic Research.

[27] Buss, D. M. (1989), "Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures.", *Behavioral and brain sciences*, 12(1):1-14.

[28] Cairncross, S., Shordt, K., Zacharia, S., and Govindan, B. K. (2005), "What causes sustainable changes in hygiene behaviour? a cross-sectional study from Kerala, India." *Social science & medicine*, 61(10):2212-2220.

[29] Caldwell, B. K., Caldwell, J. C., Mitra, S., and Smith, W. (2003), "Tubewells and arsenic in bangladesh: challenging a public health success story." *International Journal of Population Geography*, 9(1):23-38.

[30] Calzolari, G. and Nardotto, M. (2011), "Nudging with information: A randomized fi eld experiment." Available at SSRN 1924901.

[31] Carson, R. T., Koundouri, P., and Nauges, C. (2010), "Arsenic mitigation in Bangladesh: a household labor market approach.", *American Journal of Agricultural Economics*, 93(2):407-414.

[32] Chakrabarti, S., Khan, M. T., Kishore, A., Roy, D., and Scott, S. P. (2019), "Risk of Acute Respiratory Infection from Crop Burning in India: Estimating Disease Burden and Economic Welfare from Satellite and National Health Survey Data for 250 000 persons.", *International Journal of Epidemiology*

[33] Chakravorty, U., Pelli, M., and Risch, A. (2014), "Far Away from the Forest? Fuelwood Collection and Time Allocation in Rural India.", ISI Delhi Working Paper

[34] Chakravorty, U., Pelli, M., and Risch, A. (2015), "Who is Buying? Fuelwood Collection in Rural India.", Unpublished manuscript.

[35] Chay, K. Y. and Greenstone, M. (2003a), "Air Quality, Infant Mortality, and the Clean Air Act of 1970.", Technical report, National Bureau of Economic Research.

[36] Chay, K. Y. and Greenstone, M. (2003b), “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession.”, *The Quarterly Journal of Economics*, 118(3):1121-1167.

[37] Chen, Y., van Geen, A., Graziano, J. H., Pfaff, A., Madajewicz, M., Parvez, F., Hussain, A. I., Slavkovich, V., Islam, T., and Ahsan, H. (2007), “Reduction in urinary arsenic levels in response to arsenic mitigation efforts in Araihazar, Bangladesh.”, *Environmental Health Perspectives*, 115(6):917-923.

[38] Chiappori, P.-A., Oreffce, S., and Quintana-Domeque, C. (2012), “Fatter attraction: anthropometric and socioeconomic matching on the marriage market.”, *Journal of Political Economy*, 120(4):659-695.

[39] Chowdhury, S., Dey, S., Tripathi, S. N., Beig, G., Mishra, A. K., and Sharma, S. (2017), “Traffic Intervention Policy Fails to Mitigate Air Pollution in Megacity Delhi.”, *Environmental Science & Policy*, 74:8-13.

[40] Chowdhury, S., Krause, A., and Zimmermann, K. F. (2016), “Arsenic contamination of drinking water and mental health.”, DEF-Discussion Papers on Development Policy, (222).30

[41] Cook, C. J., Fletcher, J. M., and Forgues, A. (2018), “Multigenerational Effects of Early Life Health Shocks.”, Technical report, National Bureau of Economic Research.

[42] Council, E. S., Redon, J., UK, A. D., Narkiewicz, K., Nilsson, P. M., Burnier, M., Viigimaa, M., Ambrosioni, E., UK, M. C., Coca, A., et al. (2013), “2013 esh/esc guidelines for the management of arterial hypertension.”, *European Heart Journal*, 34:2159-2219.

[43] Cummins, J., Agarwal, N., Aiyar, A., Bhattacharjee, A., Gunadi, C., Singhania, D., Taylor, M., Wigton-Jones, E., et al. (2017), “Month of Birth and Child Height in 40 Countries.”, *Economic Letters*, 157, 10-13.

[44] Currie, J. and Neidell, M. (2005), “Air Pollution and Infant Health: What Can We Learn from Californias Recent Experience?”, *The Quarterly Journal of Economics*, 120(3):1003-1030.

[45] Currie, J. and Vogl, T. (2013), "Early-life Health and Adult Circumstance in Developing Countries.", *The Annual Review of Resource Economics*, 5(1):1-36.

[46] Currie, J. and Walker, R. (2011), "Traffic Congestion and Infant Health: Evidence from E-Z Pass.", *American Economic Journal: Applied Economics*, 3(1):65-90.

[47] Currie, J., Zivin, J. G., Mullins, J., and Neidell, M. (2014), "What Do We Know about Short-and Long-Term Effects of Early-Life Exposure to Pollution?", *The Annual Review of Resource Economics*, 6(1):217-247.

[48] Curtis, V., Kanki, B., Cousens, S., Diallo, I., Kpozehouen, A., Sangare, M., and Nikiema, M. (2001), "Evidence of behaviour change following a hygiene promotion programme in Burkina Faso.", *Bulletin of the World Health Organization*, 79:518-527.

[49] Dandona, L., Dandona, R., Kumar, G. A., Shukla, D., Paul, V. K., Balakrishnan, K., Prabhakaran, D., Tandon, N., Salvi, S., Dash, A., et al. (2017), "Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the global burden of disease study.", *The Lancet*, 390(10111):2437-2460.

[50] de Walque, D. (2007), "How does the impact of an HIV/AIDS information campaign vary with educational attainment? evidence from rural Uganda.", *Journal of Development Economics*, 2(84):686-714.

[51] Deaton, A. and Dreze, J. (2009), "Food and Nutrition in India: Facts and Interpretations.", *Economic and Political Weekly*, 42-65.

[52] Del no, R. J., Brummel, S., Wu, J., Stern, H., Ostro, B., Lipsett, M., Winer, A., Street, D. H., Zhang, L., Tjoa, T., et al. (2009), "The relationship of respiratory and cardiovascular hospital admissions to the southern california wildfires of 2003.", *Occupational and environmental medicine*, 66(3):189-197.

[53] Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., Reif, J. (2016), "The mortality and medical costs of air pollution: Evidence from changes in wind direction", (No. w22796). National Bureau of Economic Research.

[54] Dey, S. and Di Girolamo, L. (2010), "A Climatology of Aerosol Optical and Microphysical Properties over the Indian Subcontinent from 9 Years (2000-2008) of Multiangle Imaging Spectroradiometer (MISR) Data.", *Journal of Geophysical Research: Atmospheres*, 115(D15).

[55] Dey, S., Girolamo, L., Van Donkelaar, A., Tripathi, S., Gupta, T., and Mohan, M. (2012), "Decadal Exposure to Fine Particulate Matters (PM 2.5) in The Indian Subcontinent Using Remote Sensing Data.", *Remote Sens Environ*, 127:153-161.

[56] Do, Q.-T., Joshi, S., and Stolper, S. (2018), "Can Environmental Policy Reduce Infant Mortality? Evidence from the Ganga Pollution Cases.", *Journal of Development Economics*, 133:306-325.

[57] Douglass, R. L. (2008), "Quantification of the Health Impacts Associated with Fine Particulate Matter due to Wildfires.", PhD dissertation, Nicholas School of the Environment and Earth Sciences, Duke University.

[58] Duflo, E., Dupas, P., and Kremer, M. (2015), "Education, HIV, and early fertility: Experimental evidence from Kenya.", *American Economic Review*, 105(9):2757-97.

[59] Dupas, P. (2011a), "Do teenagers respond to hiv risk information? Evidence from a fi eld experiment in Kenya.", *American Economic Journal: Applied Economics*, 3(1):1-34.

[60] Dupas, P. (2011b), "Health behavior in developing countries.", *Annual Review of Economics*, 3(1):425-449.

[61] Elder Jr, G. H. (1969), "Appearance and education in marriage mobility.", *American Sociological Review*, 519-533.

[62] Field, E. and Ambrus, A. (2008), "Early marriage, age of menarche, and female schooling attainment in Bangladesh.", *Journal of political Economy*, 116(5):881-930.

[63] Finlay, S. E., Mo at, A., Gazzard, R., Baker, D., and Murray, V. (2012), "Health impacts of wildfires.", *PLoS currents*

[64] Flanagan, S. V., Johnston, R. B., and Zheng, Y. (2012), "Arsenic in tube well water in Bangladesh: health and economic impacts and implications for arsenic mitigation.", *Bulletin of the World Health Organization*, 90:839-846.

[65] Fletcher, J. M., Green, J. C., and Neidell, M. J. (2010), "Long Term Effects of Childhood Asthma on Adult Health.", *Journal of Health Economics*, 29(3):377-387.

[66] Foster, A., Gutierrez, E., and Kumar, N. (2009), "Voluntary compliance, Pollution Levels, and Infant Mortality in Mexico.", *American Economic Review*, 99(2):191-97.

[67] Foster, A. D. and Rosenzweig, M. R. (2003), "Economic Growth and the Rise of Forests.", *The Quarterly Journal of Economics*, 118(2):601-637.

[68] Frankenberg, E., McKee, D., and Thomas, D. (2005), "Health consequences of forest fires in Indonesia.", *Demography*, 42(1):109-129.

[69] Galasso, E., Wagsta , A., Naudeau, S., and Shekar, M. (2016), "The Economic Costs of Stunting and How to Reduce Them.", Policy Research Note World Bank, Washington, DC.

[70] Gale, D. and Shapley, L. S. (1962), "College admissions and the stability of marriage.", *The American Mathematical Monthly*, 69(1):9-15.

[71] Gallagher, K. M. and Updegraff, J. A. (2011), "Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review.", *Annals of behavioral medicine*, 43(1):101-116

[72] Goldstein, M., Zivin, J. G., Habyarimana, J., Pop-Eleches, C., and Thirumurthy, H. (2008), "Health worker absence, hiv testing and behavioral change: Evidence from western Kenya." Mimeo, Department of Economics, Columbia University.

[73] Goss, P. E., Strasser-Weippl, K., Lee-Bychkovsky, B. L., Fan, L., Li, J., Chavarri-Guerra, Y., Liedke, P. E., Pramesh, C., Badovinac-Crnjevic, T., Sheikine, Y., et al. (2014), "Challenges to effective cancer control in China, India, and Russia.", *The lancet oncology*, 15(5):489-538.

[74] Goyal, N. and Canning, D. (2017), "Exposure to Ambient Fine Particulate Air Pollution In-Utero as a Risk Factor for Child Stunting in Bangladesh.", *International Journal of Environmental Research and Public Health*, 15(1):22.

[75] Greenstone, M. and Hanna, R. (2014), "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India.", *American Economic Review*, 104(10):3038-72.

[76] Gupta, P. K., Sahai, S., Singh, N., Dixit, C., Singh, D., Sharma, C., Tiwari, M., Gupta, R. K., and Garg, S. (2004), "Residue Burning in Rice-Wheat Cropping System: Causes and Implications.", *Current Science*, 1713-1717.

[77] Gupta, R. (2014), "Low-Hanging Fruit In Black Carbon Mitigation: Crop Residue Burning In South Asia.", *Climate Change Economics*, 5(04):1450012.

[78] Gupta, R. et al. (2011), "Agro-environmental revolution in Punjab: Case of the happy seeder technology." Technical report.

[79] Gupta, S., Agarwal, R., and Mittal, S. K. (2016), "Respiratory Health Concerns in Children at Some Strategic Locations from High PM Levels during Crop Residue Burning Episodes.", *Atmospheric Environment*, 137:127-134.

[80] Habluetzel, A., Diallo, D., Esposito, F., Lamizana, L., Pagnonl, F., Lengeler, C., Traore, C., and Cousens, S. (1997), "Do insecticide-treated curtains reduce all-cause child mortality in burkina faso?", *Tropical medicine & international health*, 2(9):855-862.

[81] Haikerwal, A., Akram, M., Del Monaco, A., Smith, K., Sim, M. R., Meyer, M., Tonkin, A. M., Abramson, M. J., and Dennekamp, M. (2015), "Impact of fine particulate matter (pm 2.5) exposure during wildfires on cardiovascular health outcomes.", *Journal of the American Heart Association*, 4(7):e001653.

[82] Hassan, M. M., Atkins, P. J., and Dunn, C. E. (2005), "Social implications of arsenic poisoning in Bangladesh.", *Social Science & Medicine*, 61(10):2201-2211.

[83] He, G., Fan, M., and Zhou, M. (2016), "The Effect of Air Pollution on Mortality in China: Evidence from The 2008 Beijing Olympic Games.", *Journal of Environmental Economics and Management*, 79:18-39.

[84] Holstius, D. M., Reid, C. E., Jesdale, B. M., and Morello-Frosch, R. (2012), "Birth Weight following Pregnancy during The 2003 Southern California Wildfires.", *Environmental Health Perspectives*, 120(9):1340.

[85] Isen, A., Rossin-Slater, M., and Walker, W. R. (2017), "Every Breath you Take Every Dollar youll Make: The Long-Term Consequences of The Clean Air Act of 1970.", *Journal of Political Economy*, 125(3):848-902.

[86] Jain, N., Bhatia, A., and Pathak, H. (2014), *Emission of Air Pollutants from Crop Residue Burning in India.*, *Aerosol and Air Quality Research*, 14(1):422-430.

[87] Jakariya, M. (2007), "Arsenic in tubewell water of Bangladesh and approaches for sustainable mitigation." PhD thesis, KTH.

[88] Jalan, J., Somanathan, E., and Chaudhuri, S. (2009), "Awareness and the demand for environmental quality: survey evidence on drinking water in urban India." *Environment and Development Economics*, 14(6):665-692.

[89] Jayachandran, S. (2009), "Air Quality and Early-Life Mortality Evidence from Indonesias Wildfires.", *Journal of Human Resources*, 44(4):916-954.

[90] Jose, A. P., Awasthi, A., Kondal, D., Kapoor, M., Roy, A., and Prabhakaran, D. (2019), "Impact of repeated blood pressure measurement on blood pressure categorization in a population-based study from India.", *Journal of human hypertension*

[91] Kahn, R. A. and Gaitley, B. J. (2015), "An Aanalysis of Global Aerosol Type as Retrieved by MISR.", *Journal of Geophysical Research: Atmospheres*, 120(9):4248-4281.

[92] Kalmijn, M. (1998), "Intermarriage and homogamy: Causes, patterns, trends.", *Annual review of sociology*, 24(1):395-421.

[93] Kaskaoutis, D., Kumar, S., Sharma, D., Singh, R. P., Kharol, S., Sharma, M., Singh, A., Singh, S., Singh, A., and Singh, D. (2014), "Effects of crop residue burning on aerosol properties, plume characteristics, and long-range transport over northern India.", *Journal of Geophysical Research: Atmospheres*, 119(9):5424-5444.

[94] Keskin, P., Shastry, G. K., and Willis, H. (2017), "Water quality awareness and breastfeeding: evidence of health behavior change in Bangladesh.", *Review of Economics and Statistics*, 99(2):265-280.

[95] Kinniburgh, D. and Smedley, P. (2001), "Arsenic contamination of groundwater in Bangladesh." British Geological Survey and Department of Public Health Engineering (BGS and DPHE)

[96] Kirby, D. B., Baumler, E., Coyle, K. K., Basen-Engquist, K., Parcel, G. S., Harrist, R., and Banspach, S. W. (2004), "The safer choices intervention: its impact on the sexual behaviors of different subgroups of high school students.", *Journal of Adolescent Health*, 35(6):442-452

[97] Kochi, I., Donovan, G. H., Champ, P. A., and Loomis, J. B. (2010), "The economic cost of adverse health effects from wild re-smoke exposure: a review.", *International Journal of Wildland Fire*, 19(7):803-817.

[98] Kunii, O., Kanagawa, S., Yajima, I., Hisamatsu, Y., Yamamura, S., Amagai, T., and Ismail, I. T. S. (2002), "The 1997 Haze Disaster in Indonesia: Its Air Quality and Health Effects.", *Archives of Environmental Health: An International Journal*, 57(1):16-22.

[99] Lai, W., Li, Y., Tian, X., and Li, S. (2017), "Air Pollution and Cognitive Function: Evidence from Crop Production Cycles.", Available at SSRN 3039935.

[100] Luby, S. P., Kadir, M. A., Yushuf Sharker, M., Yeasmin, F., Unicomb, L., and Sirajul Islam, M. (2010), "A community-randomised controlled trial promoting waterless hand sanitizer and handwashing with soap, Dhaka, Bangladesh.", *Tropical medicine & international health*, 15(12):1508-1516.

[101] Madajewicz, M., Pfaff, A., Van Geen, A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R., and Ahsan, H. (2007), "Can information alone change behavior? response to arsenic contamination of groundwater in Bangladesh.", *Journal of development Economics*, 84(2):731-754.

[102] Maitra, P., Menon, N., and Tran, C. (2018), "The Winters Tale: Season of Birth Impacts on Children in China.", Monash University, Discussion Paper 09-18.

[103] Marini, M. M. (1978), "The transition to adulthood: Sex differences in educational attainment and age at marriage.", *American sociological review*, 483-507.

[104] Mendez, M. A. and Adair, L. S. (1999), "Severity and Timing of Stunting in the First Two Years of Life Affect Performance on Cognitive Tests in Late Childhood.", *The Journal of Nutrition*, 129(8):1555-1562.

[105] Milton, A. H., Smith, W., Rahman, B., Hasan, Z., Kulsum, U., Dear, K., Rakibuddin, M., and Ali, A. (2005), "Chronic arsenic exposure and adverse pregnancy outcomes in Bangladesh.", *Epidemiology*, 82-86.

[106] Mukherjee, A., Sengupta, M. K., Hossain, M. A., Ahamed, S., Das, B., Nayak, B., Lodh, D., Rahman, M. M., and Chakraborti, D. (2006), "Arsenic contamination in groundwater: a global perspective with emphasis on the asian scenario.", *Journal of Health, Population and Nutrition*, 142-163.

[107] Neidell, M. J. (2004), "Air Pollution, Health, and Socio-Economic Status: The Effect of Outdoor Air Quality on Childhood Asthma.", *Journal of Health Economics*, 23(6):1209-1236.

[108] ODonnell, M. and Behie, A. (2015), "Effects of Wildfire Disaster Exposure on Male Birth Weight in an Australian Population.", *Evolution, Medicine, and Public Health*, 2015(1):344- 354.

[109] Opar, A., Pfaff, A., Seddique, A., Ahmed, K., Graziano, J., and van Geen, A. (2007), "Responses of 6500 households to arsenic mitigation in Araihazar, Bangladesh.", *Health & place*, 13(1):164-172.

[110] Oster, E., Shoulson, I., and Dorsey, E. (2013), "Optimal expectations and limited medical testing: Evidence from Huntington disease.", *American Economic Review*, 103(2):804-30.

[111] Prabhakaran, D., Jeemon, P., Sharma, M., Roth, G. A., Johnson, C., Harikrishnan, S., Gupta, R., Pandian, J. D., Naik, N., Roy, A., et al. (2018), "The changing patterns of cardiovascular diseases and their risk factors in the states of India: the global burden of disease study 1990-2016.", *The Lancet Global Health*, 6(12):e1339-e1351.

[112] Prinz, D., Chernew, M., Cutler, D., and Frakt, A. (2018), "Health and Economic Activity Over the Lifecycle: Literature Review.", Technical report, National Bureau of Economic Research.

[113] Pullabhotla, H. (2018a), "Fires, wind, and smoke: Air pollution and infant mortality.", Job Market Paper.

[114] Rahman, A., Vahter, M., Ekstrom, E.C., Rahman, M., Golam Mustafa, A. H. M., Wahed, M. A., Yunus, M., and Persson, L.A. (2007), "Association of arsenic exposure during pregnancy with fetal loss and infant death: a cohort study in Bangladesh.", *American journal of epidemiology*, 165(12):1389-1396.

[115] Rangel, M. A. and Vogl, T. S. (2018), "Agricultural fires and health at birth.", *Review of Economics and Statistics*

[116] Ranta, E., Lindstrm, K., & Salojrvi, K. (1992), "Water quality, fishing effort and fish yield in lakes.", *Fisheries research*, 15(1-2), 105-119.

[117] Ravenscroft, P., Brammer, H., and Richards, K. (2009), "Arsenic pollution: a global synthesis", volume 28. John Wiley & Sons.

[118] Rukumnuaykit, P. (2003), "Crises and Child Health Outcomes: The Impacts of Economic and Drought/Smoke Crises on Infant Mortality and Birth Weight in Indonesia", Mimeo, Michigan State.

[119] Sahu, S. K., Zhang, H., Guo, H., Hu, J., Ying, Q., and Kota, S. H. (2019), "Health Risk Associated with Potential Source Regions of PM 2.5 in Indian Cities.", *Air Quality, Atmosphere & Health*, 1-14.

[120] Sanders, N. J. (2012), "What Doesnt Kill You Makes You Weaker - Prenatal Pollution Exposure and Educational Outcomes.", *Journal of Human Resources*, 47(3):826-850.

[121] Sharma, A. R., Kharol, S. K., Badarinath, K., and Singh, D. (2010), “Impact of Agriculture Crop Residue Burning on Atmospheric Aerosol Loading - A Study over Punjab State, India.”, *Annales Geophysicae*, 28(2).

[122] Shyamsundar, P., Springer, N., Tallis, H., Polasky, S., Jat, M., Sidhu, H., Krishnapriya, P., Skiba, N., Ginn, W., Ahuja, V., et al. (2019), “Fields on fire: Alternatives to crop residue burning in India.”, *Science*, 365(6453):536-538.

[123] Simeonova, E., Currie, J., Nilsson, P., and Walker, R. (2018), “Congestion Pricing, Air Pollution and Childrens Health.”, Technical report, National Bureau of Economic Research.

[124] Singh, C. P. and Panigrahy, S. (2011), “Characterisation of Residue Burning from Agricultural System in India Using Space based Observations.”, *Journal of the Indian Society of Remote Sensing*, 39(3):423.

[125] Singh, R. P. and Kaskaoutis, D. G. (2014), “Crop residue burning: a threat to south asian air quality.”, *Eos, Transactions American Geophysical Union*, 95(37):333-334.

[126] Smith, A. H., Lingas, E. O., and Rahman, M. (2000), “Contamination of drinking-water by arsenic in Bangladesh: a public health emergency.”, *Bulletin of the World Health Organization*, 78:1093-1103.

[127] Sohel, N., Persson, L. A., Rahman, M., Streat eld, P. K., Yunus, M., Ekstrom, E.-C., and Vahter, M. (2009), “Arsenic in drinking water and adult mortality: a population-based cohort study in rural Bangladesh.”, *Epidemiology*, 824-830.

[128] Soo, J.-S. T. and Pattanayak, S. K. (2019), “Seeking Natural Capital Projects: Forest Fires, Haze and Early-Life Exposure in Indonesia.”, *Proceedings of the National Academy of Sciences*

[129] Spears, D., Dey, S., Chowdhury, S., Scovronick, N., Vyas, S., and Apte, J. (2019), “The Association of Early-life Exposure to Ambient PM2.5 and Later-childhood Height-for-age in India: An Observational Study.”, *Environmental Health*, 18(1):62.

[130] Spivey, C. (2010), “Desperation or desire? the role of risk aversion in marriage”, *Economic Inquiry*, 48(2):499-516.

[131]

[132] Srivastava, P. and Garg, A. (2013), “Forest Fires in India: Regional and Temporal Analyses.”,*Journal of Tropical Forest Science*, 228-239.

[133] Tan-Soo, J.-S. and Pattanayak, S. K. (2019), “Seeking natural capital projects: Forest fires, haze, and early-life exposure in Indonesia.”, *Proceedings of the National Academy of Sciences*, 116(12):5239-5245.

[134] UNICEF (2008), “Arsenic mitigation in Bangladesh.”

[135] UNICEF et al. (2007), “Predicting the global extent of arsenic pollution of groundwater and its potential impact on human health.”, Report Prepared by Peter Ravenscroft, Consultant Hydrogeologist, Cambridge, UK.

[136] Vadrevu, K. P., Lasko, K., Giglio, L., and Justice, C. (2015), “Vegetation Fires, Absorbing Aerosols and Smoke Plume Characteristics in Diverse Biomass Burning Regions of Asia.”, *Environmental Research Letters*, 10(10):105003.

[137] van Donkelaar, A., Martin, R., Verduzco, C., Brauer, M., Kahn, R., Levy, R., and Villeneuve, P. (2010), “A Hybrid Approach for Predicting PM 2.5 Exposure: van Donkelaar et al. respond.”, *Environmental Health Perspectives*, 118(10):A426-A426.

[138] Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., and Villeneuve, P. J. (2010), “Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application.”, *Environmental Health Perspectives*, 118(6):847.

[139] Van Geen, A., Ahsan, H., Horneman, A. H., Dhar, R. K., Zheng, Y., Hussain, I., Ahmed, K. M., Gelman, A., Stute, M., Simpson, H. J., et al. (2002), “Promotion of well-switching to mitigate the current arsenic crisis in Bangladesh.”, *Bulletin of the World Health Organization*, 80:732-737.

[140] Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., Sachdev, H. S., Maternal, Group, C. U. S., et al. (2008), “Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital.”, *The Lancet*, 371(9609):340-357.

[141] Vijayakumar, K., Safai, P., Devara, P., Rao, S. V. B., and Jayasankar, C. (2016), “Effects of Agriculture Crop Residue Burning on Aerosol Properties and Long-Range Transport over Northern India: A Study Using Satellite Data and Model Simulations.”, *Atmospheric Research*, 178:155-163.

[142] World Bank (2007), “Implementation completion and results report (ida-31240 swtz-21082) on a credit in the amount of sdr 24.2 million (USD 44.4 million equivalent) to Bangladesh for arsenic mitigation water supply.”

[143] Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., Clement, D. L., Coca, A., De Simone, G., Dominiczak, A., et al. (2018), “2018 esc/esh guidelines for the management of arterial hypertension.”, *European Heart Journal*, 39(33):3021-3104.

[144] Zheng, S., Wang, J., Sun, C., Zhang, X., and Kahn, M. E. (2019), “Air Pollution Lowers Chinese Urbanites Expressed Happiness on Social Media.”, *Nature Human Behaviour*