2021

Health, Environment, And Inequality In India

Aashish Gupta
University of Pennsylvania

Health, Environment, And Inequality In India

Abstract

This dissertation makes three scientific contributions to understand the ongoing epidemiological transition in India. The first chapter documents local externalities of solid fuel use for adult lung function. The use of solid fuels for cooking and heating is rooted in poverty and gender inequality within households. However, harms from solid fuel use are more widespread. In neighborhoods with high solid fuel use, the lungs of those who do not use solid fuels can be as obstructed as the lungs of those who use solid fuels. Because it contributes to both infectious disease among children and chronic diseases among adults, the use of solid fuels complicates the epidemiological transition in India. The second chapter observes that Indian infants face higher mortality risks in the summer, monsoon, and winter months compared to the spring months. Using birth history data, the chapter develops an innovative demographic approach which estimates and adjusts infant mortality by calendar month. The chapter highlights that Indian infants face multiple environmental threats that are less salient for a limited period within a year. It finds that although seasonal variation has declined, it remains a concern in rural areas and among more disadvantaged households. The last chapter provides the first estimates of life expectancy by social class in India for the period 1990-2016. It develops methods to directly estimate life tables from survey data. The chapter documents persistent and stark mortality disparities in a period of robust economic growth and changing disease profiles. It finds progress in reducing levels and differentials in child mortality. However, patterns in the working ages are concerning, with slower progress and little reduction in inequality. The three chapters make both substantive and methodological contributions to the study of health and mortality in low- and middle-income countries. They show that addressing social inequalities and environmental risks are essential for population health improvements in India.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Demography

First Advisor
Irma T. Elo

Keywords
Demography, Environment, Epidemiological transition, Health Disparities, India, Mortality

Subject Categories
Demography, Population, and Ecology | Public Policy | Sociology

HEALTH, ENVIRONMENT, AND INEQUALITY IN INDIA

Aashish Gupta

A DISSERTATION

in

Demography and Sociology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

_____

Irma T. Elo, Professor of Sociology

Graduate Group Chairperson, Demography

_____

Irma T. Elo, Professor of Sociology

Graduate Group Chairperson, Sociology

_____

Jason Schnittker, Professor of Sociology

Dissertation Committee
Irma T. Elo, Professor of Sociology
Jere R. Behrman, William R. Kenan, Jr. Professor of Economics, Professor of Sociology
Michel Guillot, Professor of Sociology
Atheendar S. Venkataramanni, Assistant Professor of Medical Ethics and Health Policy

HEALTH, ENVIRONMENT, AND INEQUALITY IN INDIA

ACKNOWLEDGMENTS

ABSTRACT

HEALTH, ENVIRONMENT, AND INEQUALITY IN INDIA

Aashish Gupta

Irma T. Elo

This dissertation makes three scientific contributions to understand the ongoing epidemiological transition in India. The first chapter documents local externalities of solid fuel use for adult lung function. The use of solid fuels for cooking and heating is rooted in poverty and gender inequality within households. However, harms from solid fuel use are more widespread. In neighborhoods with high solid fuel use, the lungs of those who do not use solid fuels can be as obstructed as the lungs of those who use solid fuels. Because it contributes to both infectious disease among children and chronic diseases among adults, the use of solid fuels complicates the epidemiological transition in India. The second chapter observes that Indian infants face higher mortality risks in the summer, monsoon, and winter months compared to the spring months. Using birth history data, the chapter develops an innovative demographic approach which estimates and adjusts infant mortality by calendar month. The chapter highlights that Indian infants face multiple environmental threats that are less salient for a limited period within a year. It finds that although seasonal variation has declined, it remains a concern in rural areas and among more disadvantaged households. The last chapter provides the first estimates of life expectancy by social class in India for the period 1990-2016. It develops methods to directly estimate life tables from survey data. The chapter documents persistent and stark mortality disparities in a period of robust economic growth and changing disease profiles. It finds progress in reducing levels and differentials in child mortality. However, patterns in the working ages are concerning, with slower progress and little reduction in inequality. The three chapters make both substantive and methodological contributions to the study of health and mortality in low- and middle-income countries. They show that addressing social inequalities and environmental risks are essential for population health improvements in India.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

In 1947, in his first speech to an independent India, Jawaharlal Nehru listed the several tasks before the nation. A prominent part of the speech were endeavors "to fight and end poverty and ignorance and disease" (Nehru, 1983). Mindful of the enormity of the task, Nehru reiterated that fulfilling these endeavors would require "incessant striving". It would also require monitoring progress. So in 1953-54, for its seventh survey round, India's National Sample Survey fielded a module on mortality (Government of India, 1962). As the survey report admitted, there was likely "considerable under-reporting" of deaths in the survey (pg. 5). The survey estimated an infant mortality rate of 151 infant deaths per 1,000 births. In contrast, indirect approaches used by the United Nations inter-agency group estimated an infant mortality rate of 180 deaths per 1,000 live births for 1953-54 (UN-IGME, 2020).

People themselves did not need surveys to know that mortality was high, or that fighting diseases is easier said than done. Veteran *Dalit* activist Babytai Kamble, who was born before independence, discussed the scale of mortality matter-of-factly in her autobiography, *The Prisons We Broke*:

> "Many young girls on the threshold of life succumbed to death. One in every
> ten lost their lives during childbirth. Infants died as well."

> – Kamble (1990)

To better understand how the country was doing in its fight against disease, India instituted the Sample Registration System (SRS) in the late 1960s. In thousands of randomly selected villages and urban blocks, the SRS monitored all births and deaths. Although by no means eliminating them, the SRS did help reduce uncertainty in mortality estimates (Bhat et al., 1984). It also noted some progress. Infant mortality was estimated to be 140 deaths per 1,000 births in 1975 by the SRS (ORGI, 2016). The UN-IGME estimate for this year is

close: 131 infant deaths per 1,000 births. These rates are lower than the infant mortality rates estimated for the 1950s. But they were still too high. The SRS also helped generate the first direct and reliable estimates of mortality across the life course. If a hypothetical cohort experienced the age-specific mortality rates that the SRS observed in 1970-75, they would have lived, on average, to the age of 49.7 years. About 50 years.

As the infant mortality rate indicates, many would have died earlier. Sugna, for instance, who died in infancy, in Gorakh Pandey's 1975 poem *Mehnat Ke Barahmasa* - "Twelve months of hard work". Her mother reminds her father:

<div align="center">

हमरी सुगना को ले गइल बुखार सजना,
नाहीं दवा-दारू नाहीं उपचार सजना,
तोर मेहनत-मजूरी सब बेकार सजना,
अस जिनगी जियल धिरकार सजना।

</div>

<div align="center">

Our Sugna was taken by the fever, O beloved,

Neither medicine nor healthcare, O beloved,

Your hard work and labour is pointless, O beloved,

Living such a life is worthless, O beloved.

</div>

<div align="right">

– Pandey (2014)

</div>

Mortality has reduced further since then. A large number of factors, among them people's own struggles for a better life, economic growth, government initiatives, social movements, and scientific progress have contributed to this decline (Caldwell et al., 1988). In 2018, the most recent years for which estimates are available, the infant mortality rate was 32 deaths per 1,000 births (Drèze et al., 2021). Estimated period life expectancy at birth was 70.7 years for women and 68.2 years for men in 2014-2018 (Government of India, 2020).

Within the broader demographic literature, these changes have been organized within ideas of the health transition (Caldwell, 1993), or the epidemiological transition (Horiuchi,

1999). Deaton (2013) has called these reductions in mortality "The Great Escape". Babytai Kamble also recognized the escape from premature mortality and suffering. But she also recognized the role of "incessant struggles" that Nehru invoked. Particularly the struggles waged by those like her who were at the margins of society – "The Prisons *We* Broke."

However, this progress in mortality reduction may still be too slow. According to Drèze and Sen (2013), survival rates in India "are far from flattering", especially when compared to peer countries. As they note, India's GDP per capita is thrice as high as Nepal, and twice as high as Bangladesh, yet child survival and life expectancy are higher in these countries. Why might this be the case? According to Drèze and Sen (2013):

> "It is not surprising that class, caste, and gender inequalities have played an important role in preventing India from becoming a social development leader in South Asia."

Similar concerns were echoed by Dr. B.R. Ambedkar around the time of the first National Sample Survey on mortality, when he noted that "the health of the untouchable is the care of nobody" (Ambedkar, 1989), and before that, that "caste has killed public spirit." (Ambedkar, 1936). Coffey et al. (2017) have noted that India's caste system may prevent the use of simple toilets that prevent diseases, which explains why Indian children are shorter than children in Sub-Saharan Africa (Spears, 2018). Coffey (2015b) finds that Indian women finish pregnancy at weights lower than the weights at which women in Sub-Saharan Africa start pregnancy. Consequently, neonatal mortality in India is higher than what GDP per-capita would predict (Coffey, 2015a).

In other words, these authors are pointing out that social inequalities may slow the epidemiological transition in India. The more the number of prisons, the harder the escape.

Gender inequalities and environment in the epidemiological transition

Gender inequality in household wok is particularly severe in India. Men in Indian households do little cooking, for instance (Hirway and Jose, 2011; Sharma, 2018). Women also have the responsibility to prepare the solid fuels used for cooking. The use of these solid fuels is reinforced by gender inequality within the household (Vyas et al., 2020). The first essay of this dissertation, *"Solid fuel externalities, gender, and adult respiratory health"* shows that solid fuel use affects not just women – its harms are much more widespread.

The chapter is concerned with lung obstruction and Chronic Obstructive Lung Disease (COPD). COPD is a leading cause of death in India (Jha et al., 2005). In 2007-08, the WHO SAGE conducted spirometry, a test to measure lung function and obstruction in six large Indian states. The WHO-SAGE also measured behaviors such as smoking and the type of fuel used in kitchens by households. As expected, the WHO-SAGE found that those who smoked had higher lung obstruction.

As did those who used solid fuels, the chapter shows. In addition, the chapter documents local negative externalities of solid fuel use. This idea of local air pollution contributions of solid fuel use and subsequent harms to lung obstruction had been theorized, for instance, by Chafe et al. (2014). The chapter provides an empirical test of this hypothesis. A household and its members may not use solid fuels. But if they were around other households that used solid fuels, their lungs can be equally obstructed. This is because even in Indian villages, settlements can be dense, and household use of solid fuels can contribute to local air pollution. Similarly, men, even if they did not smoke, had their lungs obstructed if they were members of households that used solid fuels, or residents of neighborhoods that used solid fuels.

These harms for adult respiratory health are in addition to already documented effects of solid fuel use on child health and mortality (Venkataramani and Fried, 2011). The chapter

also shows that once household and local solid fuel use was accounted for, there was no relationship between household income and lung obstruction. Given these negative externalities of solid fuel use, what can public policy do? Negative externalities are a classic case for public intervention. In 2016, the Indian government initiated the *Ujjwala* program. It encouraged access to subsidized LPG cylinders. Poverty is certainly a factor preventing solid fuel adoption, and the *Ujjwala* program reduced the use of solid fuels (Gupta et al., 2020). However, instead of tackling gender inequality and encouraging both men and women to cook with LPG, the *Ujjwala* program reinforced the role of women as cooks in the household in its messaging. Despite the program, survey evidence from rural India shows that solid fuel use remains common, even if families own LPG stoves and cylinders (Gupta et al., 2020; Vyas et al., 2020).

This folly, of not targeting root causes that lie in social inequality, was recognized by the chief architect of India's constitution, Dr. B.R. Ambedkar. He argued that "to leave inequality between class and class, between sex and sex" and "to go on passing legislation relating to economic problems is to make a farce of our Constitution and to build a palace on a dung heap" (Ambedkar, 1990). Dung heaps that are recycled to fuel *chulhas* remain common in rural India, and Dr. Ambedkar's warning remains relevant.

Multiple environmental threats and infant mortality

Gender inequality contributes to solid fuel use, and the use of solid fuels contributes to air pollution. But other things contribute to air pollution too (Spears, 2019a). In the winter months, people often make fires and burn stuff to get some warmth and respite from the cold. Before the lighting of fire crackers in the winter festival of Diwali, farmers in north India burn crop stubble to prepare the fields for the next crop. Air pollution itself gets trapped near the ground in the winter, a meteorological phenomenon called inversion. Do children have worse health and higher mortality in these months?

But the winter is harsh not just because of the pollution. As Munni and Halku knew well

in Premchand's "A January Night", the cold itself can be brutal (Orsini et al., 2004). And beyond the winter, the summer brings heat. Soon after that, the monsoon brings humidity. Both can be fatal (Geruso and Spears, 2018a). The monsoon may also spread infectious diseases (Pascual et al., 2002). Add floods to this, then droughts, and then the vagaries of agriculture, and there may hardly be a time when child mortality risks are not high. Consequently, as Behrman and Deolalikar (1989) document, "important seasonal variations in nutritional and health status" are "persistent themes in the development literature."

Does infant mortality vary by season too? To answer this, the second essay in the dissertation, *"Seasonal variation in infant mortality in India"*, examines mortality by calendar month. It uses data from the National Family Health Surveys, which visited all Indian states in 1992-93, 1998-99, 2005-06, and 2015-16. These surveys interview women in the reproductive ages 15-49 years. They ask them if they have ever had children, the month and year of the birth of their child if they had children, and the age at death of their child, if their child died. Using this information, the chapter estimates the number of infants alive and infant deaths for each calendar month up to four years before the survey.

Although women interviewed by the NFHS list all the births they have had, this study considers birth in the last five years. This is to reduce recall errors. Another concern is seasonality in births. If some calendar months have more births, then the average infant in them would be younger than in other months. Among children, mortality is highest close to birth, and the risk of death declines as children age (Guillot et al., 2012). To adjust for variation in ages of children within a month, the chapter uses demographic standardization techniques. It adjusts the extent to which each child 'weighs' in the data, such that the distribution of children's age in each calendar-month is the same as distribution of ages as the overall survey round.

The chapter finds higher mortality in the summer, monsoon, and winter months when compared to the spring months. At an annualized rate, infant mortality in the spring months was lower by as many as 11 deaths per 1,000 infants per year in the early 1990s.

In the latest available round of the NFHS, for the period 2012-2015, mortality in the spring months was lower by 4 deaths per 1,000 infants per year. These mortality differences between spring and other seasons are substantial – higher than the overall mortality rates in some high-income countries, for instance.

What might we do to reduce seasonal variation, and bring mortality in other months to the level of spring? The chapter does not measure the changes in the environmental exposures over the seasons itself. The prior population science literature has already documented the effect of these exposures. This literature has emphasized the population health importance of reducing pollution, simple technologies such as kangaroo mother care, improving indoor temperature control in rural houses, and preventing exposure to heat and humidity (Spears, 2019a). The chapter shows that the overall contribution of these approaches can lead to important reductions in child mortality. It emphasizes priorities for healthcare settings too. Hospitals and health systems can plan for seasonal fluctuations in health and in births. These may reduce health system burdens in particular months. The chapter also documents that seasonal variation in mortality is higher among poorer households, and in rural areas.

These results will not be surprising to Gorakh Pandey. This was the theme of his poem *Mehnat ke Barahmasa* – "Twelve months of Hard Work" (Pandey, 2014) – that there is almost no month in the year which brings any respite. Except perhaps a brief spring, when the weather was less harsh and hunger was less common.

Mortality disparities across the life course

Can inequalities between "class and class" that Dr. Ambedkar invoked in his resignation speech, be observed in overall mortality as well? In an essay in the *Annual Review of Sociology*, Irma Elo noted that "among the many advantages afforded to the well-educated, affluent, and well-positioned is a longer and healthier life" (Elo, 2009). As the previous two chapters hint at, the poor are more likely to be exposed to environmental harms, from

pollution to heat. These higher exposures, and the lesser ability of those who are poor to fight the diseases these exposures cause would predict higher mortality among the poor. A large demographic literature, using the birth-history data used in chapter 2, has found higher mortality among poorer children in India and elsewhere (Chao et al., 2018; Chalasani and Rutstein, 2014). A separate literature has found that insults in the childhood ages have consequences in the adult ages (Barker, 1995). If poorer children are more likely to be exposed to environmental harms, and poor children grow up to be poorer adults, then poorer adults may also have higher mortality.

Because the link between class and mortality has been so persistent across time and space, Link and Phelan (1995) have called social conditions a "fundamental cause" of disease. However, in contrast to high-income countries, and in contrast to child mortality, stark gradients in adult mortality have not been documented in low- and middle-income countries (Sudharsanan, 2019; Beltrán-Sánchez et al., 2020; Luo and Xie, 2014; Nikoi and Odimegwu, 2013; Sudharsanan et al., 2020). The fact that adult mortality risk factors, such as obesity and cardiovascular disease, are not more common among poorer adults helps explain these modest-to-no gradients in low- and middle-income countries.

To what extent does socioeconomic status pattern mortality in India? And how does this relationship vary by age? Chapter 3, *"Household wealth and life expectancy in India: 1990-2016"* investigates this. Information on mortality beyond the childhood ages is sparse in most low- and middle-income countries. However, in addition to asking about child mortality, the NFHS also implemented a simple question about recent deaths in the household in its first, second, and fourth rounds. The NFHS asked the hundreds of thousands of households that it visited if there were any deaths of usual household members in the last couple of years. If a death had occurred, the NFHS interviewers also asked about the sex, age, and the month and year of death of the person who died. Demographers have worried about the accuracy of this question (Timæus, 1991; Hill, 1991) but the chapter shows that all-India age-specific mortality rates constructed from this approach correspond closely

with the official Indian life tables as estimated by the Sample Registration System.

Surveyors also asked the household about assets they owned - from cycles to cars, from clocks to washing machines. The chapter combines information on these assets to create a household wealth index. The NFHS are multi-stage sample surveys. First, the NFHS organizers randomly selected urban blocks and villages. Then, NFHS interviewers randomly selected households within these villages and urban blocks. The chapter develops an approach to cluster-bootstrap standard errors. This measures uncertainty in the estimates of life expectancy or mortality probability by wealth.

The chapter finds disparities in period life expectancy across ages, regions, and periods in India. In 1990-99, the wealthiest ten per cent of Indian women and men had life expectancy at birth of 71.8 years and 70.1 years, respectively. This is higher than the life expectancy estimates for India more than two decades later in 2014-2018. In other words, this level of overall life expectancy has not yet been achieved for India as a whole. The poorest decile of Indians, meanwhile, had a life expectancy at birth of 57.0 years among women and 54.0 years among men in 1990-1999. Average life expectancy at such levels were seen in India in the 1980s, about fifteen years b efore. Life expectancy gaps between the wealthiest and the poorest Indians in 1990-1999 were thus larger than the life expectancy gains in India over three decades.

The period between 1990-99 and 2012-16 in India saw declines in poverty, but also increases in economic inequality (Himanshu, 2019). For mortality, however, the chapter observes declines in both overall mortality as well as disparities in mortality. Disparities declined in both absolute and relative terms. These declines reflect the extent to which India's poorest were able to break "prisons". They were also aided, to some extent, by improvements in living standards in this period, and state initiatives. As Drèze and Khera (2017) wrote, "There has been a major expansion of social security programs in India during the last 15 years or so, along with wider recognition of economic and social rights."

However, differences are still quite large. For life expectancy at birth, differences between the top and bottom decile were 8.3 years among women and 13.5 years among men. These disparities were not just contributed by differences in child mortality between the wealthy and the poor. At age 15, the richest women had 3 to 7 more years in life expectancy than the poorest women, and the richest men had 6 to 9 more years than the poorest men.

Ironically, the decline in life expectancy disparities is less of a puzzle than their presence in the first place. Why is India not like other low and middle-income countries, where mortality disparities in the adult ages are modest or non-existent? Countries such as China and South Africa, without strong gradients in adult mortality (Sudharsanan et al., 2020; Nikoi and Odimegwu, 2013; Luo and Xie, 2014), often have more economic inequality than India (Ravallion, 2008). It is not the case that cardio-vascular disease is more common among the poor in India, either (Arokiasamy et al., 2021). So economic inequality or cardio-vascular disease are unlikely to explain adult mortality disparities in India.

Environmental exposures may provide an explanation – pollution from solid fuel use harms both children and adults, and could explain adult mortality disparities to some extent. Chronic Obstructive Pulmonary Disease and Tuberculosis are still leading causes of death in India, and are known to affect the poor more. Ultimately, social inequalities, such as those of gender and caste may be more severe in India, even more than other unequal contexts, and may be the ultimate cause of class-based adult mortality disparities. Marginalized social groups in India, such as *Dalits* and *Adivasis* also have lower life expectancy than higher-caste Hindus (Gupta and Sudharsanan, 2020; Vyas et al., 2021).

These inequalities may constrain overall population health as well. In Uttar Pradesh, for instance, the chapter finds that life expectancy at birth in the richest quartile of women was lower than the life expectancy of women in the poorest quartile in South India. Dreze and Gazdar (1997) guessed that this would happen. Comparing Uttar Pradesh, India's largest state, with South India, they pointed out:

"Underlying the specific failures relating to public services, elementary education, and women's agency, is a deeper failure to achieve the kind of basic social change that facilitates progress in these fields. One aspect of the 'inertia' that accounts for slow social progress in Uttar Pradesh is the apathy of the state, but an equally important factor is the failure of civil society to challenge oppressive patterns of caste, class, and gender relations. The society of Uttar Pradesh remains steeped in traditional inequalities, which makes it that much harder to achieve widespread literacy, to run efficient public services, or to promote the agency of women in social and political matters."

The epidemiological transition in India

Arundhati Roy once wrote that "India lives in several centuries at the same time" (Roy, 2002). Indeed, mortality risks as high as the ones faced by India's marginalized populations were observed in high-income countries more than a century ago. In a similar vein, Drèze and Sen (2013) described India as a "sea of Sub-Saharan Africa with islands of California." A substantial section of the Indian population – the bottom decile, which is a population that is larger than that of most countries – has life expectancy equivalent to that of Sub-Saharan Africa. But there isn't an AIDS epidemic in India. Instead, it is social inequality which is rampant.

This dissertation in Demography and Sociology suggests that these inequalities and environmental exposures are key constraints on population health improvements in India. However, these constraints are not unbreakable. As Babytai Kamble's experience shows, it is possible to escape from prisons, confront social inequality, improve population health, and reduce health disparities. Although it is a much smaller concern than these goals above, this dissertation also suggests that it is possible to measure population health better.

CHAPTER 1:

Solid fuel externalities, gender, and adult respiratory health in India

Abstract

Chronic respiratory conditions are a leading cause of death. Using data on lung obstruction from the WHO Survey of Global AGEing and Adult Health (WHO-SAGE 2007-08), this paper studies the determinants of respiratory health in India, home to a third of all deaths from Chronic Obstructive Pulmonary Disease. First, we find that smokers and members of households that use solid fuels (wood, biomass, coal or dung) for cooking have higher lung obstruction. Second, even if a respondent's household uses clean fuels, their lung obstruction is higher if their neighbors use solid fuels. In neighborhoods with high solid fuel use, the lungs of members of households that use clean fuels can be as obstructed as lungs of members of households that use solid fuels. These negative externalities of solid fuel use are robust to additional controls for neighborhood socioeconomic status, falsification tests, tests with placebo measures, and tests using alternative measures of respiratory health as outcomes. Third, the influence of the determinants is patterned by gender. Smoking tobacco is an important influence on lung obstruction among men. Confirming non-linear dose-response relationships, we find that women from households that use solid fuels are the only group not further harmed by neighborhood solid fuel smoke, possibly because of high exposure to pollutants while cooking. The study improves our understanding of behavioral, social, and environmental determinants of respiratory health in India. Importantly, it makes a case for greater public investments to promote the adoption and use of cleaner fuels.

1.1 Introduction and background

Chronic Obstructive Pulmonary Disease (COPD), "a lung ailment that is characterized by a persistent blockage of airflow from the lungs" (World Health Organisation, 2015) is among the leading causes of death in India (Gomes et al., 2017). Although less developed countries have a higher burden of disease attributable to chronic respiratory diseases and adult respiratory health is becoming increasingly important for overall population health, empirical research on the determinants of respiratory health at the population level in developing countries is limited (Mannino and Buist, 2007). In high-income countries, smoking tobacco is the principle risk factor for COPD (Currie, 2010; Pauwels and Rabe, 2004). Emerging evidence suggests that smoke from the use of solid fuels (wood, animal dung, coal, charcoal, and crop residues), which are used for cooking and heating, may be an equally important determinant in less developed countries (Salvi and Barnes, 2009, 2010).

Solid fuel smoke can harm the respiratory health of women (Po et al., 2011; Ezzati and Kammen, 2002), as they are the household members primarily responsible for cooking (Hirway and Jose, 2011). Other household members can be harmed because of exposure to indoor air pollution (Venkataramani and Fried, 2011; Hu et al., 2010; Mishra et al., 2005). Researchers have hypothesized that because solid fuel smoke contributes to local air pollution, neighbors can be harmed as well, even if they themselves use clean fuels (Smith, 2002; Torres-Duque et al., 2008; Balakrishnan et al., 2014; Chafe et al., 2014). This points to important negative externalities, or negative effects of solid fuel use on third parties (Pigou, 1924; Fairbrother, 2016). However, existing research has not yet tested or quantified the association between neighborhood externalities of solid fuel use and lung obstruction empirically. Although these negative externalities can extend beyond the neighborhood as well, we investigate local externalities as they are likely to be stronger. Finally, the patterning of the exposures by key social determinants such as gender has not been studied at the population level for developing countries. This paper intends to fill these research gaps, for the important case of India.

Both smoking and cooking are patterned by gender in India, a large section of the Indian population is exposed to solid fuel smoke, and respiratory health is poor. Few women in India smoke (Jha et al., 2002; Mishra et al., 2016), while men do little household work or cooking (Hirway and Jose, 2011; Sharma, 2018). Using data from the only representative survey that measured time use in India, Hirway and Jose (2011) found that while men spent less than an hour per week in India on cooking and related household activities, women spent an average of 25 hours per week. Thus, women who live in households that use solid fuels have high exposure to direct smoke while cooking (Parikh et al., 1999). More than 67.2% of Indian households used solid fuels for cooking (Government of India, 2011a; Gupta et al., 2019), and about one-third of the world's population which relies on solid fuels was estimated to live in India (Bonjour et al., 2013). Finally, in 2016, India was estimated to be home to one-third of all deaths from chronic respiratory conditions (GBD Causes of Death Collaborators, 2017), and almost half of Indian adults had COPD (World Health Organisation, 2013) by the GOLD (2016) criteria. COPD was also among the top three causes of death (Krishnan et al., 2011; Patel et al., 2011). These factors make India an important and urgent context to examine lung obstruction determinants.

The paper informs this literature by considering, for the first time, neighborhood externalities of solid fuel use in an empirical framework, by using population representative survey data on lung obstruction, by examining the gendered patterning of the exposures, and by investigating both dichotomous and continuous measures of lung obstruction as outcomes. We use population level survey data on lung obstruction collected by the World Health Organization's Survey of Global AGEing and adult health (WHO-SAGE) 2007-08 (Kowal et al., 2012).

From a social science perspective, this study contributes evidence from a developing country context to the literature on environment and population health, as well as the literature on how social and behavioral determinants mediate this relationship. From a health policy perspective, the study presents evidence of negative externalities, which are

a classic argument for public intervention.

We discuss policy and research implications of this study in section 3.5. The next subsection provides background information on lung obstruction. Section 1.2 discusses the measures and the empirical strategy, and section 3.3 presents results.

Causes, prevalence, and diagnosis of lung obstruction

The lung is the body's organ of gas exchange. The trachea or the windpipe branches into the bronchi inside the lungs, which further divide into smaller branches, called bronchioles. The bronchioles end in microscopic air sacs called alveoli. Long-term exposure to noxious particles and gases produces inflammation in the lungs (Currie, 2010). This inflammation results in the narrowing of the bronchioles, and is accompanied by lung tissue destruction (emphysema), disruption of repair and defense mechanism in the lungs (bronchiolitis), and secretion of mucous and increased cough (chronic bronchitis) (Calverley and Walker, 2003). These changes increase resistance to airflow in the lungs. Persistent respiratory symptoms and lung obstruction characterize Chronic Obstructive Pulmonary Disease (GOLD, 2016). Lung obstruction can result in respiratory or heart failure (Holguin et al., 2005), increases the risk of lung infections and pneumonia (Prescott et al., 1995) while exacerbating the risk of death from co-morbidities such as heart disease and lung cancer (Sin et al., 2006).

Smoking tobacco is a well-established cause of lung obstruction. In a systematic review, Forey et al. (2011) pool odds-ratios from 133 studies and find higher odds of COPD in ever smokers, current smokers, and ex-smokers when compared to non-smokers. While smoking cessation cannot reverse damage, a systematic review found that compared to smoking, quitting was associated with slower increases in lung obstruction (Godtfredsen et al., 2008). Inhaling smoke from solid fuels is the other big risk factor for lung obstruction. Torres-Duque et al. (2008) and Kurmi et al. (2010) review evidence on the impact of solid fuel use on respiratory diseases, including COPD. Kurmi et al. (2010) find higher odds of having COPD among households that use solid fuels. In India, McKay et al. (2012);

4

Moschovis et al. (2015); Dave et al. (2017) have found sex, smoking status, and indoor air pollution to be significant predictors of lung obstruction. However, some studies (Amaral et al., 2017; Sana et al., 2018) have failed to find significantly higher odds of COPD among those who used solid fuels. These studies have not considered solid fuel externalities. The comparison groups in these studies, those not using solid fuels, may also have high lung obstruction because of exposure to neighborhood air pollution.

COPD is diagnosed using a non-invasive physiological test, spirometry, which measures lung obstruction. A spirometer measures the proportion of the volume of air (liters) one can exhale in 1 second to the total volume of air one exhales, after inhaling as deeply as possible (Cotes et al., 2009). This ratio is called Forced Expiratory Volume in 1 second percent (FEV1%). Higher values of FEV1% signify better respiratory health, and imply lesser obstruction of airflow. Existing research on solid fuel use and respiratory health relies on dichotomized COPD as the outcome, based on the Global initiative for Obstructive Lung Disease (GOLD) criteria of diagnosing COPD if FEV1% is less than 70% (GOLD, 2016). The GOLD cut-off over-diagnoses older patients, and under-diagnose younger patients (Pellegrino et al., 2005). While categorization has clinical relevance, such dichotomization does not capture the severity of the disease and reduces statistical power and precision (Weinberg, 1995; Royston et al., 2006). Consequently, we use continuous FEV1% as the outcome, but confirm that the results are robust to using dichotomized COPD.

1.2 Materials and methods

Data and measures

The WHO-SAGE (2007-08) is representative of the adult population aged above 18 years in six states of India: Rajasthan, Uttar Pradesh, Assam, Karnataka, Maharashtra and West Bengal. These states had a combined population of 570 million in 2011, representing 47% of India's total population (Government of India, 2011b). The WHO-SAGE is unique among

surveys carried out in India in that it conducted spirometry to measure lung obstruction for its 9,551 respondents (World Health Organisation, 2013). The WHO-SAGE also collected information on other individual and household characteristics. Table 1–1 provides summary statistics for the main measures used in this analysis of WHO-SAGE data.

Outcome

The primary outcome is Forced Expiratory Volume in 1 second percent (FEV1%), the proportion of total air inhaled that a person can exhale in one second. FEV1% is a direct measurement of lung obstruction. The outcome is also called FEV1/FVC (ratio of Forced Expiratory Volume in 1 second to Forced Vital Capacity). We use the continuous measure to preserve the full range of variation in lung function and improve the precision of our estimates. Our results were qualitatively unchanged when using a dichotomous outcome. Average FEV1% was 70.8% in the WHO-SAGE sample. Mean levels of FEV1% as low as these, very close to the GOLD criteria of FEV1% of 70% to clinically diagnose someone as having COPD, reinforce the impaired state of respiratory health in India. Men and women had similar mean values of lung obstruction (Table 1–1).

Exposures

The main independent variables that measure exposure to risk factors for lung obstruction are smoking; household solid fuel use, and neighborhood solid fuel use.

Smoking: Smoking status was self-reported. Smoking is coded as a dummy variable (a very small fraction of the sample consisted of former smokers, and second-hand smoke was not observed). About 17% of the SAGE respondents smoked (Table 1–1). While 31% of the men smoked, only 2% of the women smoked.

Household use of solid fuels: Respondents were asked the fuel their household uses for cooking. Coal/charcoal, wood, agricultural residue/shrubs and animal dung were classified as solid fuels. Kerosene, gas, and electricity were classified as clean fuels. In robustness tests, we examine if the result change by considering only gas and electricity as clean

6

fuels (Table 1A–3). Household use of solid fuels is coded as a dummy variable. About 76% of all households used solid fuels for cooking.

Neighborhood solid fuel use: Neighborhood solid fuel use is measured as the proportion of households, excluding the respondent's household, in a Primary Sampling Unit (PSU) that use solid fuels. A PSU is an urban block or a rural village. Neighborhood solid fuel use is coded as a continuous variable between 0 and 1. The SAGE survey interviewed respondents in 375 different clusters or PSUs, selecting on average 26 households per cluster. Nine out of ten clusters had more than 19 respondents (a histogram of the cluster size is shown in Figure 1A–1). Solid fuel use in the neighborhood is hypothesized to capture the deterioration in air quality in the neighborhood due to solid fuel use. On average, 76% of a respondent's neighbors used solid fuels. Accordingly, the respondent would be assigned a value of 0.76 as a measure of neighborhood solid fuel use.

Although a larger sample from each PSU would be desirable, such specifications have been used productively in research on neighborhood disease exposure, such as sanitation (Geruso and Spears, 2018b) and vaccination (McGovern and Canning, 2015). The neighborhood solid fuel use measure may introduce measurement error due to extreme values (everyone using solid fuels or nobody using solid fuels). This measurement error biases our estimates on neighborhood solid fuel use towards zero, thus making these estimates conservative. An additional concern is the correlation between neighborhood and household solid fuel use. This estimated correlation between these variables was .6969, and the Variation Inflation Factor based on the r-squared in a regression was 1.955. We find similar results when we use district level solid fuel use estimated from the 2011 Indian census instead of neighborhood solid fuel use estimated from the WHO-SAGE 2007-08 (Table 1A–3).

Other explanatory variables

In all regressions, we adjust for age dummies (in years), sex of the respondent, state fixed effects, and a dummy for living in a rural area. This specification of age has the advantage of accounting for non-linear relationships between age and lung obstruction. In the SAGE sample, average age was 40.4 years. Men were slightly older than women, and 47% of the sample was female. Almost 70% of the respondents lived in rural areas (Table 1–1). Apart from demographic and place controls, we also control for two measures of socioeconomic status, namely, years of education and number of household assets owned. We use dummies for years of education and count of assets, a stronger control than linear controls for wealth and education. Years of education ranged from 0 to 18. Mean years of education were 5.9 years, 7.4 years for men and 4.2 years for women, reflecting an aspect of gender inequality well-known for India (Drèze and Sen, 2013). The WHO SAGE asked respondents the number of chairs, tables, and cars their household owned, whether the household had a cycle, clock, bucket, cot, fridge, fixed line phone, mobile phone, television, computer, radio, livestock, sewing machine, a motorized two-wheeler, bullock cart, and non-homestead agricultural land. Based on these indicators, we created a variable for the count of assets a respondent's household owned. The mean number of assets was 13.2. Using district identifiers in the WHO-SAGE, we matched the respondents with the proportion of households who were using solid fuels in their district in the 2011 census (Government of India, 2011a). Districts in India are a third level of administrative unit, below the federal and the state level, and above the level of rural villages or urban municipalities. India has 727 districts, and the WHO SAGE was conducted in 139 of them. Mean district level solid fuel use was 74%. We also used estimates of particulate matter pollution (PM 2.5) derived by Van Donkelaar et al. (2016). Since geographic coordinates of SAGE respondents or their clusters were not available, we estimated mean district level exposure in ArcGIS. Mean PM 2.5 exposure was 52.0 μg / m^3, which is relatively very high.

Empirical strategy

The empirical strategy consists of several steps, including descriptive graphs of cumulative density function of lung obstruction by exposure type, Ordinary Least Squares (OLS) regressions, and multiple robustness checks. The OLS linear regression equation is specified as:

$$respiratory\ health_{ij} = \beta_1 smoke\ tobacco_{ij} + \beta_2 household\ solid\ fuel\ use_{ij} +$$

$$\beta_3 neighborhood\ solid\ fuel\ use_{ij} \tag{1.1}$$

$$+ \alpha place_{ij} + \theta demographic_{ij} + \lambda ses_{ij} + \varepsilon_{ij},$$

where respiratory health is a measure of obstruction of airflow of respondent i living in primary sampling unit j, measured as FEV1%. We cluster standard errors at the level of the primary sampling unit, to account for the WHO SAGE's multi-stage sample design, and use survey weights throughout. The coefficients of interest are β_1, of whether the respondent smokes, β_2, of whether the respondent's household uses solid fuels, and β_3, of the proportion of other households in a respondent's primary sampling unit who use solid fuels. β_3 is a continuous measure between 0 and 1. Since regression coefficients show the effect of a one-unit change, β_3 indicates the difference between respondents who live in neighborhoods where no other households use solid fuels and respondents who live in neighborhoods where all other households use solid fuels.

$place_{ij}$ controls for state fixed effects and a dummy for rural residence; $demographic_{ij}$ adjusts for sex of the respondent and age dummies, in years; and ses_{ij} controls include dummies for both years of education and count of household assets.

In addition to these controls for SES variables, we conduct further robustness checks, adjusting for additional household and neighborhood SES variables, such as log household expenditure and average neighborhood assets. In placebo tests, we test whether

measures similar to the exposures, such as chewing tobacco, drinking alcohol, household ownership of toilets, household electrification, neighborhood open defecation or neighborhood electrification are associated with the outcome. In falsification tests, we test whether other health outcomes, such as Forced Vital Capacity of lungs or adult heights are associated with the exposures. Finally, we test if the exposures are associated with alternative measures of respiratory health, such as self-reported lung disease and Forced Expiratory Volume in 1 second. In placebo tests, it is expected that there will be no association between the placebo variables and the outcome. In falsification tests, it is expected that there will be no association between the falsification outcomes and the exposures. We expect to find an association between the alternative measures of respiratory health and the exposures. These results are available in the supplementary appendix.

1.3 Results

The results section is organized as follows. We begin by descriptive graphs in section 1.3, first examining the cumulative density function of FEV1% by the three exposures. We then see the extent to which the differences in the exposures can be explained by socioeconomic status. Thereafter, we examine whether household and neighborhood solid fuel use jointly influence lung obstruction. Section 1.3 presents regression results, documenting the association between the outcome and the exposures while adjusting for other explanatory variables, and examining the implications of the gendered patterning of the exposures. Finally, section 3.4 presents the results from robustness checks, including the placebo and falsification tests, and tests with alternative measures of respiratory health as outcomes.

Descriptive graphs

Figure 1–1 presents cumulative density function of FEV1% for those exposed to the three risk-factors of lung obstruction (smoking, household solid fuel use, neighborhood solid fuel use) and those not exposed. The cumulative distributions show differences in lung obstruction by exposure at all levels of the distribution of FEV1%. p-values and D-values

(absolute maximum distance between the two distributions) from Kolmogorov-Smirnov tests, and p-values and z-statistics from Wilcoxon rank-sum test are also shown to compare the distributions in each of the sub-figures. The p values, which are all less than .001, indicate that the cumulative density functions of those exposed to the risk factors and those not exposed are significantly different. Smokers, members of households that use solid fuels, and residents of neighborhoods in which other members use solid fuels have higher lung obstruction across the distribution, than non-smokers, members of households that use clean fuels and residents of neighborhoods that have lower levels of solid fuel use, respectively. The figure reveals that differences by household and neighborhood solid fuel use exposure are larger than or similar to the differences by smoking.

Figure 1–2 provides graphical evidence that these differences cannot be explained by differences in socioeconomic status. In fact, measures of socioeconomic status, such as years of education and household assets themselves have a weak relationship with lung obstruction. In Figure 1–2 the left panel shows differences in years of education, and the right panel shows differences by count of household assets using non-parametric local polynomial regressions with 95% confidence intervals shaded blue. At all levels of education and household assets, respondents whose households or neighbors use solid fuels have higher lung obstruction and lower values of FEV1%.

Figure 1–3 shows differences in lung obstruction between members of households that use solid fuels and members of households that use clean fuels by neighborhood solid fuel use. The x-axis shows the proportion of households in a primary sampling unit using solid fuels, and the y-axis shows FEV1%. Among those respondents whose households use clean fuels, there is a much steeper decline in FEV1% as neighborhood solid fuel use rises. Among those respondents whose households use solid fuels, as neighborhood solid fuel increases, a modest decline in FEV1% can be seen. That respondents who use cleaner fuels show a steeper decline in FEV1% as neighborhood solid fuel use increases reinforces existing evidence of a downward sloping concave concentration response function. This im-

11

plies diminishing marginal impacts of increases in pollution (Burnett et al., 2014; Pope III et al., 2015; Gupta and Spears, 2017). The pattern of non-linear decline among respondents whose households use clean fuels also supports these results. Respondents already exposed to direct or indoor smoke from solid fuel use are not much harmed by neighborhood solid fuels.

Regression results

Table 1–2 provides results from OLS regressions. The results indicate that average differences by the exposures are statistically significant, in magnitude similar to the ones in the descriptive graphs, and robust to demographic, place, and SES controls. The first 3 models present results for the overall sample, models 4 and 5 restrict the sample to women, and models 6 and 7 consider only men. All regressions adjust for age dummies, sex, state fixed effects and rural residence. All models except model 1 also adjust for socioeconomic status, by using dummies for years of education and number of household assets owned. Models with linear controls for age and measures of socioeconomic status are available in appendix table 1A–1. Appendix table 1A–1 shows that assets and education are not associated with lung obstruction, and that dummies for age and socioeconomic status improve model fit.

Model 1 of table 1–2 shows that those who smoke, whose household uses solid fuels, and whose neighbors use solid fuels have lower FEV1% and higher lung obstruction. Model 2 adjusts for socioeconomic status but the point estimates or the significance of the coefficients do not change substantially. Compared to a respondent who does not smoke, whose household does not use solid fuels, and whose neighbors do not use solid fuels, a respondent who smokes, or a respondent whose household uses solid fuels, is, on average, able to breathe out 3.7 percentage points less oxygen in the first second. Holding smoking and household solid fuel use constant, a respondent living in a neighborhood in which all her/his neighbors use solid fuels is able to breathe out 6 percentage points less FEV1%. Models 1 and 2 reinforce the results from the descriptive graphs, that household solid fuel

use and smoking tobacco are equivalent risks, and neighborhood solid fuel use is an even greater risk for lung obstruction. In other words, a person who is a member of a household that does not use solid fuels, does not smoke, but lives in a neighborhood in which half the households use solid fuels is likely to have the same level of lung obstruction as a person who is a tobacco smoker.

Model 3 shows the interaction between household and neighborhood solid fuel use. Predicted FEV1% for household and neighborhood solid fuel use are available in figure 1–4. The interaction results and predicted FEV1% closely replicate the results from figure 1–3: as neighborhood solid fuel rises, FEV1% declines more steeply among members of households that use clean fuels.

Gender, exposures, and lung obstruction

Women and men differ both in the extent of their exposure to solid fuel smoke and smoking tobacco (Jha et al., 2002; Hirway and Jose, 2011), as well as their physiological responses to the exposures (Aryal et al., 2013), making a case for within-sex comparisons. The next set of models in table 1–2 present stratified regressions for women (models 4 and 5) and men (models 6 and 7).

Model 4 of table 1–2 shows that while smoking has a negative association with FEV1% among women, the effect size is much smaller than that of men (model 7), and not significant. This is likely due to selection, lesser intensity of smoking among women, a very small number of women for whom smoking is observed, and possible reporting bias because women may be less willing to say that they smoke. Household solid fuel smoke is significantly and negatively associated with FEV1% among women. While neighborhood solid fuel use has a negative association with FEV1%, this effect isn't statistically significant (p-value 0.128). However, model 5, which interacts household and neighborhood solid fuel use for women, shows that neighborhood solid fuel use has a significantly negative association for women whose households don't use solid fuels. On the other hand, women who are members of households that use solid fuels are not further harmed by neighborhood

solid fuel use. This result can be more clearly seen through predicted margins, reported in figure 1–4.

Among men, smoking is associated with higher lung obstruction and lower FEV1%. The association of household use of solid fuels has a similar magnitude as the association of smoking, while the neighborhood solid fuel use coefficient is even higher than that of smoking (model 6). Model 7 reports coefficients from an interaction of household solid fuel use and neighborhood solid fuel use. For men, the interaction is not significant. These patterns, consistent with Burnett et al. (2014) and Pope III et al. (2015), underscore the high levels of exposure to smoke among women exposed to solid fuel smoke from direct cooking.

Robustness tests

As described previously, we next examine the robustness of the results. The regression tables are shown in the supplementary appendix.

Table 1A–2 adds controls for log monthly expenditure, log monthly expenditure squared, average neighborhood assets, and average neighborhood assets sequentially to the main specification from Table 1–2 (model 2), repeated as Model 1 in appendix table 1A–2 for comparison. The log of monthly household expenditure and the square of the log expenditure are not significantly associated with lung obstruction (model 2). Nor does the inclusion of these controls change the magnitude of the associations for the three exposures. Models 3 and 4 add controls for average neighborhood assets (model 3) and average neighborhood education (model 4). As with neighborhood solid fuel use, these variables are constructed as average of other households' average assets, or other respondents' average education. They are also not associated with FEV1%, and do not substantively change the effect sizes of the exposures.

Table 1A–3 tests for robustness in the exposures of household and neighborhood solid fuel use. In Model 1, we test if the main result is valid in the sub-sample of the

SAGE in primary sampling units with more than 19 households. These results are similar to the main result (Table 1–2, Model 1), confirming that the associations are not driven by the small proportion of primary sampling units with fewer respondents. Model 2 uses an instrumental variable approach. We randomly split the respondents in each primary sampling unit into two units. We then instrument the mean neighborhood solid fuel use in the first sub-sample of the primary sampling unit with the mean solid fuel use in the second sub-sample of the primary sampling unit. In this instrumental variable approach, the co-efficients for both household and neighborhood solid fuel use are higher than those in the main specification, suggesting that the estimates in the main model are conservative. Models 3 and 4 test for the association between district level solid fuel use and lung obstruction. We estimate district level solid fuel use from the 2011 census, which asked all households the fuel they used for cooking. Model 3 shows that the association of neighborhood and district level solid fuel use with lung obstruction is of similar magnitude. Model 4, which has all the exposures, at the household, neighborhood, and district level, shows that it is the household and neighborhood level exposures that matter. Once household and neighborhood level exposures are accounted for, district level exposures are not significant. This makes sense, because districts are much larger units. Finally, Model 5 separates clean fuels into kerosene and other clean fuels (gas and electricity). The model shows that Kerosene does not appear to be significantly worse than other cleaner fuels.

Table 1A–4 reports placebo tests for smoking tobacco, household solid fuel use, and neighborhood solid fuel use. The associations of chewing tobacco and drinking alcohol with FEV1% are compared to that of smoking tobacco. Owning a toilet and household electrification, measures of household SES, are used as placebos for solid fuel use. Measures for neighborhood electrification and neighborhood open defecation are used as placebos for neighborhood solid fuel use. As before, all regressions control for demographic, place, and other socioeconomic status variables. Panel A shows that there is no association between FEV1% and chewing tobacco, or between FEV1% and drinking alcohol. Chewing tobacco and drinking alcohol cause other health problems, but are not

15

linked to respiratory health, either theoretically or empirically in the WHO-SAGE. Regressions in Panel B find no association of FEV1% with household ownership of toilets and household electrification, and regressions in Panel C find no association of lung obstruction with neighborhood electrification or neighborhood open defecation. The effect sizes for the placebo variables are also close to zero.

In Table 1A–5, we conduct falsification tests with health measures we do not expect to be correlated with solid fuel use. It is well known that exposures such as smoking or solid fuel use harm airflow obstruction but not total lung capacity, which we confirm in model 1. The outcome in model 2 is height of the respondents. While height is influenced by early childhood disease environment, adult exposures to solid fuel use should not influence height, which is confirmed in model 2. We show that word recall, a measure of memory, and walk time, a measure of physical strength, are also not associated with the exposures.

In Table 1A–6, we confirm that the mechanism through which lung obstruction is associated with neighborhood solid fuel use is through ambient air pollution (models 1-3). We also confirm that alternative measures of respiratory health, such as dichotomized COPD and self-reported lung disease are predicted by the exposures. Model 1 examines the relationship between the outcome and the exposures, without controls for state of the respondent and rural residence, along with satellite PM 2.5 exposures. It finds that both neighborhood solid fuel use and satellite PM 2.5 are associated with lung obstruction. Model 2 introduces controls for state dummies and rural residence, as in our main specification. In this model, while neighborhood solid fuel use is significant, the association of satellite PM 2.5 is no longer significant. This could partly because of the spatial correlation of satellite measures of air pollution, which vary only over large geographic areas, such as states (Van Donkelaar et al., 2016). Still, Model 3, which has PM 2.5 as an outcome, confirms that local solid fuel use is predictive of satellite air pollution, establishing a link between neighborhood solid fuel use, pollution, and health.

Model 4 in Table 1A–6 shows that smoking tobacco, household use of solid fuels, and neighborhood solid fuel use are negatively associated with Forced Expiratory Volume in 1 second, measured in liters. Because the amount of air blown out is also associated with stature, model 4 adjusts for height in meters. Model 5 and 6 use the logistic model with binary outcomes of having COPD (following GOLD 2016 guidelines) and self-reported symptoms of lung disease, following the WHO SAGE guidelines (Arokiasamy et al., 2015). Both models show higher odds of having COPD and self-reported lung diseases among smokers, members of households who use solid fuels, and residents of neighborhoods with high solid fuels use, reinforcing the main finding.

1.4 Discussion

This paper expands evidence on the determinants of lung obstruction for the important case of India, where one-third of all deaths due to chronic respiratory conditions take place, in three key ways. First, to our knowledge, this is the first study that documents an empirical association between local solid fuel use by other households in one's neighborhood and obstruction in one's own lungs. Multiple robustness checks suggest a causal interpretation of this association. Second, this study improves our understanding of gender as a determinant of respiratory health in India. We find that women who live in households that use solid fuels are not further harmed by neighborhood solid fuel use. They are the only group not further harmed by neighborhood solid fuel use, which points to the already high exposure to indoor smoke they must have while cooking. Gender also influences smoking behavior. Finally, the study establishes that large sections of the Indian population - those who smoke, those who use solid fuels, and those who live near others who use solid fuels - have higher lung obstruction than those who aren't exposed to these determinants of lung obstruction. Higher socioeconomic status does not mitigate the influence of these exposures.

From the point of view of estimating differences in lung obstruction due to the

exposures over a person's life course, the study offers conservative estimates. Some of the respondents who were not exposed at the time of the survey would have been exposed to the exposures earlier in their lives. However, it has some limitations. The lungs of women who cook with solid fuels are more likely to be obstructed (Johnson et al., 2011), but cooking behavior is not observed in the WHO-SAGE. Future health surveys should identify household members responsible for cooking, and if possible, ask about duration of cooking, so that dose-response relationships can be understood. We are able to include only district-level satellite measures of air pollution in our analysis, as ground-level pollution monitoring in India is in a nascent stage (Guttikunda, 2017).

The study is also limited because it uses cross-sectional data. However, from a policy-makers' perspective, this study elevates the importance of promoting clean fuels to improve adult health in India. Even though currently exposed populations may not see immediate or significant improvements in FEV1% (Chapman et al., 2005; Smith-Sivertsen et al., 2009; Kurmi et al., 2012), further obstruction can be prevented. Encouraging the adoption of cleaner fuels remains a substantial challenge (Mobarak et al., 2012; Peel et al., 2015; Jeuland et al., 2015). But the only viable public health strategy for reducing lung obstruction is prevention through the adoption of cleaner fuels. Existing government programs in India to promote cleaner fuels are means-tested, provide modest capital subsidies, involve significant private investments, do not have components for encouraging behavior change, and focus on private harms to women from cooking with solid fuels (Sethi and Deep, 2018; Gupta et al., 2019). There is a case for greater investments in efforts to communicate the private and neighborhood harms of solid fuel use and encouraging households to switch to cleaner fuels.

From an environmental health perspective, there is an urgent need for research that integrates social determinants, environmental exposures, measures of pollution, and respiratory health outcomes. Existing evidence points that women's lower status impedes the adoption of cleaner fuels (Kishore and Spears, 2014; Austin and Mejia, 2017; Gupta

et al., 2019; Vyas et al., 2020). Researching the role of gender in the persistence of solid fuel use in India, and investigating approaches that address this social determinant of adoption of cleaner fuels are an important research concern as well.

Table 1–1: Descriptive statistics for the sample

	Total		Women		Men	
	mean	s.e.	mean	s.e.	mean	s.e.
Outcome						
Forced Expiratory Volume 1s %	70.8	0.5	71.2	0.7	70.5	0.6
Exposures						
Respondent smokes	0.17	0.01	0.02	0.00	0.31	0.02
Household uses solid fuels	0.76	0.02	0.77	0.02	0.76	0.02
Neighbourhood solid fuel use	0.76	0.02	0.77	0.02	0.76	0.02
Explanatory variables						
Demographic						
Age	40.4	0.3	39.4	0.3	41.3	0.4
Female	0.47					
Socio-economic Status						
Years of education	5.9	0.2	4.2	0.2	7.4	0.2
# of assets owned	13.2	0.2	13.2	0.2	13.3	0.2
Contextual						
Rural	0.70	0.02	0.70	0.03	0.70	0.02
District level solid fuel use	0.74	0.00	0.74	0.00	0.73	0.00
District level PM 2.5 (ug /m3)	52.0	0.3	51.6	0.3	52.5	0.4
N	9551		5777		3774	

Source: WHO SAGE 2007-08

Note: All means are weighted using national individual weights, and standard errors are clustered at the level of the primary sampling unit. FEV1% = (FEV1s * 100) / FVC.

Table 1–2: OLS regression coefficients of the association of smoking, household solid fuel use, and neighborhood solid fuel use with lung obstruction

model:	overall sample			women		men	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
smoke tobacco	-3.403**	-3.764***	-3.812***	-0.307	-0.506	-3.510**	-3.529**
	(1.04)	(1.04)	(1.04)	(3.183)	(3.185)	(1.130)	(1.131)
household uses solid fuels	-3.226*	-3.762***	-7.407*	-3.484*	-10.510*	-4.166*	-5.460
	(1.25)	(1.17)	(3.01)	(1.460)	(4.171)	(1.725)	(3.999)
neighbourhood solid fuel use	-6.082*	-6.088**	-7.299***	-3.399	-8.831***	-5.259*	-6.153+
	(2.50)	(2.45)	(2.06)	(2.231)	(2.480)	(2.457)	(3.453)
hh X neighbourhood solid fuel use			5.79		11.091*		2.011
			(3.71)		(5.149)		(5.191)
female	-0.396	-0.859	-0.849				
	(0.84)	(0.91)	(0.91)				
demographic controls		X	X	X	X	X	X
place controls	X	X	X	X	X	X	X
ses controls		X	X	X	X	X	X
constant	72.829***	74.768***	75.497***	71.536***	73.507***	76.119***	76.403***
	(1.96)	(2.62)	(2.71)	(3.284)	(3.490)	(3.151)	(3.355)
observations	9551	9551	9551	5777	5777	3774	3774

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + p<.1, * p<.05, ** p<.01, *** p<.001. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (litres) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). Demographic controls: age dummies and sex (coefficients for female shown). Place controls: rural residence and state fixed effects. SES controls: dummies for years of education and count of assets. Appendix Table 1 compares models with linear controls for age and ses measures with fixed effects for age and ses measures. Appendix Figure 1 shows the predicted margins of the interaction of household and neighbourhood solid fuel use, from models 3, 5, and 7. hh: household.

Figure 1–1: Cumulative Density Function of lung obstruction by exposure type
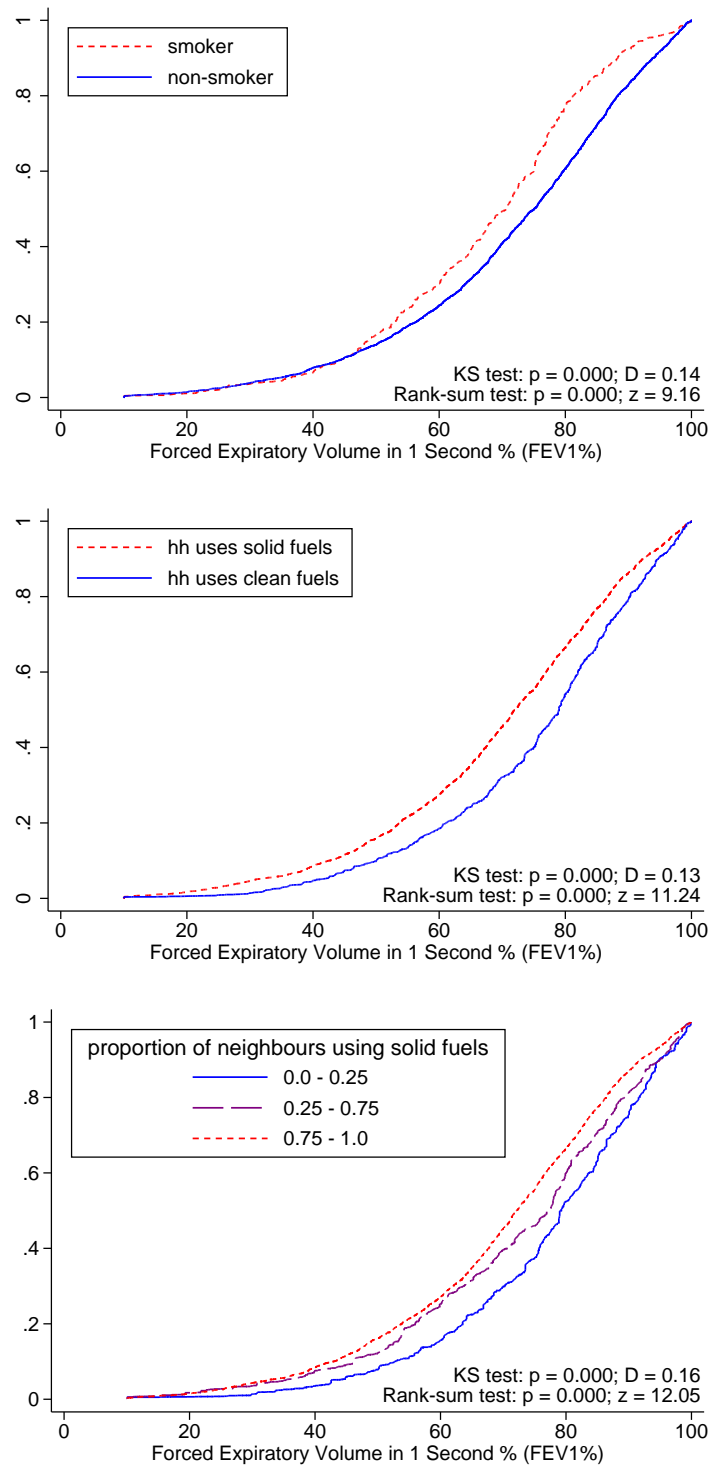
Source: SAGE 2007-08. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (liters) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). Higher values denote lesser lung obstruction. hh: household

Figure 1–2: Lung obstruction by household and neighborhood exposure, and two measures of socioeconomic status

Source: SAGE 2007-08. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (liters) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). Higher values denote lesser lung obstruction. hh: household

Figure 1–3: Lung obstruction by neighborhood solid fuel use among households that use solid fuels and households who use clean fuels

Lung obstruction by household and neighbourhood solid fuel use

Source: SAGE 2007-08. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (liters) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). Higher values denote lesser lung obstruction. hh: household

Figure 1–4: Predicted FEV1% from OLS regressions with household and neighborhood solid fuel interactions

Source: SAGE 2007-08. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (liters) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). Higher values denote lesser lung obstruction. Graphs show local polynomial regression. hh: household

25

1.5 Supplementary Appendix

Appendix table 1A–1 examines the association of smoking, household solid fuel use, and neighborhood solid fuel use with lung obstruction. The models in the table compare the association of the exposures when considered individually (models 1 - 3) and when considered together (models 4 - 6), the magnitudes of association for linear controls of age and socioeconomic status (models 1 - 5), and the magnitude of association for living in a rural area (models 1 - 6). Measures of socioeconomic status are not associated with lung obstruction. Model 6 in 1A–1 is the same as Model 2 in table 1–2. Operationalizating the measures of socioeconomic status and age as dummies improves model fit, as shown by the r^2. The table shows coefficients for the rural dummy as well, finding that respondents in rural areas are not significantly different from urban respondents once household and neighborhood solid fuel use are adjusted for.

Appendix Tables 1A–2 - 1A–6 examine the robustness of the main result with controls for additional measures of socioeconomic status, placebo tests, falsification tests, mechanisms checks, and tests with alternative measures of respiratory health. The main findings from these tables are discussed in section 3.4.

Table 1A–1: Association between smoking, household use of solid fuels, and neighborhood solid fuel use with lung obstruction: Models with linear controls for age, assets, and education

	(1)	(2)	(3)	(4)	(5)	(6)
smoke tobacco	-3.909***			-3.484**	-3.673***	-3.764***
	(1.069)			(1.066)	(1.062)	(1.040)
household uses solid fuels		-5.241***		-2.968*	-3.491**	-3.762**
		(1.427)		(1.261)	(1.230)	(1.170)
neighourhood solid fuel use			-9.188***	-6.127*	-6.424*	-6.088*
			(2.739)	(2.628)	(2.612)	(2.450)
female	-0.650	0.543	0.584	-0.367	-0.576	-0.859
	(0.936)	(0.828)	(0.828)	(0.877)	(0.939)	(0.908)
rural	-2.691*	-0.533	1.301	1.320	1.384	1.382
	(1.131)	(1.367)	(1.898)	(1.816)	(1.804)	(1.672)
age	-0.124***	-0.134***	-0.134***	-0.124***	-0.127***	
	(0.021)	(0.021)	(0.021)	(0.021)	(0.024)	
assets					-0.070	
					(0.086)	
education					-0.048	
					(0.092)	
state fixed effects	X	X	X	X	X	X
age in years (as dummies)						X
asset count (as dummies)						X
years of education (as dummies)						X
constant	74.108***	76.179***	77.875***	78.248***	80.334***	74.768***
	(2.044)	(1.993)	(2.125)	(2.140)	(2.556)	(2.619)
r-squared	0.048	0.053	0.053	0.060	0.060	0.090
Observations	9551	9551	9551	9551	9551	9551

Source: WHO SAGE 2007-08

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$. Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air one can exhale in 1 second to the total volume of air one can exhale after inhaling as deeply as possible.

Table 1A–2: Robustness I: Additional controls for household and neighborhood socioeconomic status

	(1)	(2)	(3)	(4)
smoke tobacco	-3.764***	-3.792***	-3.774***	-3.770***
	(1.04)	(1.04)	(1.04)	(1.04)
household uses solid fuels	-3.762**	-3.577**	-3.352**	-3.363**
	(1.17)	(1.15)	(1.14)	(1.14)
neighbourhood solid fuel use	-6.088*	-6.397*	-7.936**	-7.492*
	(2.45)	(2.59)	(2.78)	(3.18)
log monthly expenditure		1.012+	0.838	0.839
		(0.60)	(0.61)	(0.61)
log monthly expenditure squared		-0.667	0.255	0.284
		(1.21)	(1.32)	(1.31)
average neighbourhood assets			-0.342	-0.385
			(0.24)	(0.27)
average neighbourhood education				0.114
				(0.38)
demographic controls	X	X	X	X
place controls	X	X	X	X
ses controls	X	X	X	X
constant	74.768***	72.409***	71.927***	71.393***
	(2.62)	(10.95)	(10.83)	(10.80)
observations	9551	9551	9551	9551

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + p<.1, * p<.05, ** p<.01, *** p<.001.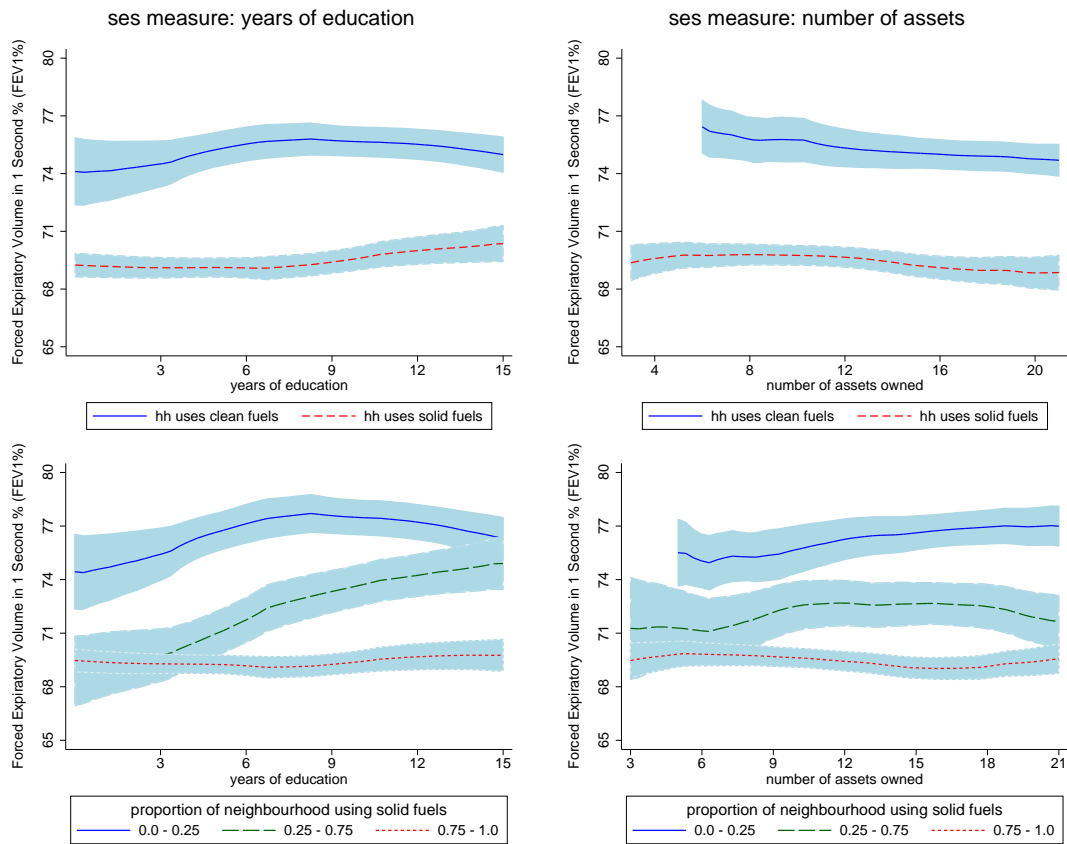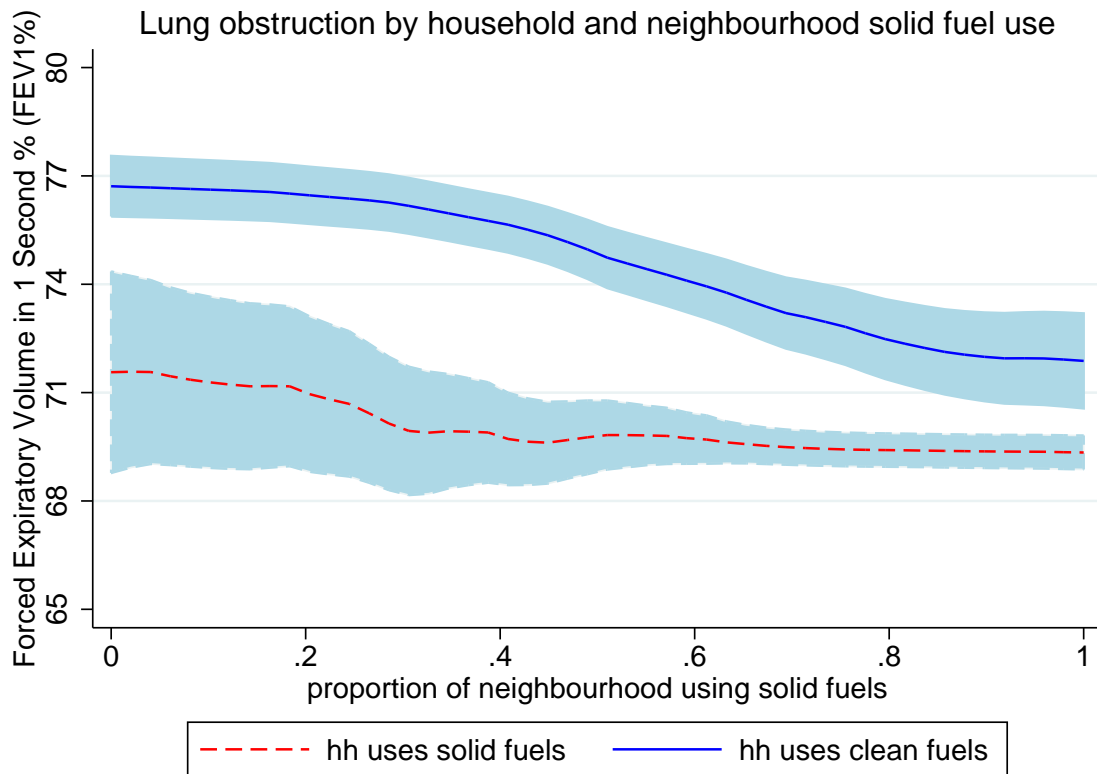 Outcome: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air one can exhale in 1 second to the total volume of air one can exhale after inhaling as deeply as possible.

Table 1A–3: Robustness II: Tests for bias in exposures

specification / sample	psu with n > 19	iv with split psu	district level solid fuel use as an exposure		kerosene separately
	(1)	(2)	(3)	(4)	(5)
smoke tobacco	-3.693**	-3.596***	-3.740***	-3.758***	-3.762***
	(1.136)	(1.069)	(1.049)	(1.041)	(1.040)
household uses solid fuels	-2.777*	-7.324***		-3.748**	-3.710**
	(1.192)	(1.817)		(1.165)	(1.176)
household uses kerosene					0.827
					(3.141)
neighbourhood solid fuel use	-6.223*	-7.595*		-5.460*	-3.710**
	(2.548)	(3.466)		(2.477)	(1.176)
proportion using solid fuels in district			-6.549*	-1.998	
			(3.242)	(2.972)	
demographic controls	X	X	X	X	X
place controls	X	X	X	X	X
ses controls	X	X	X	X	X
constant	74.702***	85.363***	74.101***	76.013***	74.702***
	(2.610)	(4.304)	(3.345)	(3.347)	(2.610)
observations	8519	9361	9551	9551	9551

Source: WHO SAGE 2007-08

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + p<.1, * p<.05, ** p<.01, *** p<.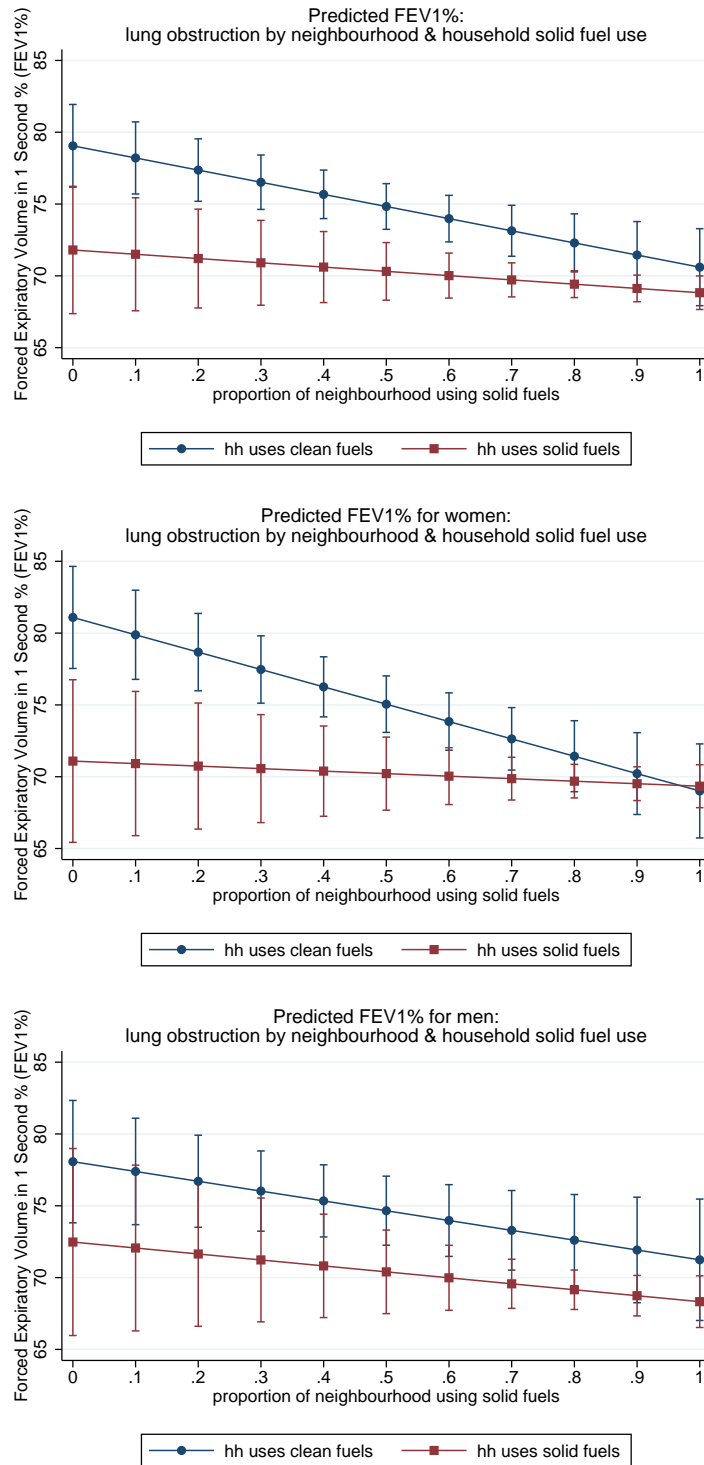001. Outcomes: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (litres) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible).

Table 1A–4: Placebo tests for smoking, household solid fuel use, and neighborhood solid fuel use

A. Smoking tobacco

	(1)	(2)
chew tobacco	-0.28 (0.86)	
drink alcohol		-1.318 (1.34)
smoke tobacco	-3.772*** (1.10)	-3.522*** (1.06)
demographic controls	X	X
place controls	X	X
ses controls	X	X
constant	77.408*** (2.99)	77.416*** (2.95)
observations	9551	9551

B. Household solid fuel use

	(1)	(2)
hh owns toilet	0.742 (0.94)	
hh electrified		0.242 (0.95)
hh uses solid fuels	-5.501*** (1.05)	-5.766*** (1.07)
demographic controls	X	X
place controls	X	X
ses controls	X	X
constant	82.242*** (3.16)	82.429*** (3.32)
observations	9551	9551

C. Neighbourhood solid fuel use

	(1)	(2)
neighbourhood open defecation	-0.178 (1.95)	
neighbourhood electrification		-0.83 (2.05)
neighbourhood solid fuel use	-7.309*** (2.20)	-7.852*** (1.83)
demographic controls	X	X
place controls	X	X
ses controls	X	X
constant	83.252*** (3.34)	84.130*** (4.00)
observations	9551	9551

Table 1A–5: Falsification tests: Association between other health outcomes and exposures

outcome:	FVC	height	word recall	walk time
	(1)	(2)	(3)	(4)
smoke tobacco	-0.038	0.002	-0.071	-0.061
	(0.058)	(0.004)	(0.105)	(0.047)
household uses solid fuels	-0.050	-0.000	-0.067	0.048
	(0.059)	(0.004)	(0.128)	(0.053)
neighbourhood solid fuel use	-0.008	0.002	-0.300	0.005
	(0.075)	(0.005)	(0.225)	(0.113)
demographic controls	X	X	X	X
place controls	X	X	X	X
ses controls	X	X	X	X
constant	2.765***	1.656***	4.725***	2.994***
	(0.414)	(0.014)	(0.275)	(0.161)
observations	9551	9551	9551	9551

Source: WHO SAGE 2007-08

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$. Outcomes: FVC, or Forced Vital Capacity is the total amount (measured in liters) of air breathed out in the spirometry test. Height is measured in meters. Word recall is based on the delayed recall of ten words. Timed walk measures the time taken (in seconds) to walk 4 meters by

Table 1A–6: Mechanism check and tests with alternative health outcomes

test:	mechanism check: satellite PM 2.5			alternative measures of respiratory health		
specification:	ols				logit	
outcome:	PM2.5	PM2.5	PM2.5	FEV1s	have COPD (FEV1% < 70%)	self-reported lung disease
	(1)	(2)	(3)	(4)	(5)	(6)
smoke tobacco	-2.962**	-3.696***		-0.189***	1.422*	1.526*
	(1.055)	(1.031)		(0.030)	(0.195)	(0.282)
household uses solid fuels	-3.314**	-3.716**		-0.139***	1.467*	1.919*
	(1.175)	(1.164)		(0.037)	(0.220)	(0.505)
neighbourhood solid fuel use	-3.733*	-6.164*	22.132***	-0.162**	1.502*	1.966*
	(1.793)	(2.449)	(6.438)	(0.061)	(0.290)	(0.668)
district level PM 2.5 (ug / m3)	-0.054*	-0.034				
	(0.022)	(0.038)				
demographic controls	X	X		X	X	X
place controls		X		X	X	X
ses controls	X	X		X	X	X
height				X		
constant	86.154***	75.595***	22.783***	0.755		
	(2.730)	(2.905)	(3.516)	(0.565)		
observations	9551	9551	9551	9551	9551	9551

Source: WHO SAGE 2007-08

Note: Robust standard errors, clustered at the level of the primary sampling unit, in parentheses. + p<.1, * p<.05, ** p<.01, *** p<.001.

Outcomes: FEV1% (FEV1s/FVC as a percentage) is the ratio of the volume of air (litres) one can exhale in 1 second to the total volume of air one can exhale (after inhaling as deeply as possible). FEV1s or Forced Expiratory Volume in 1 second (measured in liters) is the total amount of air breathed out in the first second of the spirometry test. Respondents are classified as having COPD if FEV1% was below 70%, based on GOLD criteria. Respondents classified as having lung disease based on self-reported symptoms according to WHO SAGE criteria, of having shortness of breath or experiencing coughing or wheezing for more than ten minutes and coughing sputum or phlegm for most days of a month for the past three months in a year.

Figure 1A–1: Histogram of number of respondents per cluster
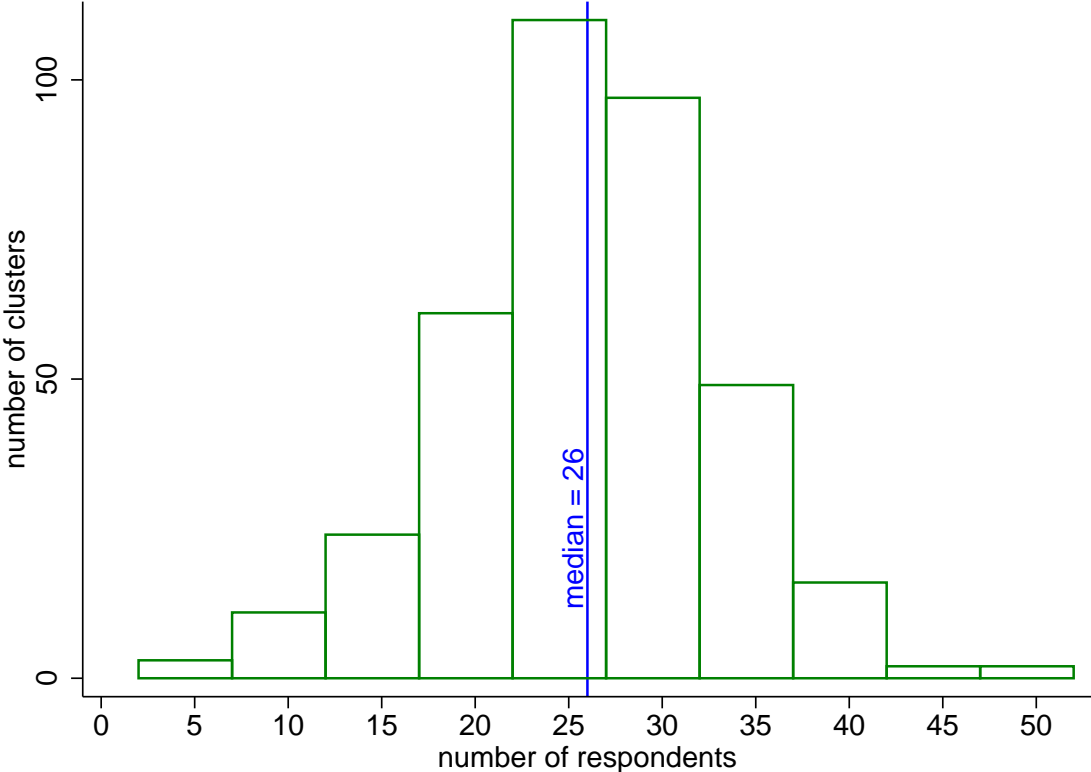
CHAPTER 2:

Seasonal variation in infant mortality in India

Abstract

Investigating seasonal variation in health helps understand interactions between population, environment, and disease. Using information on birth month-year, survival status within the first year of life, and age (in months) at death (if the infant died) of more than 330,000 children observed in the four rounds of India's Demographic and Health Surveys, I estimate period mortality rates between ages 0 and 1 ($_1m_0$) by calendar month. Relative to the spring months, infant mortality is higher in the summer, monsoon, and winter months. If the mortality conditions in spring were prevalent throughout the year, $_1m_0$ would have been reduced by 11.4 deaths per 1,000 in the early 1990s and 3.7 deaths per 1,000 in the mid-2010s. Seasonal variation in infant mortality has declined in both absolute and relative terms, but remains higher among disadvantaged children. The results highlight the multiple environmental health threats that Indian infants face, and the short time of a year when these threats are less salient.

2.1 Introduction

The recognition that seasonal variations influence health and mortality goes at least as far back as 4th century BCE. Hippocrates is said to have written that "whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year" (Adams, 1849). Seasonal influences on health were also recognized in ancient Indian and Chinese texts, such as the *Sushrut Sanhita* and *Huangdi Neijing* (Bhishagratna, 1963; Veith, 2015). In the 10th century, Ibn Sina, highlighting several mechanisms of seasonal influences in health, wrote in *al-Qanun fi al-Tibb* (canon of medicine) that "change of seasons has to do with the kind of diseases peculiar to each climate" (Gruner, 1984).

In recent years, a considerable literature in the economic, social, and health sciences has investigated monthly and seasonal variation in health (Rau, 2006b), especially early in life (Muhuri, 1996; Delaunay et al., 2001; Farrar et al., 2019). This literature has contributed to the scientific understanding of health determinants, dynamics, and the role of public policy. For instance, prior studies have documented reductions in seasonal variation along the epidemiological transition in both historical and contemporary contexts (Delaunay et al., 2001; Rau, 2006b; Schlüter et al., 2020). They have also shed light on vulnerability to climactic factors (Rayco-Solon et al., 2004; Muhuri, 1996); identified the importance of household technologies, such as those related to temperature control (Kunst et al., 1991); motivated social security interventions in lean seasons (Behrman, 1988b; Drèze, 1990; Kumar et al., 2016); and quantified the likely impact of health interventions targeting seasonal diseases, such as malaria (Etard et al., 2004), influenza (Dorélien, 2019; Ho and Noymer, 2017), and diarrhea (Farrar et al., 2019).

Investigating seasonal variation in child health in low- and middle-income countries is an important priority within this literature for several reasons. First, children are particularly vulnerable to changes in the disease environment (Jutla et al., 2013; Geruso and Spears, 2018a), and these vulnerabilities are likely to be higher in developing country contexts. Second, insults in childhood have adult health consequences (Elo and Preston,

1992; Doblhammer and Vaupel, 2001; Currie and Vogl, 2013). Thus, recognizing the timing of insults is critical to understanding interrelations of health across the life course. Third, reducing child mortality is a global health priority, as highlighted in the Sustainable Development Goals (United Nations, 2015). Highlighting seasonal variation can motivate particular policy and behavioral interventions, such as for interventions that target seasonal diseases. Given increases in the provision of maternal and child health in low and middle income countries (UNICEF, 2019), healthcare systems and interventions could also prepare for seasonal healthcare needs.

Prior research has identified four broad categories of sources of seasonality in child health in low- and middle-income contexts. First, economic and agricultural activity in developing countries varies seasonally (Devereux et al., 2013; Chambers et al., 1981). Increased economic vulnerabilities and reduced access to nutrition is linked to morbidity and mortality (Rao et al., 2009). Second, environmental exposures, such as pollution, heat, humidity, and cold weather have direct impacts on health and they vary seasonally (Geruso and Spears, 2018a; Guttikunda and Calori, 2013; Kumar et al., 2009). Third, disease dynamics, particularly for communicable diseases, are driven by individual behaviors, environmental effects on hosts and pathogens, and their interactions (Fisman, 2007). In India, pneumonia and diarrhea, leading causes of infant mortality (Fadel et al., 2017), have been documented to vary seasonally (Farrar et al., 2019). A fourth source of seasonal variation in child health is seasonal variation in births (Dorélien, 2016; Rumisha et al., 2013; Pitt and Sigle, 1998). Seasonal variation in births has a compositional effect on seasonal estimates of infant mortality - some seasons have higher births, and thus a younger age-profile among infants, leading to more deaths (Knodel, 1983). Like seasonal diseases, seasonal variation in births may also increase the demand for healthcare. Given these multiple contributors, Fisman (2007) described seasonality in health and disease as "the final common expression of environmental and population patterns."

Studies on historical patterns of seasonality in contemporary high-income coun-

tries relied on population level mortality records, which in many cases, span several centuries (Rau, 2006b). In low- and middle-income contexts, the demographic literature has documented a high-degree of seasonal variation based on sample registration (Farrar et al., 2019; Burkart et al., 2011) or surveillance site data (Muhuri, 1996; Kampe et al., 2015; Delaunay et al., 2001; Rumisha et al., 2013). Although valuable, studies from surveillance sites study localized populations, and sample registration system often lack key information, such as the population exposed to the risk of death or household characteristics. Demographic and Health Surveys (DHS), which record calendar birth month, survival, and age at death in months for infants are a relatively under-utilized source in studies of seasonal variation in infant mortality. This omission is despite several advantages of examining seasonality in child mortality from these data. The DHS are more representative than surveillance sites, they measure a greater number of child health determinants than sample registration systems, and they have increasingly larger sample sizes, allowing for more precise estimation.

Using data from the DHS, this paper presents the first estimates of all-cause infant mortality by calendar-month for India, home to every sixth birth and every fifth infant death in the world according to United Nations (2015). I estimate period mortality rates between the ages 0 and 1 ($_1m_0$) by calendar month. To account for seasonal variation in births, I compute age-standardized period-specific $_1m_0$ by calendar month. The approach developed here examines seasonality in infant mortality over two decades, a period in which overall infant and child mortality have declined. Methodologically, the paper complements Pedersen and Liu (2012), finding that the Demographic and Health Surveys can be used to estimate mortality rates for shorter periods than the five year estimates that DHS surveys typically estimate. Substantively, the paper contributes to the literature on environmental determinants of child health in developing countries. To my knowledge, this is the first paper to estimate infant mortality by calendar month using the DHS in India.

In the next section, I summarize the prior literature on seasonal variation in health and mortality. Section 2.3 presents data and methods, section 2.4 the main results, and section 2.5 presents several robustness checks. Section 2.6 concludes.

2.2 Background

Seasonal variation in mortality is a result of complex interactions between behaviors, environmental exposures, effects on hosts and pathogens, and economic conditions (Chambers et al., 1981; Fisman, 2007; Rau, 2006b). The epidemiological transition and other social developments have also shaped seasonal variation in mortality. Before the Industrial Revolution, mortality in England used to be higher in the spring and winter months, and lower in the summer months (Wrigley et al., 1997). These seasonal patterns could be seen for all ages in other European countries as well and were relatively stable before the industrial revolution, with little attenuation over time. In London, as well as in French urban areas (Ariès, 1971), however, mortality used to be the highest in summer, likely due to infectious diseases. By the early 20th century, this summer peak had vanished due to improvements in hygiene and living standards (Rau, 2006b). During the rest of the 20th century, mortality differences between summer and winter declined as well, particularly after the decrease in respiratory track infections, possibly due to a reduction in air pollution (Kunst et al., 1991). Sakamoto-Momiyama (1978) finds a similar "deseasonalition" in mortality between 1900-1970 in the United States and Japan. Even then, Ho and Noymer (2017) find that in the US, pseudoseasonal life expectancy for winter months was higher by about a year than the summer months for the period 1960-2014.

In developing countries, data limitations do not permit such a long historical perspective on seasonal variations in mortality. However, data from Health and Demographic Surveillance Sites (HDSS) in sub-Saharan Africa, stretching as far back as the 1950s, have been used by researchers to analyze patterns of seasonal variation in child mortality. Rayco-Solon et al. (2004) show that in Gambia, prior to 1975, deaths in the

'hungry' season (July-November) were 1.9 times as likely as deaths in the 'harvest' season (December-June). These odds reduced to 1.8 for the period 1975-1984, and 1.2 for the period 1984-1997 (Rayco-Solon et al., 2004). In Senegal, (Delaunay et al., 2001) found that although seasonality in mortality declined between 1960s and 1980s, it increased in the 1990s. These results suggest that seasonal patterns in mortality may not decline linearly.

To the extent that information on causes of death are available from surveillance sites, researchers have found that in Sub-Saharan Africa, along with diarrhea and pneumonia, malaria is a big component of both overall and seasonal variation in child mortality (Abdullah et al., 2007). In Senegal, mortality was highest in the rainy season of August-September (Delaunay et al., 2001). Etard et al. (2004) use verbal-autopsy information to find that the increase in seasonality was due to an increase in mortality due to malaria, diarrhea, fevers, and acute respiratory infections. Kampe et al. (2015) find a similar pattern of higher child-mortality in the rainy season months of June-October in an HDSS in Burkina Faso. These seasonal patterns were driven by malaria mortality among children, with no decline in mortality between 1998-2007. Abdullah et al. (2007) and Mutisya et al. (2010) confirm similar patterns from seven other surveillance sites in Sub-Saharan Africa, finding higher malaria mortality in the rainy seasons. Although it is well recognized that seasonal variation in Sub-Saharan Africa is large (Dorélien, 2016), studies have largely focused on the proportion of deaths in a year that occur in a month or season rather than on mortality rates.

In South Asia, studies based on both surveillance and sample registration data also find that child mortality varies seasonally. Becker (1981), Muhuri (1996), and Becker and Weng (1998) estimate seasonal mortality patterns for the Bangladesh HDSS site in Matlab for the periods 1972-1974, 1981-1982, and 1982-1990 respectively. All three papers found marked seasonality in child deaths in this area despite overall child mortality declines (Becker and Weng, 1998), as a result in part of many interventions to reduce child mortality (Chen et al., 1983). Major causes of death, such as diarrhea, respiratory diseases,

injuries, and 'others', showed similar seasonal patterns. These patterns are also confirmed in sample registration data from Bangladesh for the period 2002-2007 by Burkart et al. (2011). Burkart et al. (2011) further find higher seasonal variation in child mortality for rural areas and families with lower socio-economic status.

Farrar et al. (2019) use sample registration data with causes of death coded on the basis of verbal autopsies to examine seasonal variation in death counts at the ages 1-59 months in India. These results are based on the Million Deaths Study (Gomes et al., 2017). They show that pneumonia deaths are highest in winter months in India, but also high in the summer and monsoon months. Diarrhea deaths, on the other hand, are highest in the summer and monsoon months, but also high in the winter months. Farrar et al. (2019) find that although child mortality declined in the period 2005-2013, seasonality in both pneumonia and diarrhea mortality still remains.

Higher mortality from these infectious diseases in the summer, monsoon, and winter months confirms known seasonal patterns of environmental exposures such as heat, humidity, cold temperatures, and air pollution. Heat and humidity is highest in the monsoon months in India, and exposure to fecal germs may also be high in the monsoon months (Pascual et al., 2002). Air pollution in India is caused by both perennial factors such as industrial activity, transportation, and solid fuel use for cooking, as well as seasonal activities such as stubble burning and solid fuel use for heating (Apte and Pant, 2019). In addition to seasonal sources of pollution, meteorological factors such as inversion which trap pollution close to the ground and are more likely to be observed in the winter months (Tiwari et al., 2013) cause higher exposure to air pollution in the these months.

In contrast to these studies finding evidence of seasonal variation in child mortality, studies using data from the DHS on child heights have had more mixed findings. For 40 developing countries including India, Agarwal et al. (2017) show that there is no correlation between height-for-age and month-of-birth. However, it is worth noting that this literature is concerned with month-of-birth effects, rather than seasonal variation in

child health. Indeed, if a worse disease environment leads to higher mortality in particular seasons of a year, we may expect higher growth deficits if children were 'scarred' by the worse disease environment in particular months (Preston et al., 1998). However, to the extent that children who are more fragile die because of worse disease environment, the surviving cohort born in those months would likely be healthier. This selection mechanism linking early life circumstances and later health in (Preston et al., 1998) may in part account for the above findings.

Although there are exceptions, the social science literature on examining seasonal variation in mortality in India has preferred the approach of measuring mortality over calendar months, rather than seasons (Farrar et al., 2019; Larsen et al., 2019). This has advantages and disadvantages. Measuring mortality over longer periods will pool information, and seasonal estimates will have lower sampling variation than calendar month estimates. However, measuring mortality over calendar months allows for a finer understanding of seasonal variation. It also allows for greater flexibility, considering that seasons may not be uniform across India.

India is in the northern hemisphere. Although the intensity of the summer and winter months is influenced by wind patterns and varies over India's regions, these seasons are common across the country (Attri and Tyagi, 2010). In addition to the monsoon, Humidity is higher in India's coastal regions. The south-west monsoon varies within India, beginning in June in South India and proceeding to northern India by July, and withdrawing in September. Parts of South India receive rains from the receding monsoon, from September to November. Given the year to year variation in these seasonal patterns, as well as their variation across India, this paper also examines mortality by calendar month. To do this, the paper develops a methodological approach to estimate period $_1m_0$ by calendar month. This allows a flexible approach – we study regional patterns as well as broader seasonal patterns that are common across India.

2.3 Materials & methods

Data

The data used in this paper come from the National Family Health Surveys (NFHS), India's version of the DHS. I use the four existing waves of the NFHS, conducted in the years 1992-93, 1998-99, 2005-06, and 2015-2016. These multi-stage sample surveys, which are representative at the national and state level, obtain retrospective birth histories from women aged 15-49 years. In recalling their complete birth history, women surveyed are asked about:

- all the children they have ever had
- the month and year of birth of each child
- whether the child died
- age at death if the child died

For children who died within the first month of life, the DHS asks about the number of days the child was alive. For deaths in the first two years, the DHS asked about the number of months a child was alive. For deaths after age two years, the DHS asks about the number of years a child was alive.

To estimate mortality by calendar month (or any period), we need to estimate the number of deaths in a particular month (period), and the number of child-months (person-periods) lived in that month. Because the DHS asked the number of months a child was alive only for children who died before the age of 2, mortality rates by calendar month can only be estimated for children under the age of 2. However, because the bulk of child mortality in India is before the age of 1 (Hill et al., 2012), in this paper, I estimate period $_1m_0$ for each calendar month. I do so by using the information on children's month and year of birth; her/his survival status; and if she/he died before the age of 1, the number of months they were alive. The next sections describes this approach in detail.

Approach: Creating a child-month dataset

Using the DHS birth history module, I create estimates of infant mortality by calendar month. In the birth history module, each observation is a child. I convert this child dataset to a child-month dataset, where each observation is a child-month.

First, I assign every child 12 observations, one for each month in the first year of life they can be alive (0-11 months). This creates a child-month dataset. In this dataset, a child born in November 1989 will have 12 observations: one for November 1989, one for December 1989, one for January 1990, and so on, until October 1990. For each of these months, I also calculate the age (last month) the child attained, in months. So, the child described above will be aged 0 months in November 1989, 1 month in December 1989, 2 months in January 1990, 3 months in February 1990, and so on, until October 1990, when they will be aged 11 months.

Second, I calculate the month and year of death for children who died before the age of 1. I do this by adding the age at death (in months) to their birth month and year. So, a child born in November 1989 who lived for 5 months would have died in April 1990. A child who died in the first month of life was assumed to have died in the birth month.

Third, I create an indicator for whether a child was alive in a particular month or not. Thus, for this child who died at age 5 months in April 1990, the indicator for whether they died in a particular month will be zero for months 0 to 4. In the linear regression approach I describe below, for month 5, the indicator will be 1,000. This scaling of the binary indicator for whether a child died in that month to 0 or 1000 allows the interpretation of mortality estimates as deaths per 1,000 infants alive. I also create a dummy variable indicating 0 for alive and 1 for died in a month to use in a logistic and complementary-log-log (cloglog) regression framework.

Finally, I censor those-child months that are after the observation window, i.e. child months during or after the month-year of interview. Similarly, child-months after

the death of a child are censored – i.e. child months a child did not live. After these steps, we have a child month dataset that has the following information for each child-month:

- The month and year of observation
- The child's survival status (dead or alive) in that month and year of observation
- The child's age (in months) in that month and year of observation
- Other attributes of the child, such as sex, rural residence, or region

This approach assumes that all births in the middle of the month. Deaths are assumed to be in the calendar-month an age is attained. It is compatible with other assumptions as well, which are listed in the supplementary appendix.

Estimates of early life mortality using the DHS are typically calculated for the period five years prior to the survey, to minimize recall errors. I follow this approach, and use in estimation all children born in the last five years. But I restrict estimates of infant mortality to the four years prior to the survey so that I can calculate both numerators (infants who died in the month) and denominators (children alive in the month) correctly. This approach assures that each calendar month in the analysis period has person-months contributed from children of all ages from zero to 11 months.

Another exclusion criterion is that only those months are included in the analysis in which children from all of India's states contribute person months. The DHS surveys in India are conducted in a phased manner and interviewers visit states at different times within the survey period. Thus, for some initial months within the five year period, only a few states contribute observations. Those months are dropped. This final criteria ensures that estimates of monthly infant mortality are not driven by differences in the states that contribute deaths and person-months to the calendar months.

An example using a Lexis diagram

This approach can be visualized in the Lexis diagram. Figure 2–1 helps understand the structure of the dataset and the strategy for estimating period $_1m_0$ by calendar month. The

Lexis diagram shows lifelines for 15 children. The children were born in fifteen different months, starting from January 1991 to March 1992. The Lexis diagram follows them for each month until they reach the age of 1 year, or until they die. Out of the 15 children followed here, 4 died before reaching the age of 1 year. Thus, child 1, 2, and 3 survived until age 1 year, but child 4 died at the age of 4 months, in August 1991. Child 5 lived for a full 12 months, child 6 died in the 1st month of their life, when their age was 0 month, and so on.

We are interested in estimating period-specific $_1m_0$ for each calendar month. In this example, the first month for which we should estimate $_1m_0$ is December 1991. This is because December 1991 is the first month for which all children who could contribute child-months actually do contribute person months. In this example, however, consider $_1m_0$ for the month of January 1992. In this month, one child (child 10) died at the age of 3 months, while 10 children contributed person months. Thus, in this example, $_1m_0$ for January 1992 would be 1 death per 10 children, or 100 deaths per 1,000.

Finally, note that for January 1992, we have half a month contributed by child 1 to January 1992, and half a month contributed by child 13. In practice, the half-month contributed by child 1 is not counted, and the half month counted by child 13 is counted as a full month. In this, we have assumed that the child-months that children born in January 1991 would have contributed to January 1992 are equal to the extra person-months assigned to children born in January 1992 in of our approach (Goldman et al., 1984).

Estimating unadjusted and age-standardized $_1m_0$ by calendar month

Age-specific mortality rates for a particular period and place take the form:

$$_nm_x = \frac{_nd_x}{_nL_x} \, , \tag{2.1}$$

where $_nm_x$ is the mortality rate between ages x and $x + n$, calculated as the number of

deaths (d) in ages x to $x + n$ in the time period under consideration divided by the person-periods ($_{n}L_{x}$) lived in that time period. In our case, calculating mortality rates between the ages 0 and 1 for calendar month t, equation 2.1 can be modified to:

$$_{1}m_{0}^{t} = \frac{_{1}deaths_{0}^{t}}{_{1}child\,months\,lived_{0}^{t}},$$ (2.2)

We approximate $_{1}child\,months\,lived_{0}{}^{t}$, the number of child-months lived between the ages 0 and 1 years in month t by estimating the number of children alive in the middle of month t. This approach of using the middle of the month as an estimate of the child-months lived in the year is based on the assumption that person-months grow linearly within the month (Preston et al., 2000). Thus, in the approach followed in this paper, period mortality rates for each calendar month are estimated as:

$$_{1}m_{0}^{t} = \frac{_{1}deaths_{0}^{t}}{_{1}children\,alive\,in\,the\,middle\,of\,the\,month_{0}^{t}},$$ (2.3)

Since I use individual level data to compute these period mortality rates, I weight individual level data by survey weights that the NFHS calculated. However, the first three NFHS surveys had much smaller sample sizes than the latest NFHS survey (NFHS-4). NFHS 1, 2 and 3 were representative at the level of the states in India, whereas NFHS-4 is representative for districts in India. To avoid biasing estimates towards the NFHS-4, I give each survey an equal weight, of 0.25. Thus, for each observation in NFHS-1, I divided the survey weight by .1271 and multiplied them by 0.25, because NFHS-1 contributed 12.71% of the weighted observations in the overall sample. Similarly, weights of NFHS-2 observations were divided by .1571 and multiplied by 0.25, weights of NFHS-3 observations were divided by .1313 and multiplied by 0.25, and weights of observations were divided by 0.5909 and multiplied by 0.25. The weighted mortality rate in the age 0-1 in calendar month t can be calculated as:

$$_1m_0^t = \frac{\sum_{i=1}^{I}(died\ between\ age\ 0\ to\ 1\ in\ month\ t)_i \times weight_i}{\sum_{i=1}^{I}(children\ between\ age\ 0\ and\ 1\ alive\ in\ middle\ of\ month\ t)_i \times weight_i} \quad (2.4)$$

Births also vary seasonally in India (Appendix Figure 2A–1, Anand et al. 2000; Dyson 1991). To adjust for monthly variation in the age distribution of infants in each calendar month, I also compute age-standardized $_1m_0$ by calendar month. Seasonal fluctuations in births create a high-proportion of younger children in some calendar-months relative to other calendar months. Mortality declines rapidly even in the first year of life (Guillot et al., 2012). Thus, a higher proportion of younger children is some calendar-months relative to others can lead to a higher mortality in those months. To compute age-standardized $_1m_0$ by calendar month, I use a reweighting standardization technique developed by Di-Nardo et al. (1996) and introduced to the demography literature by Geruso (2012). This reweighting technique is similar to direct age-standardization (Preston et al., 2000), except that the reweighting technique is applied to individual level data. I re-weight the distribution of ages in each calendar month to match the overall distribution of ages in that survey round. For this, I first estimate a reweighting function for each calendar month t in survey round r. This reweighting function takes the form:

$$\psi^{t,r}(a) = \frac{f(a|r)}{f(a|t)}, \quad (2.5)$$

where f represents the probability distribution function and a represents age (in months). This function standardizes the distribution of ages in each calendar month such that each calendar-month's age distribution matches the overall age-distribution in the survey round. Age-standardized $_1m_0$* is calculated in this approach as:

$$_1m_0^{*t} = \frac{\sum_{i=1}^{I}(died\ between\ age\ 0\ to\ 1\ in\ month\ t)_i \times weight_i \times \psi^{t,r}(a_i)}{\sum_{i=1}^{I}(children\ age\ 0\ to\ 1\ alive\ in\ middle\ of\ month\ t)_i \times weight_i \times \psi^{t,r}(a_i)} \quad (2.6)$$

47

For both the estimated (equation 2.4) and age-standardized mortality rates (equation 2.6), I calculate standard errors accounting for the clustering of observations within primary sampling units.

The approach used here is compatible with other assumptions, and robust to alternative methodological approaches as well. The supplementary appendix discusses this further.

Regression based approaches to estimate relative differences in infant mortality by calendar month

Because we have a child-month dataset with a binary outcome (survived or died), regression based approaches can be used to control for covariates and examine if seasonal variation differs by observed determinants of child health. I use three regression approaches: a linear probability model which examines if there are mean differences in mortality by calendar month relative to a reference month; a logistic regression model which estimates odds-ratios for the probability of dying in a particular month compared to a reference month; and a complementary-log-log model which computes hazard-ratios of the probability of mortality in a particular month compared to a reference month.

The linear probability regressions take the form:

$$
\begin{aligned}
d_{ijt} = \sum_{m=1}^{11} \beta_m \mathbb{1}[calendar\ month_{ijt} = m] + \gamma age_{ijt} + \phi period_{ijt} \\
+ \theta place_{ij} + \lambda female_{ij} + \rho survey\ round_{ij} + \varepsilon_{ijt},
\end{aligned}
\tag{2.7}
$$

where d is a binary indicator of whether a child i in primary sampling unit j in month t died or not. The β_ms are the primary coefficients of interest, indicating the difference in the probability of death in a particular month and a reference month. The reference calendar month used in this analysis is March. In alternative specifications, I control for age-in-month dummies, period (measured as linear year, linear month-year, or year dummies),

dummies for place (such as rural residence and state or region), sex of the child, and the survey round. Note that the adjusting for dummies for age-in-months in regression based approaches is equivalent to the age-standardization approach described earlier (Case and Paxson, 2005).

Adjusting for the period allows us to account for the declining trend of infant mortality in the overall period of 1989-2014. Dummies for age in months allow us to account for the pattern of mortality by age within the first year. This controls for age distributional differences between months. It creates estimates of infant mortality by month that are analogous to age-standardized estimates in equation 2.6. Because the interview month can influence the recall of birth or death months (Larsen et al., 2019), I also control for interview month dummies. I control for dummies for survey round to account for measurement error associated with the implementation of a particular wave of the NFHS surveys.

To examine differences in seasonal variation by time characteristics such as rural-urban residence, mother's education, household wealth, sex of the child, and region, I also estimate stratified regressions. In these regressions, the sample is limited to those belonging to a particular characteristic. These regressions, for instance, compare mortality for girls in a particular calendar month against mortality for girls in a reference month. More formally, these regressions take the form

$$
\begin{aligned}
d_{ijt}|c = \sum_{m=1}^{11} \beta_m \mathbb{1}[calendar\ month_{ijt} = m] + \gamma age_{ijt} + \phi period_{ijt} \\
+ \theta place_{ij} + \lambda female_{ij} + \rho survey\ round_{ij} + \varepsilon_{ijt},
\end{aligned}
\tag{2.8}
$$

where c is the characteristic to which the sample has been limited to.

The logistic regression is estimated with the same set of covariates, and takes the form:

$$log\left(\frac{P_{ijt}}{1 - P_{ijt}}\right) = \sum_{m=1}^{11} \beta_m \mathbb{1}[calendar\ month_{ijt} = m] + \gamma age_{ijt} + \phi period_{ijt}$$
$$+ \theta place_{ij} + \lambda female_{ij} + \rho survey\ round_{ij} + \varepsilon_{ijt}, \tag{2.9}$$

P_{ijt} is the probability of death in month t for child i in primary sampling unit j. This logistic approach is exactly similar to the one followed by Rau (2006a) to understand seasonality in Sweden, except that the approach followed here uses dummies for calendar months instead of dummies for seasons. In a survival analysis sense, the model used above is equivalent to a discrete-time hazard model with a logistic link function (Singer and Willett, 2003). I report odds-ratios, which should be interpreted as the odds of death in month t divided by the odds of death in March.

Finally, I estimate a hazard model with a complementary log-log link with the same explanatory variables. This cloglog model makes an assumption of proportional hazards and not proportional odds (Singer and Willett, 2003). The cloglog model has the advantage that it allows the interpretation of hazards as if they were in continuous time. The cloglog model takes the form:

$$log(-log(1 - P_{ijt})) = \sum_{m=1}^{11} \beta_m \mathbb{1}[calendar\ month_{ijt} = m] + \gamma age_{ijt} + \phi period_{ijt}$$
$$+ \theta place_{ij} + \lambda female_{ij} + \rho survey\ round_{ij} + \varepsilon_{ijt}, \tag{2.10}$$

For this model as well, I report exponentiated coefficients, which are hazard ratios (or relative risks) relative to the month of March. An advantage of the logistic and complementary log-log approaches over the linear regression approach is that the constant is not predicted at theoretically implausible values. An advantage of the linear regression approach is that model comparisons are straightforward (Mood, 2010).

In all regressions, I cluster standard errors at the level of the primary sampling unit. Although there are multiple observations per child in this dataset, all the observed child-months for a child are still within a cluster. I also use calculated survey weights, as described in section 2.3 (equation 2.4).

2.4 Results

Summary statistics

Table 3A–1 shows summary statistics for the overall sample as well as for each of the survey rounds. Estimates for each of the NFHS rounds are weighted using the national women's weights. Overall estimates are weighted by the calculated weights, which divide the weights of the individual NFHS rounds in such a way that they contribute equally to the overall means. About 40,000 children were observed in the first three NFHS waves, and more than 200,000 were observed in NFHS-4.

Because of censoring and because some children died, we observe the average child for about 8.8 months. Overall, 6% of the children died in the four NFHS waves. The proportion of children who died declined from 7% to 4% between the 4 rounds, reflecting overall improvements in infant health (Drèze and Sen, 2013). It is worth pointing out that overall infant mortality levels observed in these rounds for close to the infant mortality rates from the Sample Registration System (ORGI, 2016), as well as infant mortality rates from the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME, 2019). Auto-correlation for the outcome of death in a month was examined, and the patterns suggested seasonal variation. Girls were slightly less than half the sample. In all the four rounds, more than 70% of the children were in rural areas.

I divide the all India sample into 4 contiguous regions: North (comprising the states of Jammu and Kashmir, Himachal Pradesh, Haryana, Punjab, and Delhi); Central (comprising the states Uttarakhand, Uttar Pradesh, Rajasthan, Madhya Pradesh, Chattisgarh, Gujarat, Bihar, and Jharkhand); East (comprising the states Odisha, West Ben-

gal, Sikkim, Manipur, Meghalaya, Arunachal Pradesh, Tripura, and Mizoram), and South (comprising the states Andhra Pradesh, Telangana, Maharashtra, Goa, Tamil Nadu, Kerala, and Karnataka). This classification, which closely follows classical patterns of Indian regional demography (Dyson and Moore, 1983), divides India into regions that are poorer and have a higher burden of disease (the central and the eastern states), and the richer and less resource-poor southern states. More than half of all children were observed in the central states, 27% in the southern states, 14% in the central states, and 6% in the northern states.

I also divide the states based on their broad climate patterns, following Farrar et al. (2019). NFHS-1, 2, and 3 do not make available the location of the primary sampling unit. I match locations of NFHS-4 primary sampling units to the 2017 Koppen-Geiger classification, based on Beck et al. (2018).

For measures of socioeconomic status, I consider mother's education and the assets a family owns. I divide mother's education into three categories: those with no education, between 1-8 years of education, and more than 8 years of education. Over the course of 25 years, education has increased. In 1992-93, 13% of mothers had more than 8 years of education, and 65% had no education. By 2015-16, 39% had more than 8 years of education, and 30% had no education. In assets, I consider the following assets: two-wheeler, car, and refrigerator. These three assets were consistently asked in the four survey rounds, are not directly linked to health, and have been consistent markers of wealth throughout this period. Because of the small proportion of households owning these assets in earlier NFHS rounds, I dichotomize the asset count into having at least one asset. About 14% of households owned any of these assets in 1992-93, 21% in 1998-99, 32% in 2005-06, and 74% in 2015-16.

Table 2–2 also shows summary statistics, but for the overall data at the level of the observations, i.e. for the child months observed in the sample. Overall, we observed more than 2.9 million child months in the four NFHS rounds, and 12,479 deaths of children

below the age of 1. The table shows the proportion of these total child-months and mean deaths that were observed in each calendar-month. The distribution of child-months over the calendar months reflect pattern of fertility, survey timing, and censoring. Similarly, the observed mean deaths by calendar months reflects patterns of seasonal variation in mortality and the age distribution of children within each calendar month.

Figure 2–2 shows estimated $_1m_0$ for each month-year included in the final analytical sample. These rates, which are not age-standardized, show a high degree of variation in $_1m_0$, especially for the first three NFHS rounds, in the 1990s and 2000s. This variation has declined substantially only by NFHS 4, although there is still evidence of seasonal variation in infant mortality.

Descriptive evidence of seasonal variation

What are the overall seasonal patterns of infant mortality in India? Are differences between calendar months, observed in Figure 2–2 substantial? Figure 2–3 answers these questions. It shows estimated raw and age-standardized rates, calculated following the approaches described for equations 2.4 and 2.6, respectively.

The blue circles, which show estimated raw rates, show a high degree of seasonal variation. Infant mortality is lowest in the spring months of February, March, and April (when summer begins) and highest in the monsoon month of August. Mortality is also high in the winter (December, January) and summer months (May, June). However, age-standardized rates, in red squares, show that at least some of the higher mortality in the monsoon months is because of a higher proportion of younger infants in those months. The age-standardized rates show highest mortality in the summer month of May and the winter month of January, which are higher than rates in the spring months of March. Although the confidence intervals overlap slightly, age standardized $_1m_0$ is also higher in the peak monsoon month of August.

From a health system perspective, both the raw and the age-standardized rates

may matter. Birth seasonality implies a higher need for care for some months. That mortality rates are higher in some months because of birth seasonality is also a relevant health policy concern. More specifically, birth seasonality implies that the health-care system would need to prepare for a higher number of births and accompanying care in some months rather than others. Higher age-standardized mortality rates in some months rather than others, on the other hand, imply a role for addressing the seasonal determinants of mortality.

Evidence from regression based approaches

In the mean $_1m_0$ by calendar month reported in Figure 2–3, some of the variation in mortality rates is due to the variation in mortality rates by survey rounds and regions. To partial out this variation, I turn to regression based approaches. Table 2–3 shows results from linear (models 1-4), logistic (model 5) and complementary-log-log (model 6) regressions of the association between calendar months and infant mortality. All models adjust for calculated survey weights and report standard errors clustered at the level of the primary sampling unit.

For the linear probability model, model 1 does not control for any covariates, replicating the raw estimated infant mortality rates from Figure 2–2. Model 2 controls for demographic covariates: dummies for the age of the child (0-11) and sex of the child. Controlling for dummies of age (in months) is equivalent to age-standardization. Model 3 additionally controls for place: the state the child was born into and rural or urban residence. Model 4 controls for time variables: the interview month, the survey round, linear month-year. Observations are child-months. The mean IMR was 4.32 infant deaths per 1,000 per month (51.8 infant deaths per 1,000 per year).

The main differences in the linear probability model are in Models 1 and 2, which replicate the differences between raw and age-standardized $_1m_0$ in Figure 2–2. Between models 2-4, adding additional controls for state, rural residence, survey round, interview month, or linear month year do not change the results. These models show that relative

to March, infant mortality is higher in the winter months of January, October, November, December; the summer month of May, and the monsoon months of July and August. Marginal differences between March and these months range from 0.7 to 1 death per 1,000 per month. This difference in substantial: it translates to at least 8.4 additional infant deaths per 1,000 per year. This difference itself is greater than the total infant mortality in high-income countries in this period (UN inter-agency group for child mortality estimation, 2019).

Models 5 and 6 confirm these findings using a logistic and complementary-log-log link in these hazard models. Both models 5 and 6 control for all the covariates that model 4 controls for. Broadly, the relative odds and relative risks for mortality are between 1.1 to 1.2 times higher for the summer, monsoon, and winter months than for the month of March. May and July have the highest mortality, followed by the winter months of November, December, and January.

Figure 2–4 examines if seasonal variation can be observed in both neonatal and post-neonatal mortality. Because neonatal mortality is much higher than post-neonatal mortality, the figure reports results from logistic regression, which examines odds of neonatal (or post-neonatal) mortality in other months relative to march. The figure finds evidence of seasonal variation in both the neonatal and post-neonatal periods. Although infectious diseases play a greater role in the post-neonatal period, they are also important for neonates (Bhalotra et al.; Fadel et al., 2017). We find that seasonal variation in both the neonatal and post-neonatal period is similar to an extent. However, compared to march, neonatal mortality is higher in May, and post-neonatal mortality is higher in July and December.

Decline and persistence in seasonal variation

How have seasonal patterns of mortality changed over time? Over the twenty-five year period between 1990 and 2015, infant mortality has nearly halved. Does this decline in infant mortality correspond to a decline in seasonal variation? Figure 2–2 provides evi-

dence from un-adjusted mortality rates for a decline in seasonal variation. Seasonality has declined particularly when comparing the latest NFHS round with earlier NFHS rounds.

Table 2–4 confirms these patterns from age standardized mortality rates for each calendar month by NFHS round. To measure the extent of variation between calendar months, it also shows the standard deviation, the coefficient of variation (standard deviation divided by the mean), and the standard deviation of the natural log of the mortality rates for each round. The coefficient of variation and the standard deviation of log mortality rates are similar measures, and measure relative variation in mortality. In contrast, the standard deviation measures absolute variation. The findings here show that relative and absolute deviation move together. The three measures of variation do not decline between the first three rounds. But they do decline in the fourth NFHS round, confirming descriptive patterns seen in Figure 2–2. In terms of absolute rates, for the first three rounds, standard errors are large, because of smaller sample sizes in the first three rounds. The estimates still reveal that there is substantial variation in mortality in all four rounds, and that seasonal variation can still be seen in the latest NFHS round, for births in the recent period of 2011-2014.

The last panel in Table 2–4 considers how much infant mortality would be lower if the mortality conditions in March were prevalent throughout the year. For this, I estimate the $_1m_0$ per year per 1,000 (estimates A) as annualized infant mortality rates observed in each of the survey rounds. Next, I calculate the $_1m_0$ in a year if the mortality conditions in March were prevalent throughout the year by calculating annualized March $_1m_0$ (estimates B). To do this, I scale observed $_1m_0$ in March by the person-months observed for the whole of the year. Finally, I calculate the difference between estimates A and B to estimate the hypothetical reduction in infant mortality if differences in infant mortality if march mortality conditions were prevalent throughout. This exercise is an upper bound on the likely mortality reductions, because some of the lower mortality in the March months could be because of 'harvesting' of more diseased infants in the earlier winter months (Dorélien,

2019). The calculation still reveals a large role of seasonal variation: a reduction of about 11 deaths per 1,000 in the early 1990s if March rates prevailed throughout the year, and about 4 deaths less per 1,000 in the mid-2010s.

Average mortality decline by calendar-month: A related question is if mortality declined in some months faster than others. Figure 2–5 reports results from an interaction between calendar year (operationalized linearly) and calendar months (operationalized as dummies from January to December). The coefficient for year in Figure 2–5 is the average decline per year in the month of March. Thus, it shows that for the month of March, infant mortality declined at the rate of 0.1 deaths per 1,000 infants per month every year. The interaction coefficients, for calendar month by year, denote the additional decline in a particular month above the 0.1 deaths per 1,000 per month decline observed for March. These estimates show that mortality has declined for all months. Additionally, they suggest that mortality in the monsoon months and in the early winter months has declined slightly faster than the mortality in winter and summer months. Compared to March and Summer mortality declines, mortality in the monsoon months declined by an additional 0.05 deaths per 1,000 per year. However, these interaction estimates are not statistically different from zero, and the hypothesis that mortality declined similarly across calendar months cannot be ruled out.

Differences by measures of socioeconomic disadvantage

Are these seasonal differences in infant mortality patterned by measures of socioeconomic status? Figure 2–6 plots regression coefficients from separate regressions by residence, sex of the child, mother's education, and by household asset ownership. The reference month is March, and the sample is limited in these regressions to the observations belonging to a particular group.

Rural residence: Panel a) compares infant mortality in other calendar months in rural areas with infant mortality in March in rural areas. It also compares mortality in a particular calendar-month in urban areas to mortality in March in urban areas. Overall, ru-

ral areas have higher seasonality than urban areas. In no calendar month are differences in urban areas statistically different from March. In rural areas, several months have higher mortality. Figure 2A–2 examines seasonality in infant mortality by round and rural residence. This figure suggests that the seasonal variation in the overall urban patterns largely comes from the earlier NFHS rounds. By NFHS-4, there is little evidence of seasonal variation in urban areas. In rural areas, however mortality in the winter and summer months is markedly higher than mortality in March. Even at a low level of overall infant mortality, differences between months in NFHS-4 can be as high as one death per 1,000 per month in rural areas.

Sex of the child: Panel b) shows differences by the sex of the child. Males are more vulnerable to disease at all ages (Austad, 2011; Drevenstedt et al., 2008). However, in India, discrimination against female children leads to higher mortality among them (Murthi et al., 1995), particularly in the ages 1-4 years. In ages 0-1 years, male and female $_1m_0$ are roughly equal (Sawyer, 2012; Rai et al., 2017). Prior research has suggested that this in itself is a reflection of gender bias, given that female mortality rates should be lower than male mortality rates (Kashyap and Behrman, 2020). Previous evidence has documented that in the presence of seasonal shocks, intra-household allocations favor boys compared to girls (Behrman, 1988a). To the extent that a worse disease environment affects boys more, we expect boys to have higher mortality rates in the calendar months where mortality is higher. Gender inequality within household in the presence of seasonal variation will predict equal mortality by sex, or higher mortality among girls. Figure 2–6 finds preliminary evidence that both factors are at play: in some months, such as July, both boys and girls have equally higher mortality. However, in other months, such as in November, boys have slightly higher mortality than girls. Because discrimination against female children and the causes of death are not directly observed here, we cannot say what the likely patterns by gender would be without discrimination. This is an crucial avenue for future research.

Maternal education: Panel c) shows mortality rates by mother's education. I use

three categories: no education, 1-8 years of education, and eight or more years of education, which corresponds to the number of years of education which are a constitutional right in India (Bhatty, 2014), and is a corresponding Sustainable Development Goal as well (United Nations General Assembly, 2015). A large literature, in the demographic and health sciences considers mother's education as a salient determinant of child health (Caldwell, 1994). The regression results suggest that although seasonal variation in mortality can be seen for all educational categories, seasonal variation is particularly salient among the children of least educated mothers. Additionally, the fact that there is seasonal variation even in children of mothers who have more than 8 years of education suggest that these patterns are not driven by measurement error. Existing evidence suggests that mothers with more education are more likely to recall their birth months correctly than mothers with less education (Manesh et al., 2008; Larsen et al., 2019), and even among these mothers, there is evidence of seasonal variation in mortality.

Household wealth: Panel d) finds that the bulk of the seasonal variation is in families that are poorer. Families that own either a motorcycle, car, or refrigerator have no seasonal variation in infant mortality. In the monsoon months, these families have lower mortality than in March, although differences are not statistically different from zero. Appendix Figure 2A–3 shows that patterns by asset ownership are consistent across the earlier and latest NFHS rounds.

Regional and climactic patterns

Figure 2–7 shows coefficients and 95% confidence intervals for infant mortality by calendar month from separate regressions for each region. The figure shows that seasonality in infant mortality is lowest in the southern states, and highest in the poorer central states of India. The southern states have lower mortality rates (Drèze and Sen, 2013) as well as a better early life disease environment (Million Death Study Collaborators and others, 2010), partly because they have better sanitation and lower solid fuel use (Government of India, 2011a). The southern states also have lesser variation across seasons, compared to north

India, which may be driving these patterns. In the eastern states, mortality is highest in the winter month of November, while in the northern states, mortality is highest in the winter month of December and the summer month of June.

Figure 2–8 shows seasonal variation in mortality by Koppen-Geiger climate regimes, using data from the fourth round of the NFHS. In this round, we observe higher mortality in the tropical monsoon regions (areas of western ghats in India) in the month of August, when the monsoon is at its peak. Mortality is higher in April, May, and June in the Humid Subtropical region, which corresponds to the northern Indian plains. Although seasonal variation is seen in other regions as well, such as with higher mortality in June in the arid and desert regions, these patterns are less pronounced.

2.5 Robustness

Appendix figure 2A–4 tests if the estimates in Table 2–3 Model 4 are robust to alternative specifications for time-trends (panel a) or age (panel b). In panel a), I consider 4 alternative ways to control for time trends: survey rounds, linear month years, linear year, and dummies for years. The regressions include all existing controls: for age dummies, sex of the child, state, and rural residence. Although the estimated coefficients and confidence intervals vary slightly, they are broadly similar for all regressions. Panel b) confirms patterns already observed in raw and age-standardized estimates (in figure 2–3). It also shows that linear age controls are similar to no controls for age, while controlling for age and a squared term for age is similar to dummies for age in months.

Appendix figure 2A–6 examines seasonal variation by access to household environmental technologies. The regression controls for household asset ownership. The figure suggests higher seasonal variation among households who do not use toilets or clean fuels for cooking. This is consistent with findings from the prior literature (Geruso and Spears, 2018b; Bassani et al., 2010), which would expect greater exposure to infectious disease among households without access to environmental health technologies.

Prior research has documented a large and implausible gap in height-for-age for children born in December in a year and children born a month later in January the next year (Agarwal et al., 2017; Larsen et al., 2019). Are measures of infant mortality by calendar month similarly biased? Appendix Figure 2A–5 reports differences in differences in period infant mortality between December and January next month. We find that there are no systematic biases in period mortality rates. Period mortality rates for December are sometimes equal to period mortality rates in January next month, sometimes higher, and sometimes lower. This suggests that unlike birth-month effects on height for age, period mortality rates by calendar month are not systematically biased. This can be inferred from figure 2–3 as well, which shows that unadjusted and age-standardized mortality rates for January are similar. Thus, it confirms that January's mortality rates are not higher because it has a higher proportion of neonates than other months.

Finally, I consider the effects of age-misreporting and heaping. Panel a) of appendix figure 2A–7 shows the number of deaths by month of death. Most deaths are in the first month, confirming patterns of high neonatal mortality in India (Sankar et al., 2016). The panel shows evidence of heaping around 6th months, when the number of deaths are higher than mortality in the 5th month. To test if this age-heaping biases mortality rates, panel b) in appendix figure 2A–7 compares $_1m_0$ with $_1m_0$ that exclude 6 month olds. The $_1m_0$ estimates which exclude 6-month olds exhibit slightly higher seasonality than $_1m_0$, suggesting that this age-misreporting biases estimates for calendar-months towards zero. Thus, if age-misreporting was lower, it is likely that estimates of seasonal variation would be even higher.

2.6 Discussion

This paper contributes evidence to the literature on environmental determinants of child health for the important case of India, home to every sixth birth and every fifth infant death in the world, in several ways. To my knowledge, this is the first paper to estimate

period infant mortality rates by calendar month for India. The methodological approach outlined here, of estimating period mortality rates between ages 0 and 1 could for periods shorter than a year, using routinely collected survey data, can be used to study other time-varying determinants of child health or policy interventions. The approach estimates unadjusted and age-standardized mortality rates below the age of 1 for calendar months. Age-standardized mortality rates by calendar month in ages below 1 adjust for compositional differences created by seasonal variation in births (Knodel, 1983). The approach here is compatible with several assumptions about the exact occurrence of births and deaths. Although the development of alternative estimation strategies, such as those that assign death dates to replicate age patterns of mortality in the infant ages, would be valuable, the approach used here is robust to alternative estimation strategies as well, such as using randomly assigned birth-dates within the birth-month.

We cannot rule out that some of the period-mortality estimates we observe could be because of lagged effects of disease environment prevailing in previous months. Another concern about the methodological approach followed here is that measurement error in calendar birth month and age at death may still bias estimates of mortality by calendar month. We are able to address these concerns to some extent. First, the paper documents seasonal variation in both the neonatal and post-neonatal period. This implies that overall seasonal variation is not just because of neonatal mortality in misreported birth months. Second, we show that unadjusted and age-standardized mortality rates are similar in January, a month which often gets a disproportionate number of misreported births (Larsen et al., 2019). Similarly, we do not find a systematic difference between December and next-year January in terms of infant mortality. Finally, although more deaths are reported at age six months than deaths at five or seven months, dropping six-month old children from the sample leads to similar results.

Substantively, the analysis presented here concludes that although seasonal variation in mortality has declined over the period 1989-2014. In particular, seasonal variation

was considerably less in the recent period, 2011-14. Similar patterns have been observed in the context of the epidemiological transition in contemporary high and low-income countries as well (Rau, 2006a; Schlüter et al., 2020). In India, the period 2005-2014 saw declines in infant mortality, progress on other aspects of human development (Drèze et al., 2021), and increases in child health provision (Singh et al., 2019). Household ownership of basic assets, such as cleaner fuels, pucca homes, fans and coolers also increased to an extent. It is likely that these improvements contributed to declines in seasonal variation in infant mortality. An important priority for future research is to investigate the causes of declines in infant mortality and seasonal variation in India.

Despite these declines, the paper documents that seasonal variation remains a challenge, particularly in rural areas, and particularly among disadvantages sections of Indian society. Why this is the case, and what can be done about this is an urgent question for policy and research. The pattern of seasonal variation documented in this paper reveals a limited period within the year – the spring months of February and March – where threats to child health are less salient. This evidence has relevance for studies of the late-life-legacy of early life conditions as well as for studies assessing the association between environmental exposures and child mortality. It is consistent with the prior literature in India. This literature has found a large number of seasonally varying exposures – pollution, heat, humidity, and cold weather, and the monsoon (Spears, 2019b; Fu et al., 2018; Brown et al., 2019). The results documented here are also consistent with existing studies using cause of death data, such as Farrar et al. (2019). Farrar et al. (2019) find evidence in support of variation in diarrhea and pneumonia related mortality in the post-neonatal period in the summer and the winter months.

On the whole, we find that given current technologies and incomes, infant mortality in March is lower by about 4 infant deaths per 1000 per year. This is a sizable compared to overall infant mortality in India and elsewhere. Further research, which examines the extent to which accounting for these seasonally varying exposures can explain sea-

sonality in mortality would help further bolster the evidence on the relative importance of the mechanisms highlighted here. Seasonal variation in nutrition availability, agricultural work, and other economic determinants are also not directly examined here. Further studies examining seasonal variation in cause-specific mortality and patterns of social disadvantage, such as gender, can be helpful in understanding discrimination within the household. The NFHS surveys may not be powered to detect seasonal variation over finer geographic and time scales, even if seasons themselves vary within India in terms of timing and intensity. This is an important avenue for future research. These specific studies can contribute to overall evidence on the role of policy approaches to reduce seasonal variation.

Figure 2–1: Lexis-Diagram containing life-lines and deaths to illustrate how $_1m_0$ by month is calculated for a particular month (January 1992)
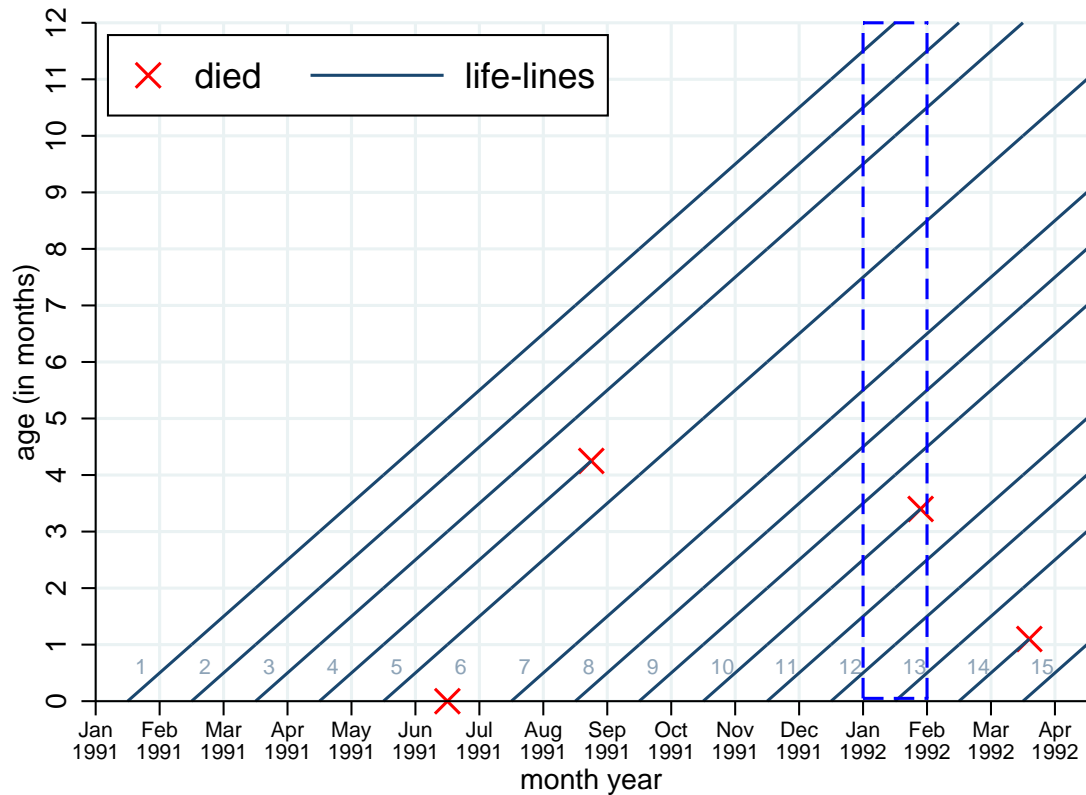
I estimate period-specific infant mortality rates, where each period is a calendar month-year. For example, to estimate the infant mortality rate for January 1992, the numerator consists of all deaths that occur to infants less than 1 year old in January 1992. The denominator consists of children below age 1 who are alive on January 15, 1992. The Lexis diagram in Figure 1 shows life-lines and deaths for 15 children, from birth to age 1 year. Of these 15, 4 died before reaching the age of 1 year. For the period January 1992, marked by blue dashed lines, there was 1 death (child id 10), and 10 children were exposed to the risk of death. The mortality rate for children under the age of 1 year for January 1992 in this hypothetical example would thus be 100 deaths per 1,000.

Figure 2–2: Estimated of infant mortality ($_1m_0$) by calendar month-year

infant deaths per 1,000

month-year
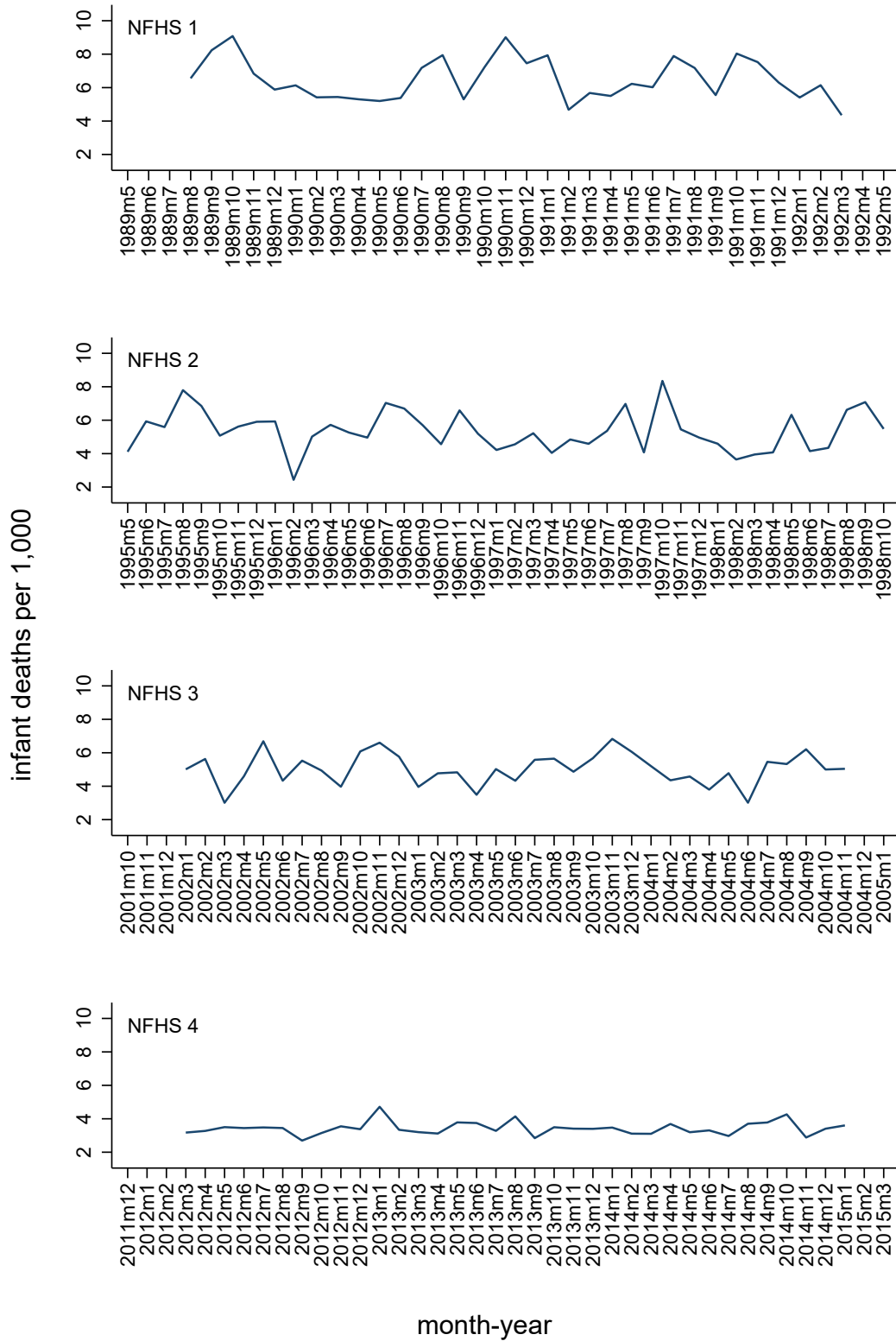
Note: Estimates account for survey weights.

Figure 2–3: Estimated raw and age-standardized $_1m_0$ with 95% confidence intervals by calendar month
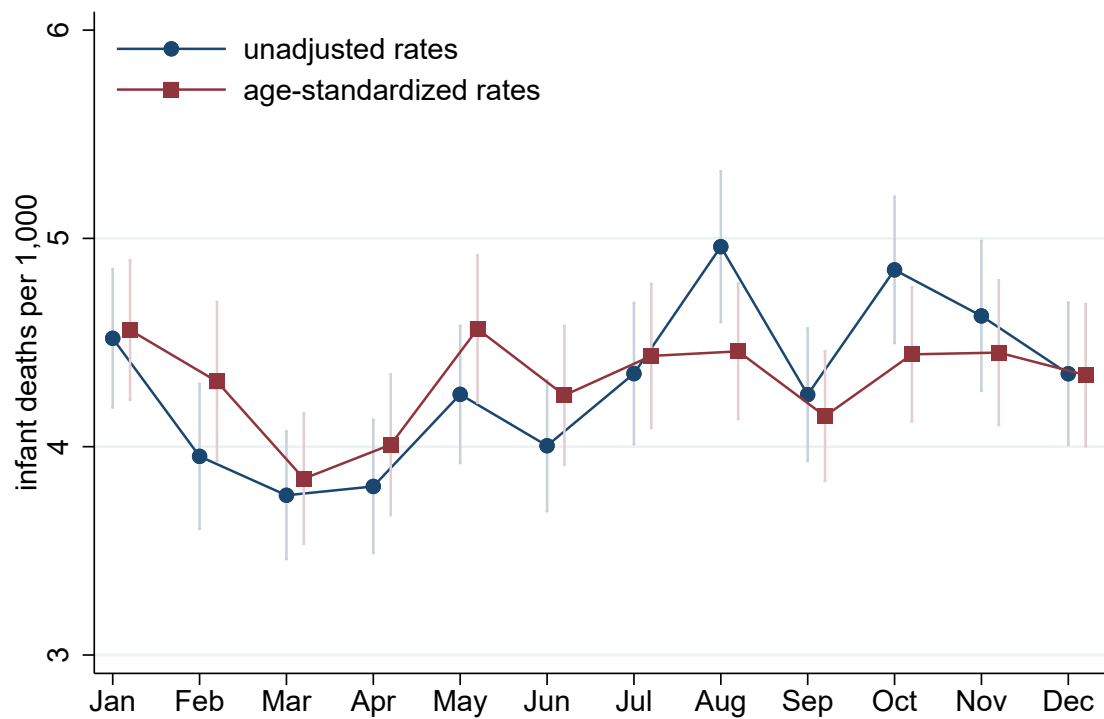
Note: Mean and 95% CI estimates account for survey weights and clustering of observations within the primary sampling unit. Age-standardized estimates match the distribution of age (in monhts) for each calendar month with the overall age distrbution. Observations are from pooled sample, for child-months from all four NFHS rounds.

Figure 2–4: Seasonal variation in neonatal and post-neonatal mortality
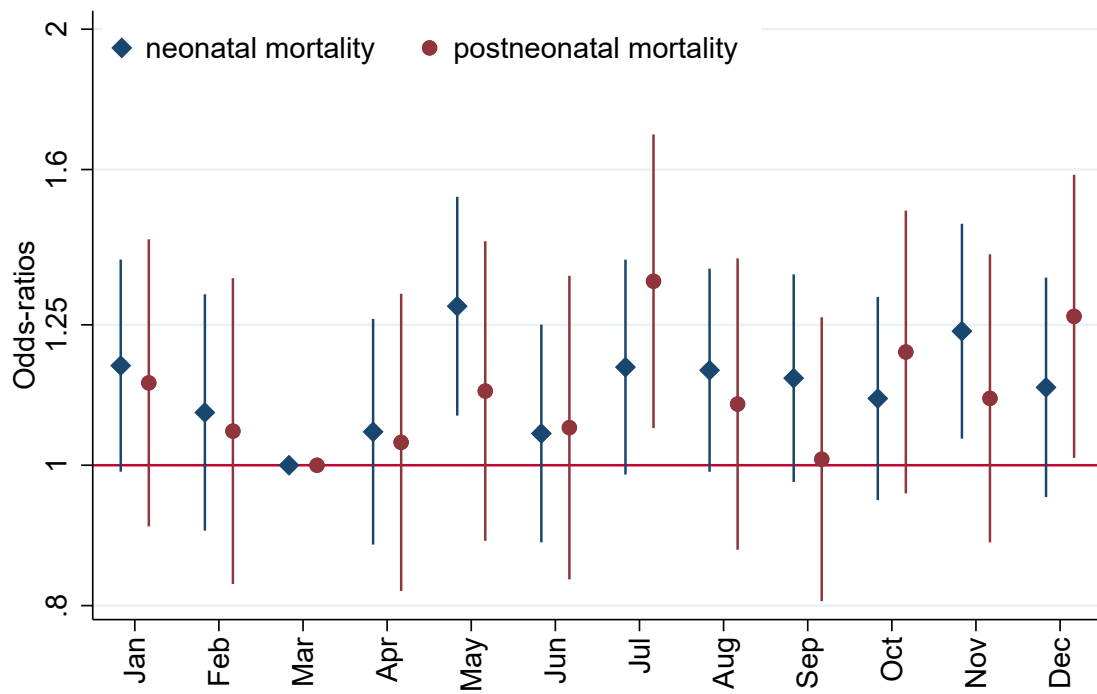
Regressions control for dummies of sex and age in months of the child,
survey round, interview month, rural residence, and state.
Estimates are for the pooled sample.

Figure 2–5: Plotted coefficients from linear probability regression interacting linear year and calendar month to examine if decline in infant mortality is higher for some calendar months
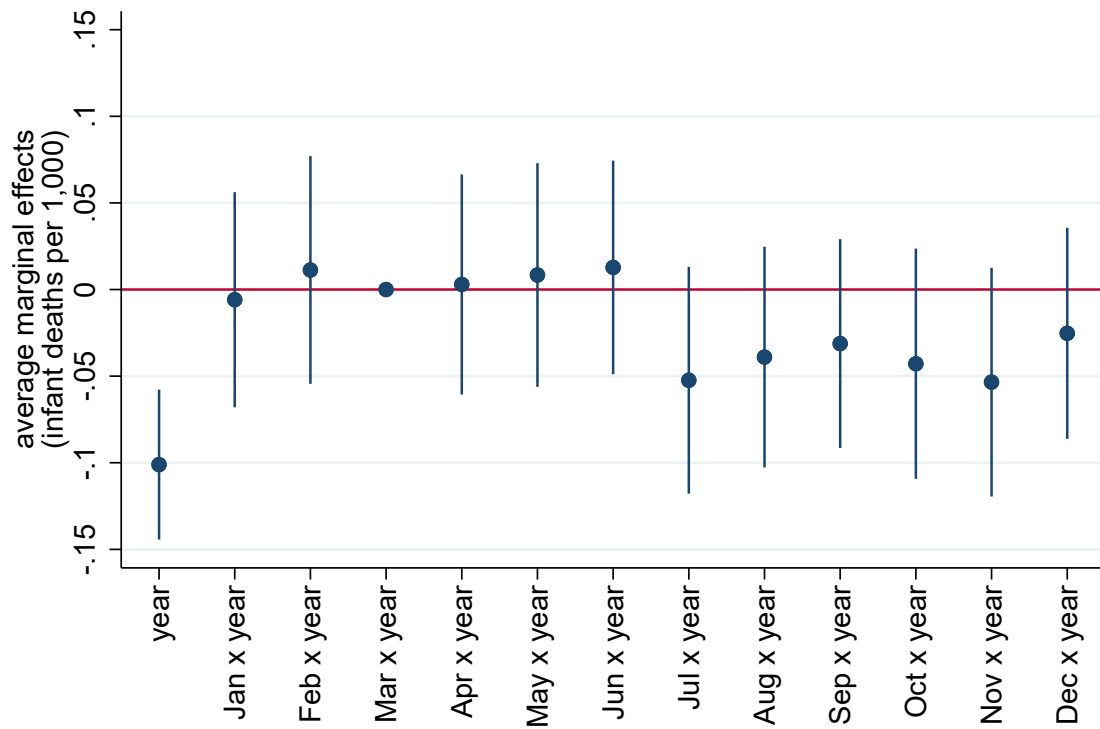
Model includes dummies for sex and age in months of the child, state, rural residence, interview month, and round. Estimates account for survey wights and clustering within the primary sampling unit.

Figure 2–6: Plotted coefficients from separate linear probability regressions for $_1m_0$ by calendar month, stratified by characteristics of children
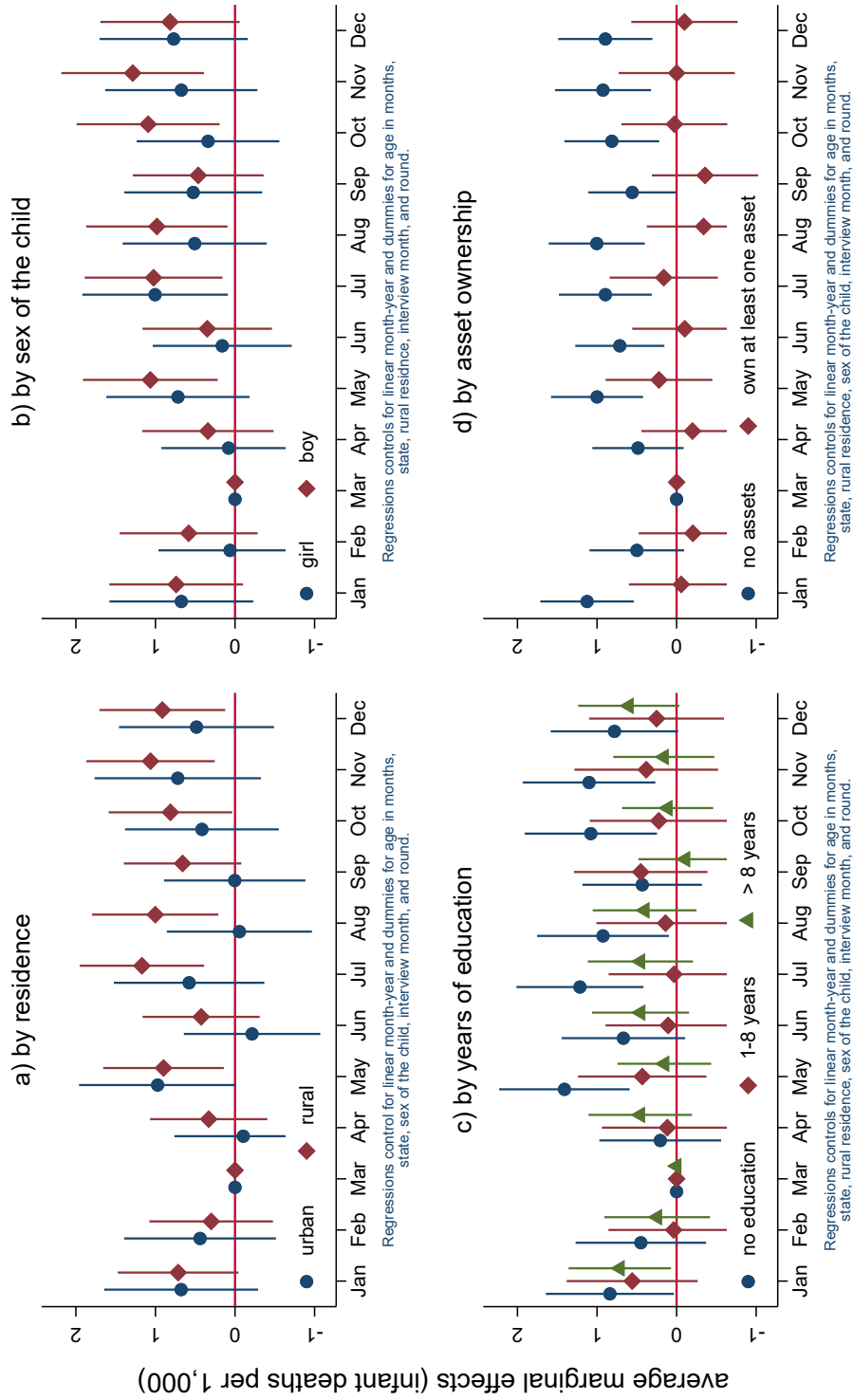
Figure 2–7: Plotted coefficients from separate linear probability regressions for $_1m_0$ by calendar month, stratified by region
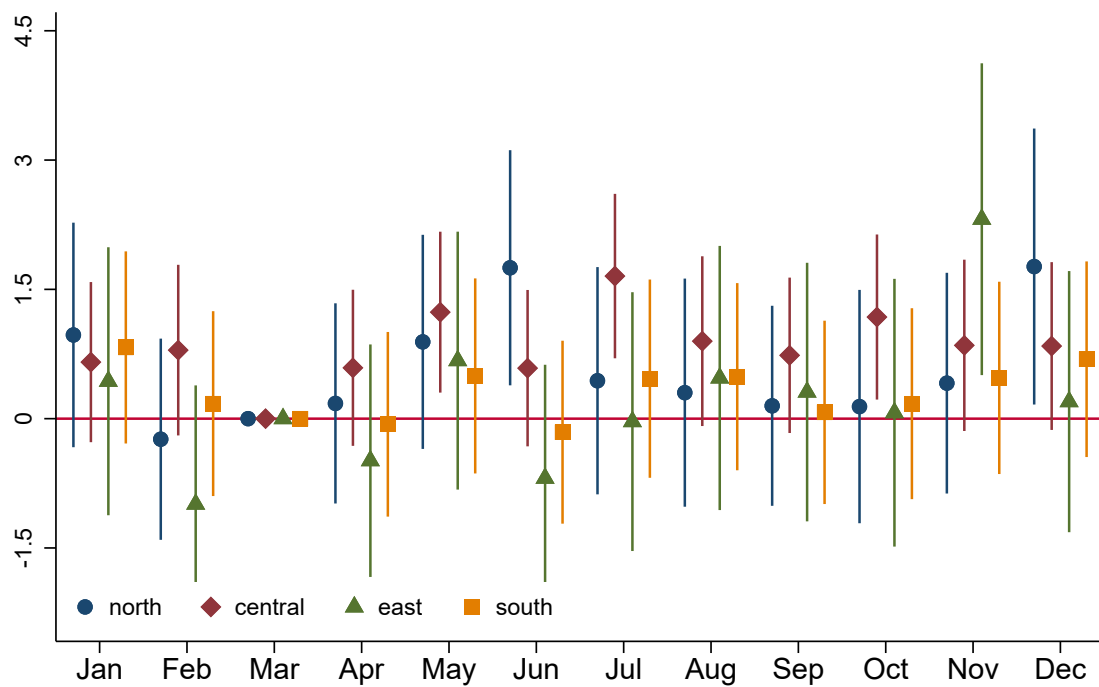
North: Jammu & Kashmir, Himachal Pradesh, Haryana, Punjab, Delhi.
Central: Uttarakhand, Uttar Pradesh, Rajasthan, Madhya Pradesh, Chattisgarh, Gujarat, Bihar, Jharkhand.
East: Assam, Odisha, West Bengal, Sikkim, Manipur, Meghalaya, Arunachal Pradesh, Tripura, Nagaland, Mizoram.
South: Andhra Pradesh, Telangana, Maharashtra, Goa, Tamil Nadu, Kerala, Karnataka.
Estimates are for the pooled sample.

Figure 2–8: Seasonal variation in $_1m_0$ by calendar month and Koppen-Geiger climate classifications, NFHS-4 (2015-16) for period 2011-2016
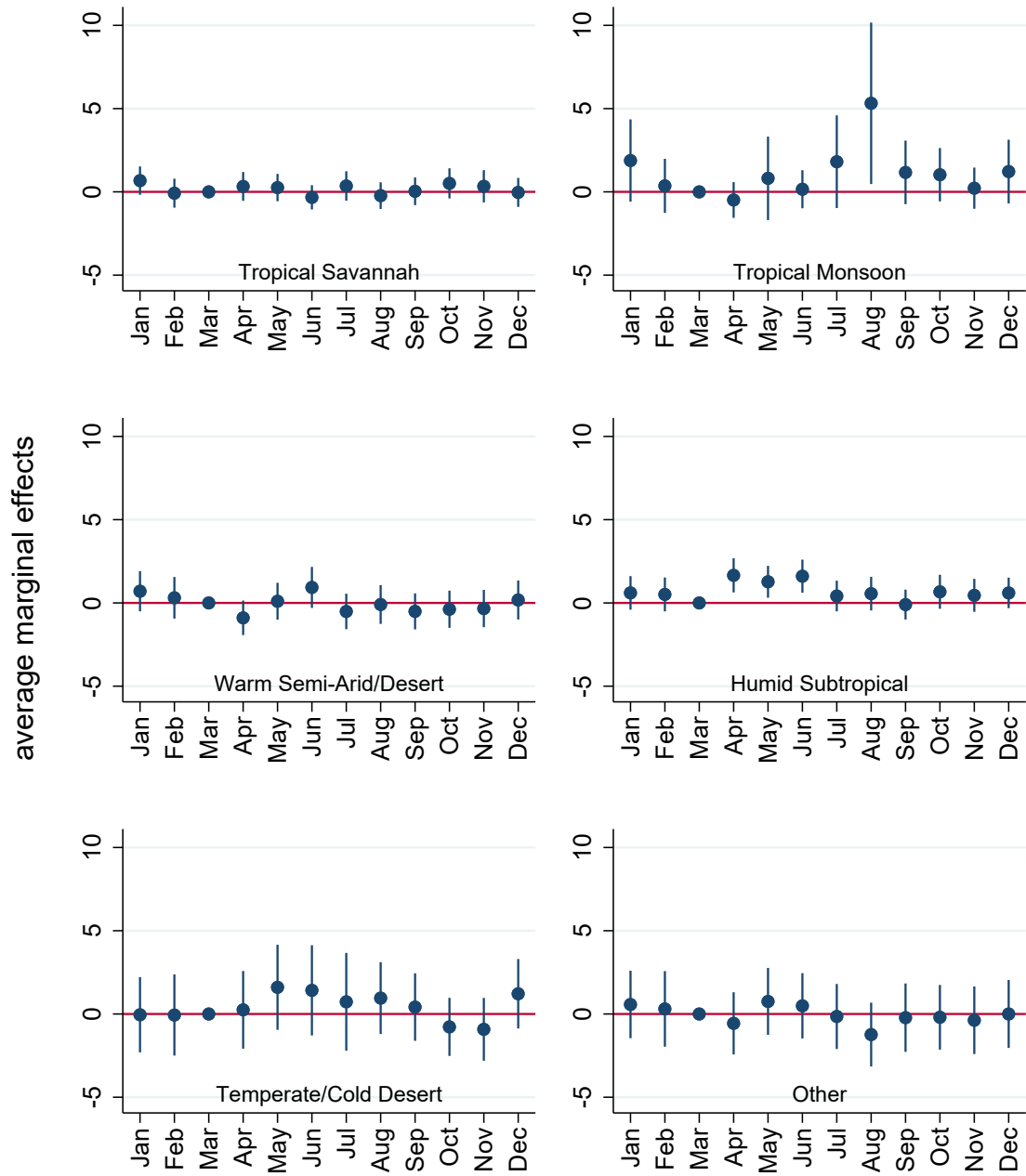
Regressions control for dummies of sex and age in months of the child,
survey round, interview month, and rural residence.

Table 2–1: Summary statistics (means or proportions) for the children in the sample

	NFHS-1 (1992-93)		NFHS-2 (1998-99)		NFHS-3 (2005-06)		NFHS-4 (2015-16)		Overall	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
Months observed	8.5	0.023	8.9	0.021	8.8	0.025	8.9	0.012	8.8	0.011
Died	0.07	0.002	0.06	0.001	0.06	0.002	0.04	0.001	0.06	0.001
Female	0.49	0.003	0.49	0.003	0.48	0.004	0.48	0.002	0.48	0.001
Rural	0.77	0.009	0.77	0.009	0.74	0.010	0.71	0.005	0.75	0.004
Region										
North	0.07	0.004	0.06	0.004	0.06	0.004	0.07	0.002	0.06	0.002
Central	0.50	0.013	0.52	0.012	0.56	0.012	0.53	0.005	0.53	0.005
East	0.16	0.008	0.14	0.008	0.15	0.008	0.13	0.003	0.14	0.004
South	0.28	0.011	0.28	0.011	0.24	0.010	0.27	0.005	0.27	0.005
Mother's education										
No education	0.65	0.007	0.57	0.007	0.51	0.008	0.30	0.003	0.51	0.003
1-8 years	0.22	0.005	0.26	0.004	0.28	0.005	0.31	0.002	0.27	0.002
> 8 years	0.13	0.004	0.17	0.004	0.21	0.005	0.39	0.003	0.22	0.002
Owns > 0 asset	0.10	0.003	0.14	0.004	0.21	0.005	0.45	0.006	0.22	0.002
Period of observation	1989m8-1992m3		1995m5- 1998m10		2002m1-2004m11		2012m3-2015m1		331,845	
Children observed	42,559		49,595		39,416		200,275		331,845	

Means and standard errors account for national women's weights and clustering of observations within primary sampling units. Units of observation in this table are children.

Table 2–2: Proportionate distribution of observed child months by calendar month in the sample, and mean observed infant deaths per child-month in each calendar month

calendar month	child-months		deaths	
	mean	s.e.	mean	s.e.
January	0.085	0.0001	0.005	0.0002
February	0.077	0.0002	0.004	0.0002
March	0.085	0.0001	0.004	0.0002
April	0.076	0.0001	0.004	0.0002
May	0.082	0.0002	0.005	0.0002
June	0.082	0.0002	0.004	0.0002
July	0.082	0.0002	0.005	0.0002
August	0.090	0.0001	0.006	0.0002
September	0.090	0.0001	0.005	0.0002
October	0.090	0.0001	0.006	0.0002
November	0.084	0.0001	0.006	0.0003
December	0.077	0.0002	0.005	0.0003
n (observations)	2,933,892		12,479	

Source: NFHS Surveys. Note: Means and standard errors account for national women's weights and clustering of observations within primary sampling units. Units of observation in this table are child-months.

Table 2–3: Estimated coefficients and odds-ratios from linear probability, logistic, and complementary-log-log models of infant mortality by calendar month

model:	linear probability				logit	cloglog
	(1)	(2)	(3)	(4)	(5)	(6)
reference: March						
January	0.764*	0.686*	0.686*	0.705*	1.157*	1.155*
	(0.319)	(0.315)	(0.315)	(0.315)	(0.081)	(0.079)
February	0.176	0.469	0.460	0.323	1.075	1.074
	(0.323)	(0.321)	(0.321)	(0.321)	(0.081)	(0.080)
April	-0.053	0.074	0.079	0.213	1.048	1.048
	(0.305)	(0.302)	(0.302)	(0.302)	(0.076)	(0.075)
May	0.623+	0.802*	0.801*	0.896**	1.229**	1.226**
	(0.319)	(0.315)	(0.315)	(0.316)	(0.087)	(0.086)
June	0.144	0.150	0.148	0.257	1.051	1.050
	(0.304)	(0.301)	(0.301)	(0.302)	(0.075)	(0.074)
July	1.035**	0.897**	0.895**	1.018**	1.219**	1.215**
	(0.326)	(0.321)	(0.321)	(0.322)	(0.085)	(0.083)
August	1.595***	0.757*	0.756*	0.753*	1.137+	1.135+
	(0.330)	(0.325)	(0.325)	(0.325)	(0.076)	(0.075)
September	0.862**	0.476	0.475	0.488	1.097	1.097
	(0.309)	(0.304)	(0.304)	(0.304)	(0.073)	(0.072)
October	1.566***	0.702*	0.702*	0.727*	1.130+	1.128+
	(0.327)	(0.320)	(0.320)	(0.322)	(0.075)	(0.074)
November	1.534***	0.948**	0.951**	0.987**	1.189*	1.187*
	(0.341)	(0.336)	(0.335)	(0.336)	(0.082)	(0.081)
December	0.986**	0.758*	0.764*	0.789*	1.177*	1.174*
	(0.330)	(0.326)	(0.326)	(0.327)	(0.083)	(0.081)
age (in month) dummies		X	X	X	X	X
sex of the child		X	X	X	X	X
state			X	X	X	X
rural residence			X	X	X	X
month of survey				X	X	X
survey round dummies				X	X	X
linear month year				X	X	X
mean imr	4.32 infant deaths per 1,000 per month					
children	331,845	331,845	331,845	331,845	331,845	331,845
n (child-months)	2,933,892	2,933,892	2,933,892	2,933,892	2,933,892	2,933,892

Source: NFHS Surveys. Estimates and clustered standard errors account for national women's weights and clustering of observations within primary sampling units. $+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

Table 2–4: Age-standardized $_1m_0$ by calendar month and survey round

	NFHS-1		NFHS-2		NFHS-3		NFHS-4	
	$_1m_0$	s.e.	$_1m_0$	s.e.	$_1m_0$	s.e.	$_1m_0$	s.e.
January	6.6	0.551	5.3	0.528	4.9	0.548	3.9	0.196
February	6.4	0.603	4.1	0.455	5.4	0.576	3.4	0.232
March	5.6	0.516	5.0	0.489	4.3	0.543	3.1	0.179
April	5.9	0.662	5.2	0.521	4.3	0.499	3.4	0.202
May	6.3	0.750	5.6	0.471	6.0	0.605	3.7	0.203
June	6.0	0.652	5.0	0.411	4.1	0.528	3.8	0.205
July	7.3	0.701	5.7	0.452	5.8	0.610	3.3	0.198
August	6.6	0.521	6.1	0.425	4.7	0.531	3.4	0.199
September	6.3	0.532	5.6	0.422	4.9	0.524	3.1	0.180
October	7.0	0.550	5.2	0.384	5.0	0.503	3.5	0.199
November	7.2	0.566	5.4	0.491	5.5	0.555	3.3	0.203
December	6.8	0.559	5.1	0.492	5.8	0.756	3.4	0.191
s.d. ($_1m_0$)	0.535		0.484		0.636		0.237	
c.v. ($_1m_0$)	0.082		0.092		0.127		0.069	
s.d. ($log_{_1m_0}$)	0.083		0.096		0.126		0.069	
A. Observed $_1m_0$, annualized	78.0		63.9		60.3		41.2	
B. March $_1m_0$, annualized	66.6		59.7		51.8		37.6	
C. Difference (A - B)	11.4		4.2		8.5		3.7	

Means and standard errors account for national women's weights and clustering of observations within primary sampling units. Units of observation in this table are child-months. s.d: standard deviation. c.v.: coefficient of variation.

2.7 Supplementary Appendix

Additional methodological notes

The approach used in this paper and outlined in the lexis diagram is compatible with several assumptions. The full list of compatible assumptions includes:

- All births are in the middle of the month (on the 15th); and all deaths are on the exact day the child attains the age (in months) at which they died
- All births are in the middle of the month (on the 15th); and all deaths are within the fifteen days of reaching a particular age (in months)
- All births are in the middle of the month (on the 15th); and all deaths are on the last day of the month, i.e. at age in months the child died + 0.49 of a month
- All births are distributed uniformly within the month; and all deaths are on the exact day the child attains the age (in months) at which they died

The approach is also robust to alternative assumptions. For instance, we can assume that within a calendar-month, births are randomly distributed across days. Then, for neonatal deaths, we know the exact month a child died because age at death in days is available. For deaths at ages greater than age zero month, we can assume that the death was fifteen days after the day a child turned a particular age-month. So a child who died at age two months can be assumed to have died 74 days after the day s/he was born. Appendix figure 2A–8 shows the unadjusted and adjusted mortality rates below age 1 for this approach (red squares). Comparing with the mortality rates using the main approach outlined in the paper (blue circles) reveals minor differences. It is worth noting that the risk mortality declines rapidly within the first year of life in India (Hill, 1995; Guillot et al., 2012). Hence, deaths are more likely to be close to the day a child turns a particular month than towards the middle or the end of that age-month. Hence, the blue circles are more closer to reality than the red squares in terms deaths, even if the red squares are a closer approximation to reality than the blue circles in terms of births.

Figure 2A–1: Seasonality in births
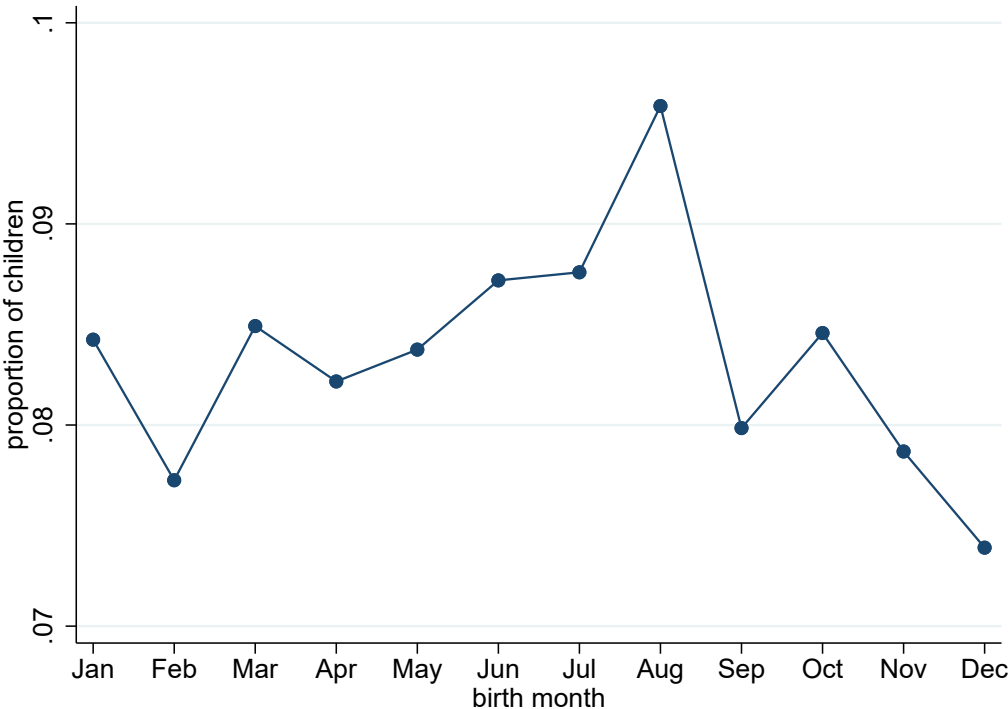
Source: NFHS Surveys. Estimates are from pooled sample.

Figure 2A–2: Plotted coefficients from separate linear probability regressions, showing $_1m_0$ by calendar month and residence, for earlier and latest NFHS rounds

NFHS 1, 2, & 3

Regressions control for linear month-year and dummies for age in months, state, sex of the child, interview month, and round.

NFHS 4

Regressions control for linear month-year and dummies for age in months, state, sex of the child, and interview month.
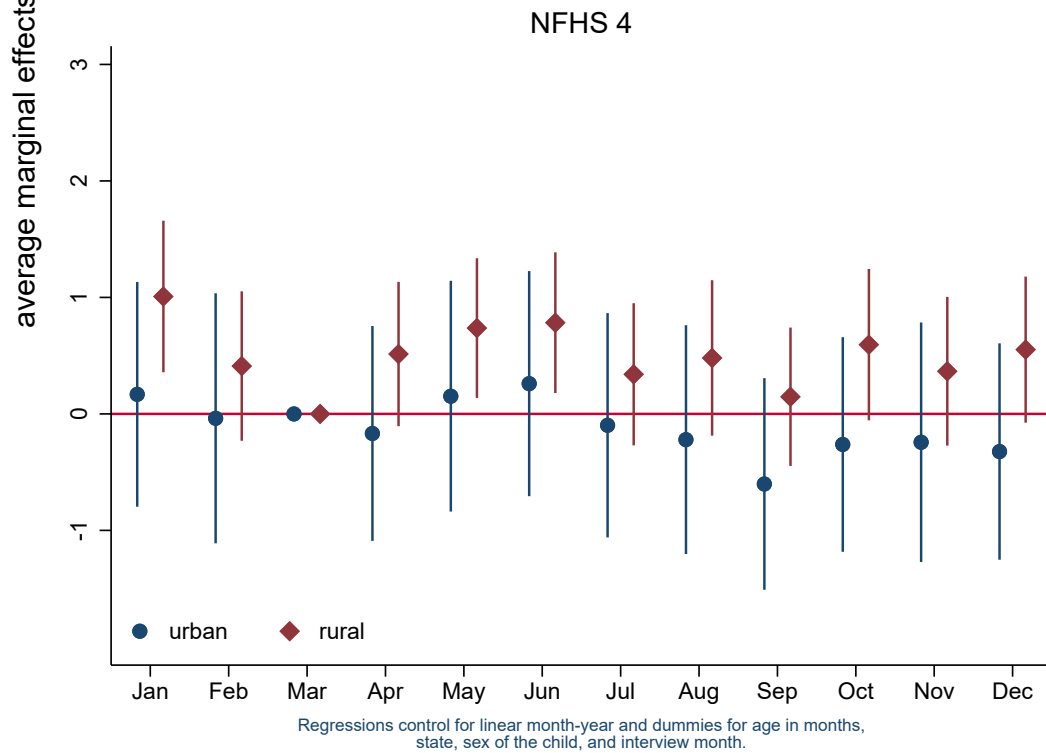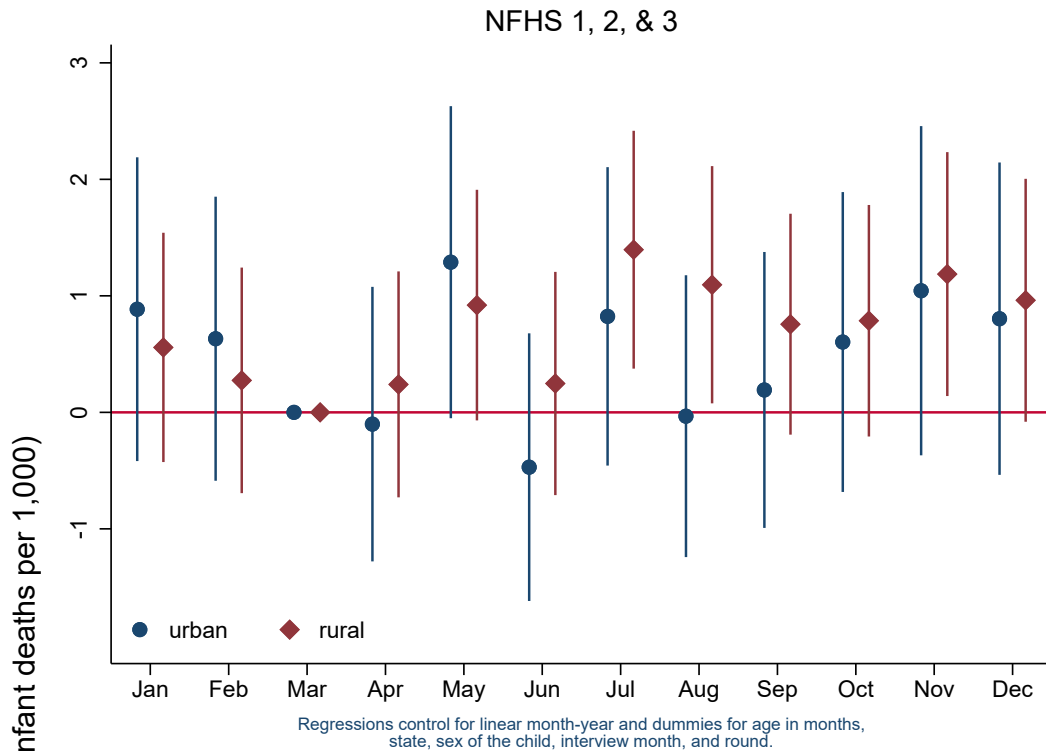
All estimates account for survey wights and clustering within the primary sampling unit.

Figure 2A–3: Seasonal variation by asset ownership, for earlier and latest NFHS rounds

Round 1, 2, and 3

no assets   own at least one asset

Round 4

no assets   own at least one asset
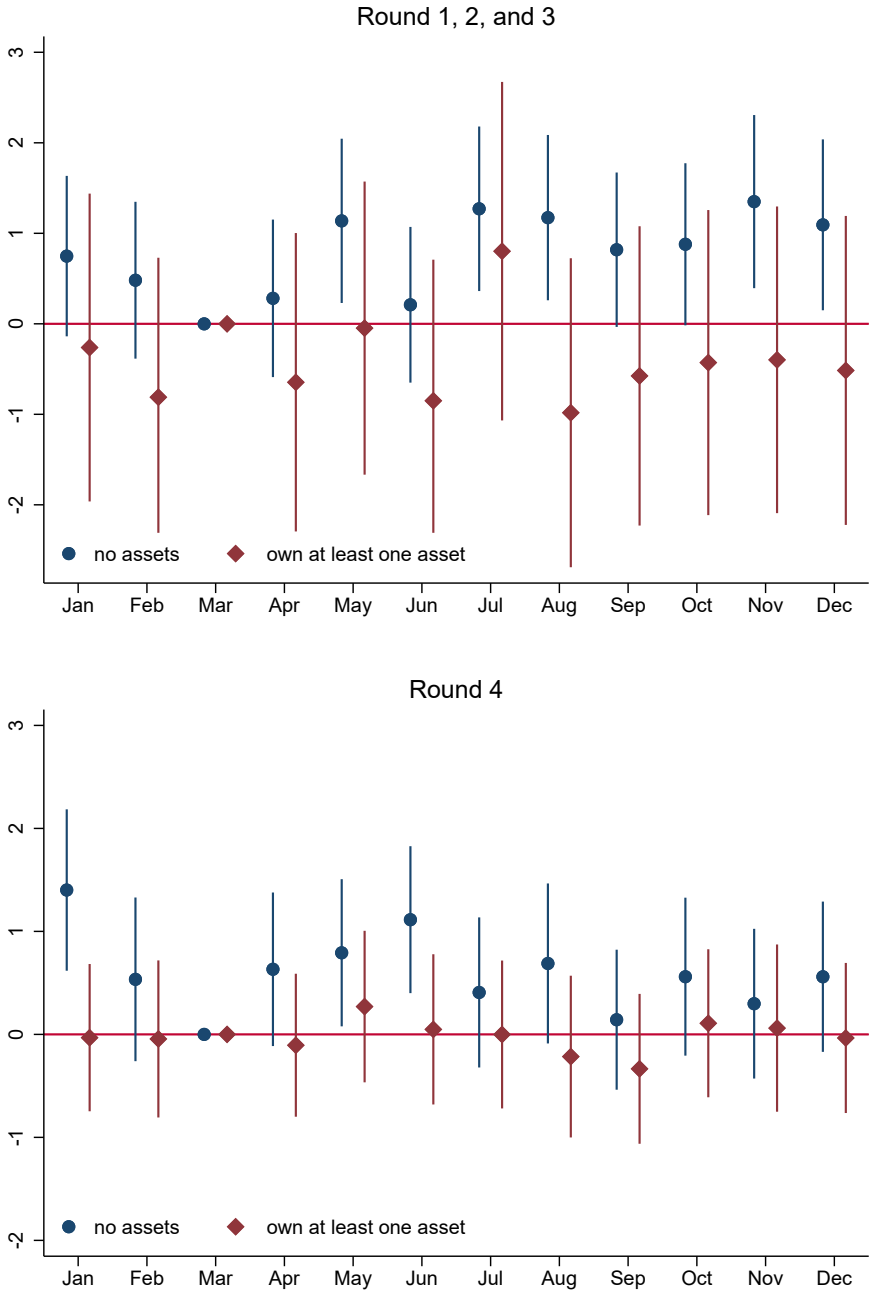
Regressions controls for linear month-year and dummies for age in months,
state, rural residence, sex of the child, interview month, and round.

Figure 2A–4: Results from alternative regression specifications to control for time and age

a) alternative specifications for time controls

All models include dummies for age in months and sex of the child, state, rural residence, interview month, and survey round.

b) alternative specifications for age controls

All models include linear month-year and dummies for sex of the child, state, rural residence, interview month, and survey round.
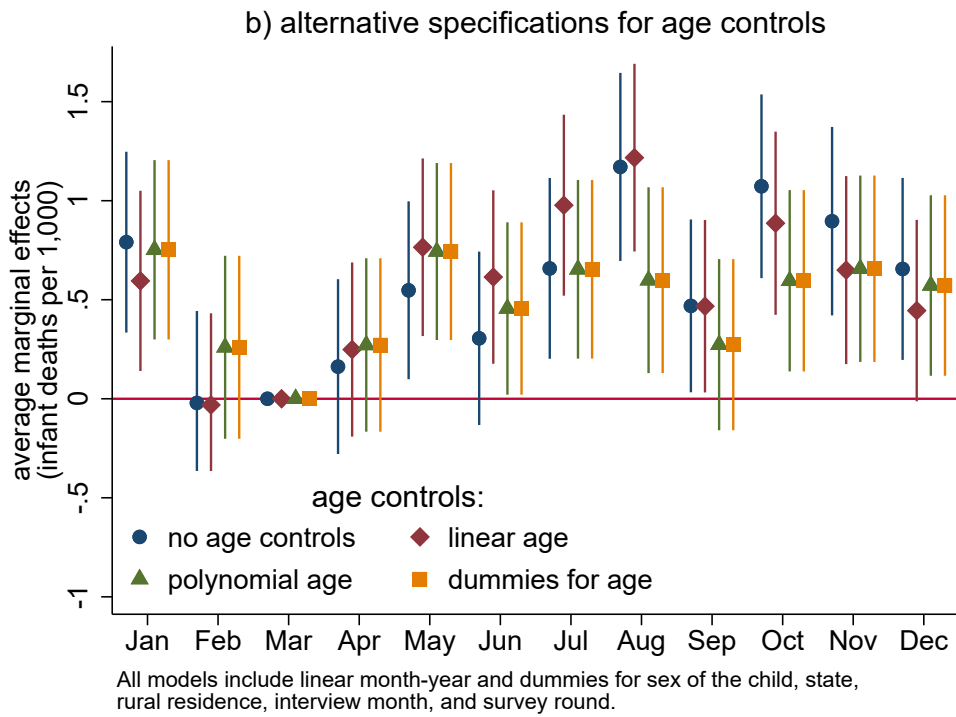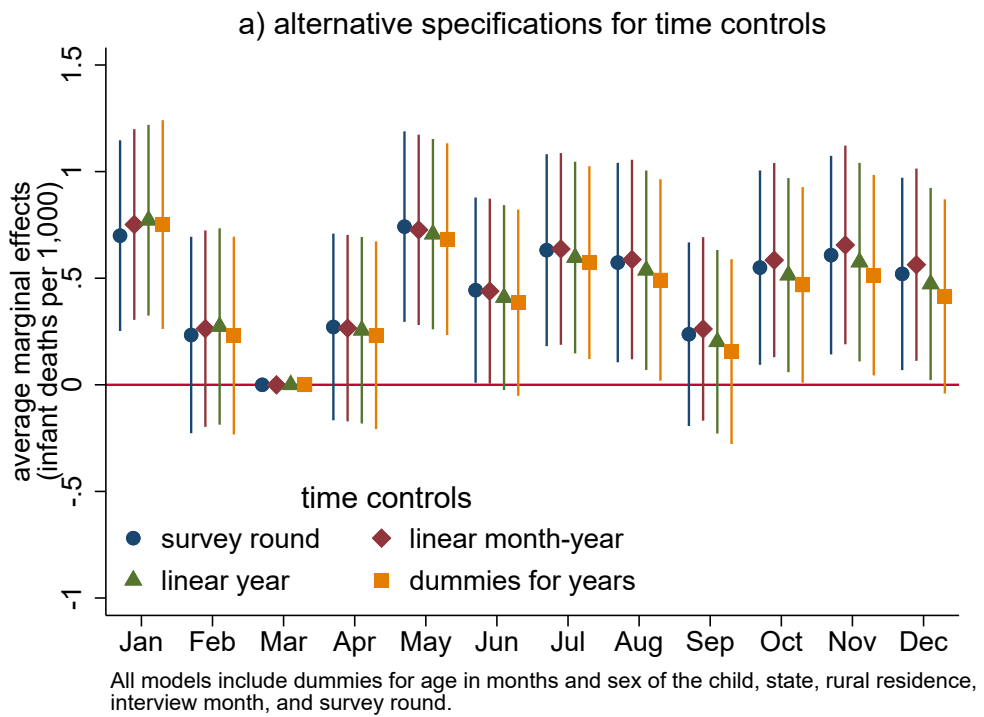
Figure 2A–5: Regression margins: Difference between January and previous-year December for infant mortality
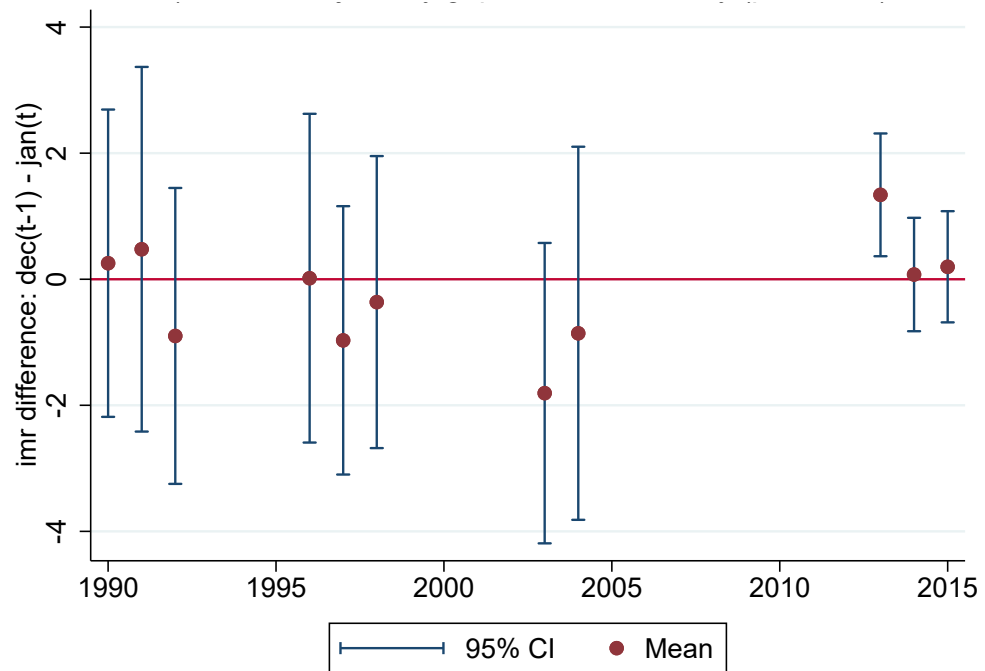
Figure 2A–6: Seasonal variation by access to household environmental technologies
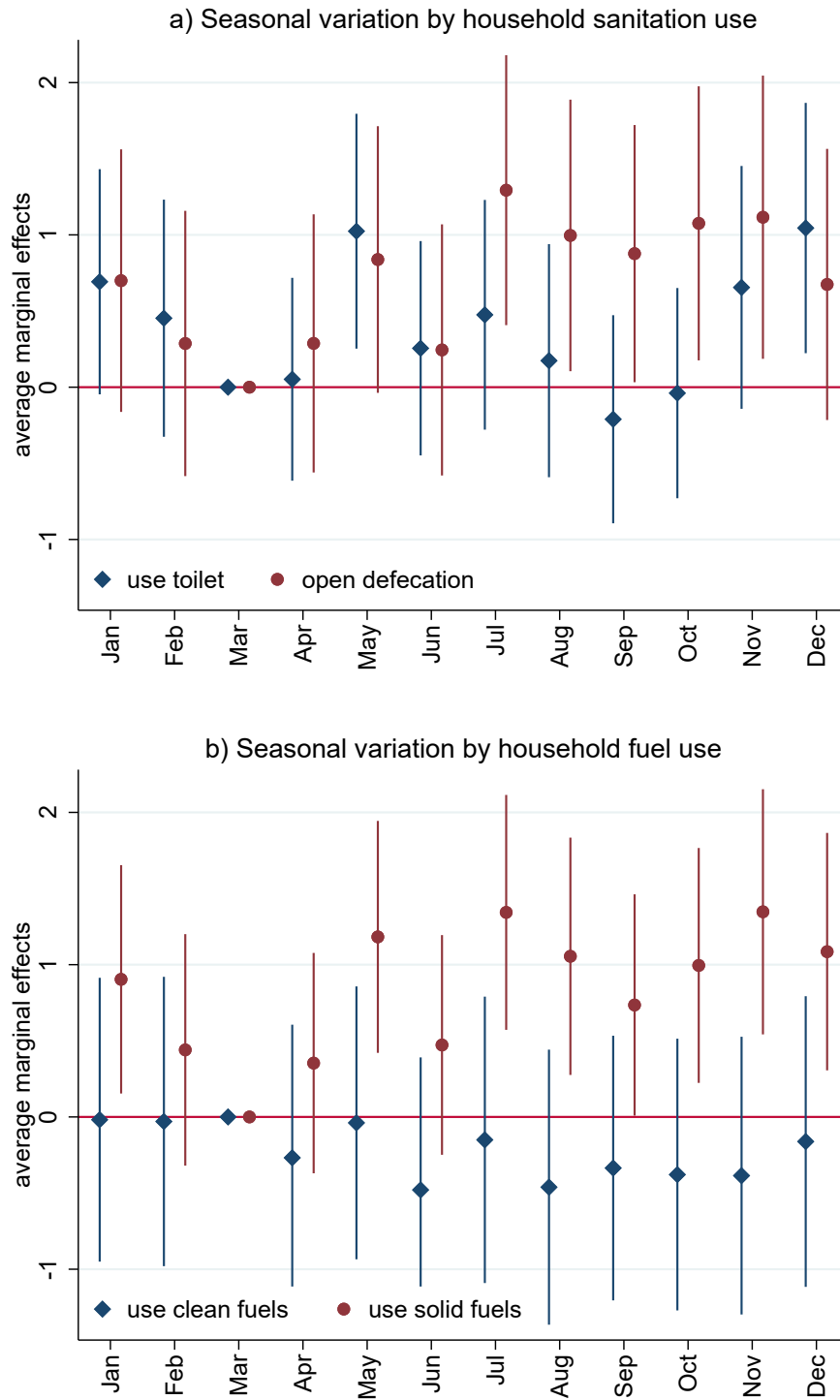
a) Seasonal variation by household sanitation use

b) Seasonal variation by household fuel use

Regressions control for dummies of sex and age in months of the child,
survey round, interview month, rural residence, and asset counts.

Figure 2A–7: Heaping in reporting age at death and mortality estimates with and without mortality at age 6 months
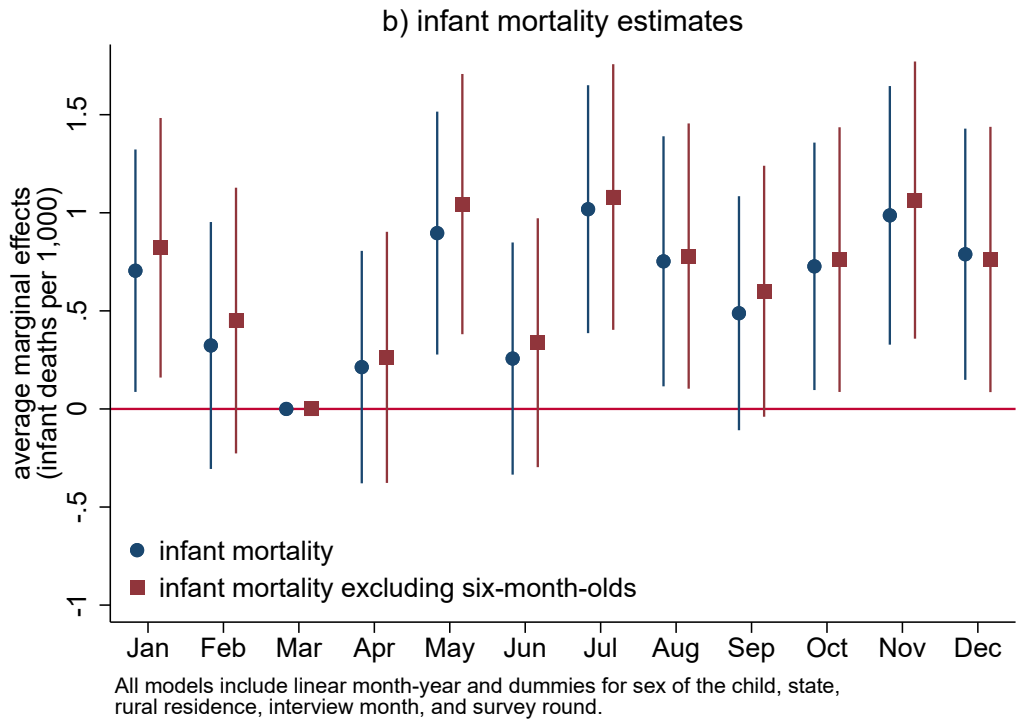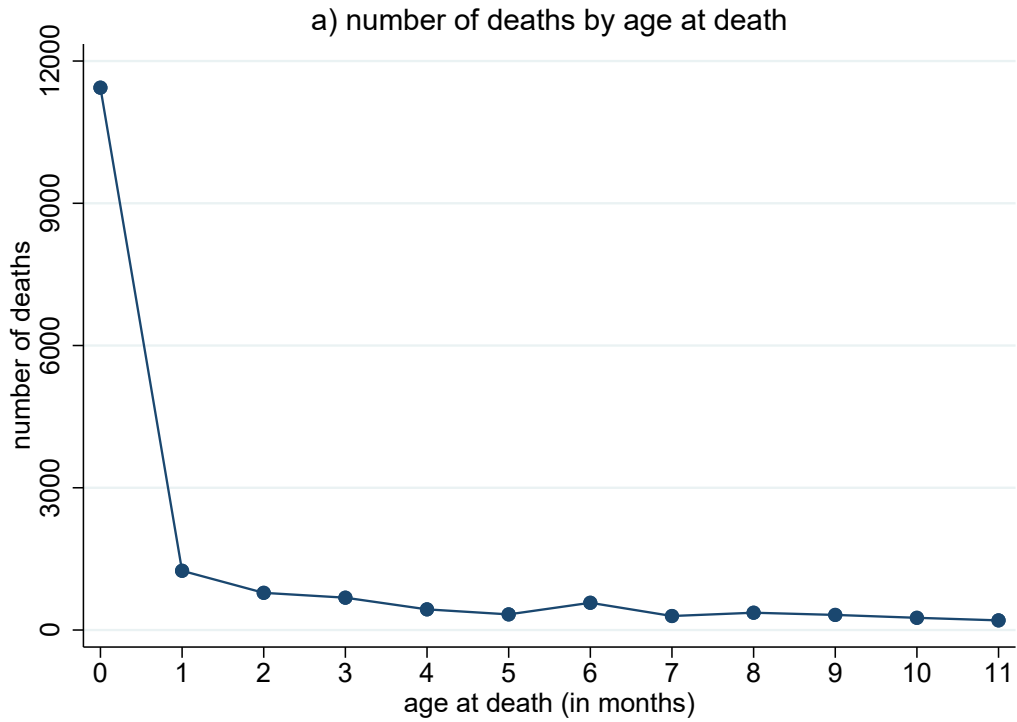
Figure 2A–8: Unadjusted and age-standardized $_1m_0$ based on alternative estimation approaches

unadjusted rates

Note: Mean and 95% CI estimates account for survey weights and clustering of observations within the primary sampling unit.

age-standardized rates

Note: Mean and 95% CI estimates account for survey weights and clustering of observations within the primary sampling unit.
Estimates are age-standardized to match the distribution of age (in monhts) for each calendar month with the overall age distrbution.

Note: Observations are from pooled sample, for child-months from all four NFHS rounds.

CHAPTER 3:

Household wealth and life expectancy in India: 1990-2016

Abstract

Research on life expectancy gradients in low- and middle-income contexts is constrained by data limitations. Using data from three Demographic and Health Surveys in India, I generate the first direct estimates of period life table quantities and their standard errors by household wealth in India. These estimates reveal large and persistent disparities in mortality. In 2013-16, life expectancy at birth for the wealthiest women was 9.2 years higher than the poorest women. Differences among men were even starker: 14.1 years higher among the wealthiest octile when compared to the poorest. Disparities have declined between the 1990s and the mid 2010s, and life expectancy at birth has improved across wealth classes. To a large extent, these patterns are driven by declines in child mortality levels and differentials. In contrast, disparities in adult mortality have not declined, and among men, even overall declines in adult mortality are not observed. We find that life expectancy disparities are higher in urban areas compared to rural areas. Regionally, life expectancies of those in the poorest quartiles are lowest in the Hindi-speaking states. In fact, the poorest women who live in South India have similar or higher life expectancies than the richest women who live in Uttar Pradesh. These findings highlight the persistence of disparities and deficits in overall well-being in India. They have several implications for the scientific literature on measurement of mortality, health system priorities, and understanding the evolution of health disparities in the ongoing epidemiological transition.

3.1 Introduction and background

A broad scientific literature has observed social class differentials in health and mortality in a variety of contexts (Elo, 2009). Apart from improving understanding of the distribution of mortality within populations, this literature has informed policy and health system priorities (Gwatkin, 1993; Deaton, 2002; Heuveline et al., 2002). Identifying the sections of a population that have higher mortality relative to others is the first step in understanding the causes of deficits and appropriate social responses (Preston and Taubman, 1994). Given that life expectancy is a critical human development outcome, evidence on mortality disparities is also central to public discussions on overall social progress (Sen, 2001; Krieger, 2011; Marmot, 2015). However, data limitations in low- and middle-income countries, especially on mortality beyond the childhood ages, prevent systematic examination of life expectancy disparities (Krueger and Burgard, 2011; Tabutin et al., 2017).

Child mortality differentials have been observed in studies of contemporary high income countries (Chen et al., 2016), historical patterns in high income countries (Preston and Haines, 2014), as well as in contemporary low- and middle-income countries (Gwatkin et al., 2007; Chao et al., 2018). Chao et al. (2018) show that absolute mortality differentials have reduced in low- and middle-income countries between 1990 and 2016. These comparisons of child mortality in households in the poorest and richest asset classes in low- and middle-income countries show that declines are partly driven by improvement in living standards and increased access to maternal and child healthcare (Boerma et al., 2018; OPHI and UNDP, 2020). Still, even when absolute disparities declined, relative child mortality disparities were stable in in this period, suggesting an even greater emphasis on improving child health in poorer populations within countries.

Unlike child mortality, patterns and trajectories of adult mortality gradients in low- and middle-income countries are not well-understood. This is despite the growing importance of mortality in the adult ages to overall population health. As child mortality declines, a greater proportion of the population and deaths are in the adult ages (Heligman

et al., 1993; Meslé and Vallin, 2011). Studies from the United States and some other countries have found increases in adult mortality disparities in recent periods (Shkolnikov et al., 2012; National Academies of Sciences, 2015; Chetty et al., 2016; Bosworth, 2018; Venkataramani et al., 2020; Committee on Population, 2021). Studies on trends in adult mortality disparities in low-and middle-income countries are less common. Single-period findings have found modest gradients in adult mortality in these countries (Chapoto et al., 2012; Sudharsanan, 2019; Beltrán-Sánchez et al., 2020; Luo and Xie, 2014; Nikoi and Odimegwu, 2013; Sudharsanan et al., 2020). A key explanation of these modest-to-no gradients has been that chronic diseases are emerging health risks in these countries, and may even be more prevalent among more advantaged sections (Rosero-Bixby and Dow, 2009; Sudharsanan and Ho, 2020).

A high-burden of risk-factors such as air pollution, on the other hand, may affect socially disadvantaged individuals more (Gupta, 2019). Disparities may also be shaped by a high-burden of accidental and tuberculosis mortality, both of which are more likely to affect poorer individuals in their productive ages (Jagnoor et al., 2012; Pai et al., 2017). Finally, adult consequences of child health insults in the presence of child mortality gradients will also predict adult mortality gradients (Mosley and Gray, 1993).

What are the overall impacts of declines in mortality (Deaton, 2013), changing health-risk profiles along with the epidemiological transition (Horiuchi, 1999; Caldwell and Caldwell, 1993), and generalizing cardio-vascular disease burdens (Patel et al., 2019) on gradients in life expectancy at birth? This paper addresses these questions for the case of India, the world's largest democracy and home to more people than the combined populations of North America and Europe. Using data from three rounds of the National Family and Health Surveys (NFHS), I investigate life expectancy disparities by household wealth. These three NFHS rounds, fielded in 1992-93, 1998-99, and 2015-16, included a module on recent deaths in the household. This is in addition to the birth history module that the Demographic and Health Surveys, of which the NFHS is the Indian version,

implement.

Combining data from the birth history module for ages 0-5, household deaths for mortality in ages beyond 5, and persons alive from household rosters, I directly estimate period life table quantities. Following a principal component approach, I estimate household wealth categories. This approach creates a wealth index using the observed durable goods that a household owns. I estimate standard errors and 95% Confidence Intervals for these life table quantities by using a cluster-bootstrap approach. The approach used here builds on methods to estimate age-specific mortality rates (Gupta and Sudharsanan, 2020) and their standard errors from clustered survey data (Vyas et al., 2021), and estimating household wealth indexes using data on household assets (Pollitt et al., 1993; Filmer and Pritchett, 2001).

The paper makes methodological contributions to studying the determinants of mortality using survey data in developing countries, where mortality registration is incomplete (Mikkelsen et al., 2015). To my knowledge, these are the first direct estimates of social class disparities in abridged life table quantities with accurate standard errors that account for the clustering of observations within primary sampling units. Substantively, the paper helps understand disparities in both overall life expectancy as well as mortality across the life course. A unique contribution of this study is that the data allow us to examine mortality disparities across the life course, unlike previous studies which focus on child or adult ages separately. Unlike the previous literature, this study is also able to examine trajectories in life expectancy disparities.

The paper intervenes in three key debates. First, it informs the India specific literature on the evolution of inequality in human development in the period 1900-2016. This period saw a robust economic growth (Drèze and Sen, 2013), a widening of economic inequalities (Himanshu, 2019), a decline in poverty (Alkire et al., 2021), and an expansion in social and health policy (Drèze and Khera, 2017). In contrast to patterns of increasing economic inequality despite poverty declines, the paper finds overall health improvements

and declines in mortality disparities between 1990 and 2016. Declines in child mortality, particularly among the poor, contributed to the reduction in overall life expectancy. However, mortality disparities in the adult ages were persistent in this period. Among men, even overall declines were not observed, pointing to a lost decades in terms of adult health improvements.

Second, it informs a policy debate on health system priorities, which has worried about equity considerations of allocating resources to addressing cardiovascular disease (Subramanian et al., 2013; Narayan and Ali, 2013; Patel et al., 2019). In documenting both child and adult mortality gradients, the paper suggests that investments in reducing the burden of adult mortality risk factors, particularly in the working ages, is not likely to increase inequality in health spending.

Third, in contrast to existing studies which document modest mortality gradients in adult mortality, the paper finds persistent disparity across the life course. In this, the paper suggests that India is different from other low- and middle-income countries where adult mortality gradients are modest. This finding highlights the relevance of fundamental cause perspectives (Link and Phelan, 1995) in India.

The next section describes the data and empirical approach followed here. We describe the results after that. Because prior demographic literature has emphasized the importance of sex, region, and residence in India (Murthi et al., 1995; Subramanian et al., 2006; Saikia and Ram, 2010; Saikia et al., 2013) and elsewhere (Elo et al., 2019; Mackenbach et al., 2003; De Walque and Filmer, 2011; Zajacova and Hummer, 2009), I break down key results by these factors. I discuss the results of several robustness tests in section 3.4. Complementary papers have discussed disparities by social group (Gupta and Sudharsanan, 2020) and the intersection of social-group and economic disparities (Vyas et al., 2021). This paper focuses on socioeconomic status as measured by household wealth. Section 3.5 discusses the strengths and limitations of the methodological approach followed here as well as implications for research. Section 3.6 concludes.

3.2 Data and methods

I use data from three rounds of the National Family Health Surveys (NFHS), India's version of the Demographic and Health Surveys (DHS). Four sets of measures are used from these data.

First, I use a question on recent deaths in the household to estimate mortality rates beyond the childhood ages. Out of the four available rounds of the NFHS surveys, three rounds – those conducted in 1992-93, 1998-99, and 2015-16 – implemented this module on recent deaths in the household. This question was asked in the household questionnaire of the surveys.

Second, I use information on usual household members listed in the household roster in the household interview questionnaire to estimate person years lived by those who did not die.

Third, the household questionnaire also asks respondents about the durable assets and household goods that it owns. I use this information to create a household wealth index.

According to the interviewer manual of NFHS-4, the NFHS household interview module can be answered by "any adult member of the household who is capable of providing information needed to fill in the Household Questionnaire" (IIPS and ICF, 2014). The NFHS surveys visit randomly selected households within randomly selected primary sampling units. The households are randomly selected from a list of all households that is prepared for the survey. Appendix figure 3A–1 shows that the household sizes of selected households in the NFHS surveys is similar to that observed in the decennial Indian censuses. Reports compiled by the NFHS surveys (IIPS, 1995; IIPS and ORC Macro, 2001; IIPS and ICF, 2017) also confirm that other characteristics measured in the NFHS surveys, such as age distribution, marital status, and sex ratios are similar to observed characteristics in the censuses and the Sample Registration System in India.

Finally, child mortality is estimated using the birth history module. This module is included in the women's questionnaire of the surveys. This question is answered by all eligible women in the household – women between the ages 15-49 – where they list all the children they have ever had and the survival status of their children.

Using this data, I create wealth- and sex-specific life-lines for dead and alive household members. I use standard life table methods (Preston et al., 2001) to estimate life tables by household wealth categories. To estimate standard errors and 95% confidence intervals for estimated life expectancies, I use a cluster-bootstrap approach. These data and methods are described in detail below.

Estimating mortality

The NFHS surveys are multi-stage sample surveys that are representative at the state and national levels in India. Given their scale, these surveys are implemented in phases, surveying several states in each phase. In Phase 1 of NFHS-1, information on deaths in the household were asked for the period January 1990 till the date of the interview. For phase 2, information on deaths between January 1991 and the interview date was asked. NFHS-2 had three phases, which asked about deaths between the periods January 1996, 1997, or 1998 and the interview date, respectively. In NFHS-4, the period for which information on deaths were asked was January 2012 till the survey date in the first phase, and January 2013 till the survey date in the second phase. To minimize recall biases, I use information on deaths in the last two years in these surveys. Person years are also estimated for this period. Thus, for NFHS-1, the estimates are for the period 1990-1993, for NFHS-2, they are for the period 1996-1999, and for NFHS-4, they are for the period 2013-2016.

The question on recent deaths was implemented in the household questionnaire of the surveys. I follow the approach developed in Gupta and Sudharsanan (2020) to estimate mortality rates. For all deaths recorded in the household death question, respondents were asked about the age and sex of the deceased, and the month and year when they died. I use this information to create life-lines for all dead individuals.

I use information from the household roster to create life-lines for alive individuals as well. For this, I use information on the age of alive individuals at the time of the interview, and create a random month of birth for them. These life-lines for alive or dead individuals start at age 5.

Then, I estimate the number of deaths, persons years lived by dead individuals, and person years lived by alive individuals in abridged life table age-groups in the estimation period. Although researchers have not used these data to directly estimate life expectancies or age-specific mortality rates, these data have been used in regression-based or indirect-estimation approaches to evaluate specific determinants of mortality (Saikia and Ram, 2010; Subramanian et al., 2006; United Nations Population Division, 2011; Asaria et al., 2019). Regression based approaches have been used to study the association between socioeconomic status and mortality using data from the India Human Development Survey as well (Po and Subramanian, 2011; Barik et al., 2018).

For ages below 5, I use the birth history module of the NFHS surveys, using standard approaches (Moultrie et al., 2013). I estimate age-specific mortality rates in the ages 0-1 and 1-4 years by using information provided by mothers who list their birth history in the DHS surveys. I create life-lines for children between the ages 0-1 and 1-4 in the estimation period by using information on their month and year of birth, month and year of death if they died, and the interview month and year. For consistency, these rates are estimated for the same period for which adult mortality rates are estimated.

Deaths and the exposures are first estimated at the level of the calendar month, which are then summed into deaths and person-years in the estimation period. The entry month for children is at the beginning of the estimation period if they were born before the beginning of the estimation period. If they were born in the estimation period, their entry date is the month-year they were born. Exit dates for children are the month-year they turned 59 months if the date is before the end of the estimation period; month-year of death if they died in the estimation period; and the month before the month-year of interview if

they did not turn 5 and did not die in the estimation period. For adults aged 5 and above, the entry date is the month-year they turned 5, or the beginning of the estimation period if they turned 5 before the estimation period began. Exits are on the month-year of death or the month before the month-year of the interview.

Although an older demographic literature has emphasized the unreliability of the question on recent deaths in the household (Hill, 1991; Timæus, 1991), Gupta and Sudharsanan (2020) show that the mortality rates estimated using this approach match closely with those of India's national life tables estimated by the Sample Registration Survey. More recent assessments from other contexts, such as those from Sub-Saharan Africa, also finds increased reliability of the recent deaths in the household approach (Masquelier et al., 2017). In Bangladesh, El Arifeen et al. (2014) and Hill et al. (2006) use the question on recent deaths in the household to measure maternal mortality, and find similar results as questions on sibling survival. In Appendix Figure 3A–2, I show that the age-specific mortality rates generated using this approach correspond closely to official Indian life tables based on the Sample Registration Survey (SRS).[1]

Estimating wealth

In the literature on social gradients in mortality, a variety of measures, including occupational class (Marmot et al., 1991; Bengtsson et al., 2019), education (Elo and Preston, 1996; Luo et al., 2015), income (Dowd et al., 2011; Chetty et al., 2016), and household wealth (Laaksonen et al., 2009; Barik et al., 2018) have been used to study the relationship between socioeconomic status and health (Elo, 2009). In this paper, I use a household wealth index based on the durable goods that a household owns, created using principal component index (Pollitt et al., 1993; Filmer and Pritchett, 2001). Each measure of socioeconomic status has its own advantages and disadvantages (Elo, 2009).

---

[1]Although the SRS is also a survey, accurate standard errors that reflect the multi-stage sampling design of the SRS are not published by the SRS. The NFHS standard errors are calculated using a clustered bootstrap approach, as described in the section on empirical strategy. If individual level SRS data are made available, the cluster-bootstrap strategy used here would be applicable to the SRS data as well.

Household assets and indexes based on them have the advantage that they are available for all individuals (unlike occupational class), for all ages (unlike education), and are easier to measure in developing countries where income data are harder to collect (Filmer and Pritchett, 2001). When compared to income, reverse-causality is less of a concern with household durable assets, although households may still sell some assets after the death of a household member. Along with social group, such as caste or indigenous identity, the assets that a household own may be a key marker for social class in India. In addition to measuring relative social class, household based wealth indices may also capture social class perceptions, given that they are easier to display and observe (Ramakrishnan et al., 2020).

A limitation of the household asset index approach is that household assets are measured at the time of the interview, after deaths have occurred (Chao et al., 2018). Unfortunately, the NFHS does not ask questions on household income, and educational attainment is only collected for alive members. However, socioeconomic status as measured by wealth index of household assets is a common approach to investigating inequality in well-being using the DHS surveys, particularly for child mortality (Lu et al., 2020).

Using principal components analysis, I create a household wealth index based on the durable goods a household is reported to own. I exclude those household goods that have direct links with health and mortality, such as toilet ownership (Geruso and Spears, 2018b) and solid fuel use (Gupta, 2019). I also use those assets which are linked to higher social status and were technologically relevant in a particular period. From NFHS-1, I use the household assets radio, television, refrigerator, cycle, motorized two-wheeler, car, sewing machine, clock, sofa, fan, whether the house was pucca, and video-cassette-recorder. From NFHS-2, I use the assets television, refrigerator, mattress, motorized two-wheeler, car, land-line phone, pressure cooker, chair, cot, table, clock, fan, whether the house was pucca, and sewing machine. From NFHS-4, I use the assets television, refrigerator, car, landline phone, pressure cooker, fan, computer, air-conditioner, and washing

machine. Each subsequent round of the NFHS surveys observes a larger number of house-hold assets and there are some changes in each round about particular assets (for instance video-cassette recorders or mattress). I estimate separate wealth indexes for all households in each round's overall, rural, and urban samples.

Empirical strategy

For each household, the household wealth index is the first component of a principal com-ponent analysis of the assets it owns. This wealth index, w, is a continuous function of the assets a household owns. Although the relationship between wealth and mortality is con-tinuous, life expectancy and other life-table quantities are not individual level outcomes. Therefore, I split the wealth-index into categories, wc.

To ensure that there are sufficient number of deaths in each wealth category, I split the wealth index into octile for the overall sample in each round. For the rural and urban sample in each round, regional patterns, and sex-ratios of mortality rates, the wealth index is split by quartiles. Because of a small number of deaths in each wealth category in the first two NFHS surveys, and the short time interval between them, I create life tables after combining them. Thus, in these two surveys, although wealth quartiles are estimated by round and sample separately, mortality estimates are for both rounds together. Hence, the estimates reported here are for the period 1990-99 and 2013-2016. Taking advantage of the larger sample size of the NFHS-4 surveys, I also split the overall NFHS-4 sample into ventiles, or twenty categories of the wealth index.

For each wealth category, wc, I estimate age-specific mortality rates, $_nm_x$ in abridged life-table ages x to $x+n$, in period t, for sex g,

$$_nm_x^{wc,s,t,g} = \frac{_nd_x^{wc,s,t,g}}{_nL_x^{wc,s,t,g}} \tag{3.1}$$

where s is the overall, rural, or urban sample; g is either male or female; $_nd_x$ is the estimated

number of deaths in ages x to $x + n$ in the period, and $_nL_x$ is the number of person-years lived in ages x to $x + n$. All estimates are weighted using the national household weights calculated by the NFHS.

I borrow values for person years lived by those who died, $_na_x$, (Preston et al., 2001, pg. 45) from national Indian life tables based on the India's Sample Registration System (Office of the Registrar General and Census Commissioner of India, 2018). I use separate $_na_x$ values by period, sex, and sample for the $_nm_x \rightarrow {_nq_x}$ conversion, and calculate period life expectancies using standard life table methods.

To calculate standard errors and 95% CIs for the life expectancy (e_x) estimates that I obtain, I use a cluster-bootstrap approach (Vyas et al., 2021; Fishman, 2015). The NFHS surveys are multi-stage sample surveys that sampled households within randomly sampled primary sampling units. Primary Sampling Units were either villages in rural or census enumeration blocks in urban areas. The strata used for selection of primary sampling units were rural and urban areas within states. I repeatedly sample primary sampling units, with replacement, for each sex-, period-, sample-specific wealth categories, maintaining the distribution of primary sampling units within strata. I draw 100 bootstrap samples, Z^{*1} ... Z^{*100}. Thus, for each period, sample, sex, and wealth category, I generate 100 life expectancy estimates: $\hat{e}_x^{*1}$ ... $\hat{e}_x^{*100}$. Standard errors for life expectancy are the standard deviation of bootstrapped estimates:

$$SE(\hat{e_x}) = \sqrt{\frac{1}{100-1} \sum_{b=1}^{100} \left(\hat{e}_x^{*b} - \frac{1}{100} \sum_{r=1}^{100} e_x^{*b} \right)^2} \tag{3.2}$$

where b denotes the specific bootstrap sample drawn, between 1 and 100. I assume that the estimator for life expectancies is normally distributed. Thus, 95% Confidence Intervals are calculated as $e_x \pm 1.96 * SE(\hat{e_x})$.

I also investigate regional patterns of these disparities. As Chetty et al. (2016) examine in the US context, this analysis asks: where do those in the poorest (and other)

quartiles have the highest life expectancies in India? For these sets of regional analyses, I use household wealth quartiles estimated nationally. For SEs and CIs, I draw 100 bootstrap for each region. Additionally, because behaviors and social factors linked to mortality differ by gender (Guillot, 2002; Canudas-Romo et al., 2016), I compute male-female mortality ratios by household wealth quartile.

3.3 Results

Estimated person years and deaths by sex, sample, and wealth quartile are shown in Appendix Table 3A–1. Because of their smaller sample sizes, there are comparatively lesser number of deaths and person-years in each quartile of the first two NFHS Surveys. This is particularly true for urban areas when we split the samples by residence. Given these smaller sample sizes, I combine these earlier two rounds. We observe a larger number of deaths and person years in each category in the latest NFHS round because of its larger sample size.

Overall patterns

How large are disparities in life expectancy at birth? Figure 3–1 shows life expectancy at birth and 95% CIs by wealth octile for men and women.[2] The figure shows large gradients in life expectancy, and a flattening of wealth gradients, particularly among women. Life expectancy at birth for the poorest octile of women in the 1990s was 56.9 years, about 14 years less than the richest octile of women. By 2013-2016, the difference between the poorest and wealthiest octile of women were more than 9 years. Disparities among men were even larger: more than 14 years between the poorest and wealthiest octiles in 1990-99 and more than 12 years in 2013-2016. Despite absolute and relative reductions in differentials, these disparities remain large, especially in a comparative perspective (Marmot, 2015).[3]

---

[2]Confidence intervals for the earlier NFHS rounds are slightly larger, reflecting their smaller sample sizes. Confidence intervals also vary across octiles to some extent. Octiles are based on household wealth-index, and given variation in household size and mortality, some octile have more observations than others. This can be seen in Appendix Table 3A–1 for quartiles as well.

[3]It is possible to split the NFHS-4 sample even further, such as in ventiles, as shown in Appendix Figure 3A–3 which shows even starker disparities.

Figure 3–2 shows that overall life expectancy disparities are driven by differences in both the childhood ages and the adult ages. The figure shows overall reductions in child mortality, as well as a flattening of the gradient across asset octiles. In contrast, however, the figure shows modest to no improvements in adult mortality, and large gradients. Gradients for adult mortality observed here are in contrast to patterns observed in other low- and middle-income countries, where mortality gradients have been observed to be modest or non-existent (Sudharsanan, 2019; Beltrán-Sánchez et al., 2020).

Appendix Figure 3A–5 shows probability of death in across the life course.[4] The age ranges displayed in the figure correspond to broad life course stages – childhood (ages 0-4), schooling and adolescence ages (5-19), working ages where chronic diseases are not observed (20-39), working ages which mark the onset of chronic diseases (40-59), and older ages (60-84). Mortality disparities by wealth can be observed till age 60. Overall improvements are modest both for both working younger and older age-groups. Disparities in the older ages (60-84) are modest and do not follow clear pattern. Within the prior literature, similar patterns have been observed by Kitagawa and Hauser (1973) for the United States. These patterns are likely to be driven by a combination of selection into survival to age 60; age-misreporting, which biases mortality estimates downwards (Preston et al., 1999) and is more likely to be observed among the poor; deaths among single-person households, who are more likely to be poor (Dommaraju, 2015); and modest social and employment security provisions for the poorest old, which has expanded to an extent in India (Drèze and Khera, 2017; Reed, 2020; Gupta, 2013). To the extent that these patterns are driven by age-misreporting or missing deaths among single-person households, the overall life-expectancy disparities documented here are conservative.

Disparities by residence and region

How do these disparities vary by residence and region? Figure 3–3 shows life expectancies at birth by household wealth quartile, sex, and residence. In both rural and urban areas,

---

[4]Appendix Figure 3A–4 shows disparities at age 0, 15, and 60.

disparities in life expectancies have reduced, partly driven by mortality declines among the poorest quartiles. Still, large inequalities remain. Among women, differences between the wealthiest and poorest quartiles in 2013-2016 were 6.1 years in urban areas, and 5.3 years in rural areas. Among men, these differences were 10.2 years in urban areas and 8.6 years in rural areas.

The figure shows higher life expectancy disparities in urban areas when compared to rural areas, even when urban areas have overall higher life expectancies. These patterns are different from the ones observed by Sudharsanan and Ho (2020) for Indonesia, where rural areas have lower mortality than urban areas because of lower prevalence of cardio-vascular disease risk-factors.

Figure 3–4 shows regional patterns. A large literature in Indian demography has focused on regional patterns, following Dyson and Moore (1983) and Murthi et al. (1995). I divide India's 29 states into ten regions. Regions are defined partly by geographic conti-guity, and partly by the need to have sufficient number of deaths and person-years lived to estimate life expectancies. For this reason, patterns for the 1990s are not shown as they have large sampling variability. Maps here document where women and men from respec-tive quartiles have high and low life expectancies.

We find that in 2013-16, variation by region within quartiles is higher for women than among men. In 2013-2016, women from the lowest quartiles have life expectancies as low as about 65 years in the states of Uttar Pradesh, Bihar and Jharkhand, and Madhya Pradesh and Chattisgarh, and as high as 72 years in states of south India and north-east India. Although the confidence intervals overlap, women in the lowest quartile who live in southern and north eastern states in India in fact have higher life expectancies than women in the wealthiest quartile in Uttar Pradesh, India's most populous state.

Apart from documenting the extreme disadvantage that women in rural North India face, these patterns show that overall disparities can be observed within regions as well. The overall disparities observed for India as a whole are not because of the co-

patterning of household wealth and mortality within India's regions.

Male-female mortality ratios by wealth quartile

Although females generally have a survival advantage across the life course (Barford et al., 2006), a large literature documents female survival disadvantage in certain younger age-groups in India (Sen, 1992). This survival disadvantage has been declining in India, with regional variations (Canudas-Romo et al., 2016). Female survival advantage can be seen in most ages now. How is female survival advantage or disadvantage patterned by household socioeconomic status? Figure 3–5 shows male-female mortality ratios by age-group and wealth quartile. These mortality ratios are calculated as the age-specific mortality rate for men in an age-group and quartile divided by the corresponding age-specific mortality rate for women in the same quartile. Because these are male-female mortality ratios, values greater than 1 (shown in shades of blue in the figure) reflect higher mortality among men. Shades of red show higher mortality among women.

Three patterns are worth noting. First, overall, between 1990-99 and 2013-16, female survival disadvantage has declined, and severe male disadvantage has emerged in the middle ages from 30-60. Second, while female disadvantage in childhood ages of 1-4 can be seen in the lower quartiles in 2013-2016, female survival disadvantage in the adolescent ages can be seen in the middle quartiles in 2013-2016. Female survival advantage in the latest NFHS round in the childhood ages in the wealthiest quartile may be a reflection of higher prenatal sex-determination and sex-selective abortions in the wealthiest quartiles (Das Gupta and Mari Bhat, 1997). As sex-selection increases, the girls who are born are more wanted, and this selection influences mortality patterns in the childhood ages. Still, in 2013-2016, the higher the quartile, the lower was the survival advantage among girls in the ages 0-1. Indeed, they had no survival advantage in the wealthiest quartile. Finally, men's survival disadvantage in 2013-2016 was particularly worse in the later productive ages in the second, third, and wealthiest quartiles. These patterns merit further attention, and may be driven by harmful health behaviors such as alcohol and tobacco consumption

among men.

3.4 Robustness

Age-misreporting influences life expectancy estimates, biasing mortality estimates downwards at older ages (Preston et al., 1999). Apart from a similar distribution of mortality risk-factors at the older ages, higher age-misreporting among poorer quartiles could also make life-expectancy estimates at age 60 among the poorer quartiles closer to those from wealthier quartiles. Figure 3A–8 examines age at death in single years by sex, period, and wealth quartile. Although age-heaping is severe among all quartiles in the older ages, it is slightly more severe among poorer quartiles, and particularly worse among women. This suggests that disparities in life expectancies at older ages reported here may be conservative.

Are 100 bootstraps enough to generate reliable estimates of SEs and CIs for these life expectancy estimates? Figure 3A–9 examines this. It shows how the 95% CIs change by the number of cluster-bootstraps drawn. Estimates are for life expectancy by sex and wealth quartile. Although less reliable when a small number of bootstrap samples are drawn, CIs do not change substantially after more than 75 bootstrap samples are drawn. Although they may still vary, the estimates shown in Figure 3A–9 show that the main interpretations drawn from these 100 bootstrap samples are robust.

A concern with questions on household deaths has been that household deaths lead to household fissures (Timæus, 1991). When compared to censuses, the probability of two households reporting the same death is less in survey settings. However, it is not zero. Appendix Figure 3A–10 reports the proportion of clusters (among all clusters that reported a death in an age group and sex) in which more than one deaths were observed in the same wealth-octile, age-group, and sex. We cannot ascertain the extent to which these deaths were of the same individual or of a different person. However, Appendix Figure 3A–10 provides evidence that the proportion of such clusters and deaths was small. Indeed, the

largest proportion of such deaths was in the childhood ages, which were observed in the birth history module. To some extent, this suggests that the vast majority of these deaths are likely to be of different individuals.

3.5 Discussion

Summary of results

This paper investigates mortality and life expectancy disparities in household wealth in India, as measured by an index of assets that a household owns. It finds that mortality disparities in India are large and can be observed across the life-course, and are particularly salient in the childhood, young-adulthood, and working ages. In addition to disparities, we document a lack of overall improvements for adults as well. These disparities have persisted despite declines in child mortality and changes in mortality risk-factors as the epidemiological transition has proceeded in India. Disparities can also be observed across regions and residence, are particularly high in urban areas. The paper documents severe mortality disadvantage among women in North India. The highest quartile of North Indian women in Bihar and Uttar Pradesh have similar or lower life expectancy than women in the lowest quartile in South India.

Strengths and limitations

The paper has some limitations. Mortality and wealth have a bi-directional relationship (Deaton, 2013), and assets are measured after mortality in the NFHS surveys. However, household assets may be less sensitive to mortality in the household than income. Additionally, the asset index here is operationalized as a relative index within survey rounds. Understanding the role of absolute improvements in socioeconomic status, either at a household or more aggregate level, would be helpful in understanding the association between socioeconomic status and mortality.

The demographic literature has noted a number of concerns on the approach of

estimating mortality rates from questions on recent deaths in the household (Hill, 1991; Timæus, 1991). This paper is able to address these concerns to some extent. For instance, both Timæus (1991) and Hill et al. (2006) raise concerns about large sampling variability in survey based estimates from the recent household deaths approach. To address these concerns, I empirically estimate standard errors and confidence intervals using a cluster bootstrap approach. I also combine the earlier two NFHS rounds. Overall, we find that the NFHS surveys are powered to study mortality disparities. A second concern that this literature has noted is downward biases in mortality estimates. Timæus (1991) for instance describes surveys and censuses where a majority of the deaths were missed. In contrast, estimates from NFHS estimates correspond closely in terms of overall mortality with India's national life tables estimated from the Sample Registration Surveys. In many ages, we observe higher age-specific mortality rates in the NFHS than in the SRS. Third, Hill (1991) notes that the recent household death approach may miss childhood deaths. This is not a concern here, since childhood mortality is estimated using the birth history module. Although we cannot rule out that household splits lead to a double counting of some deaths, the overall proportion of such deaths is small, and surveys which randomly sample households are less likely than censuses to observe the same death twice. Finally, age-misreporting among poorer older individuals and not observing deaths among those single-person households likely underestimates mortality in the poorest wealth categories. Even if these last two concerns cannot be addressed directly, they make the disparities documented here conservative.

The approach used here also has several strengths. First, the paper uses large-scale data to directly estimate empirical mortality rates for an important and long period in India. This period saw robust economic growth, a decline in absolute poverty, but an increase in economic inequality. This paper shows that this period also saw declines in child mortality disparities but persistent disparities in the working ages. Second, the paper examines the reliability of survey approaches to measure mortality across the life course and generate insights that have scientific and policy implications. Together with Vyas et al.

(2021) and Gupta and Sudharsanan (2020), the paper contributes to methods on estimating mortality rates and their standard errors. This approach may be relevant in other contexts which use multi-stage surveys. Third, relative to the previous literature, which has examined adult and child mortality disparities separately, the paper is able to understand mortality disparities across the life course. Finally, the paper contributes to the literature on the forms of socioeconomic disadvantage which predict mortality. It shows that in addition to caste and social group, mortality in India is also strongly patterned by class.

Policy implications

In the ongoing epidemiological transition in India, an academic debate has been concerned about equity in health spending and health system priorities (Subramanian et al., 2013; Narayan and Ali, 2013). Reducing child and maternal mortality has become a priority in India, specially with the National Rural Health Mission which was launched in 2005 (Drèze and Sen, 2013). From a policy perspective, prioritizing adult health may mean reducing funding for child mortality, especially if overall public health spending does not increase. Even if adult mortality gets additional allocations, a worry is that since adult mortality and adult diseases are not more prevalent among the poor, the marginal increase in health spending on adult diseases increases inequality in spending. This essay does not suggest the particular diseases to tackle or the policy approaches to pursue to reduce adult mortality. However, it suggests that since both child mortality and adult mortality are patterned by socioeconomic status, additional prioritization of adult health per se will not increase inequality in health spending. The paper shows that unlike child mortality, overall declines in adult mortality disparities are also modest. This makes prioritizing adult health an even more urgent priority.

From a broader societal perspective, this research has similarities and differences with research on economic and social inequality in India. As with other research documenting overall human development in India (Himanshu, 2019), we find life expectancy improvements across wealth classes. However, instead of increases in inequality, as docu-

mented by research on consumption patterns (Subramanian, 2020), we document declines in child mortality disparities, and no declines in adult mortality disparities. We find that overall improvements in life expectancy and declines in gradients are also driven by patterns in the childhood ages. In the adult ages, in addition to persistent gradients that did not decline, even overall mortality did not reduce substantially.

Research implications

Why is mortality, particularly in the childhood and working ages, consistently patterned by household wealth? A large literature has focused on the extent and causes of child mortality disparities in India and elsewhere (Saikia et al., 2013; Chalasani and Rutstein, 2014; Chao et al., 2018). In addition to documenting child mortality disparities in several low and middle income countries (Chao et al., 2018), this literature has shown that other markers of child health, such as anemia, weight, or height also have large gradients (Subramanyam et al., 2010; Coffey et al., 2019; Onyeneho et al., 2019).

Such large adult mortality disparities have not been documented in other low- and middle-income countries (Sudharsanan, 2019; Beltrán-Sánchez et al., 2020). Because cardiovascular disease risk factors are not patterned strongly by wealth Patel et al. (2019), this is a paradox. Except underweight, anemia, and lung function among adults, other adult mortality risk-factors such as blood glucose and hypertension are not more prevalent among the poor in India (Arokiasamy et al., 2021). The rich are more likely to be overweight in the Longitudinal Aging Study of India (Arokiasamy et al., 2021). The poor are more likely to have untreated hypertension. This is consistent with patterns observed in the India Human Development Survey – although they interpret self-reported cardiovascular diseases as prevalence of cardio-vascular diseases, Barik et al. (2018) also find higher access to healthcare among the rich, and lower adult mortality among them.

Investigating factors such as access to the healthcare system (Baru et al., 2010; Balarajan et al., 2011), exposures to pollution from solid fuel use (Gupta, 2019), and tuberculosis or accidental mortality may help understand why adult mortality is patterned by

household wealth. The extent to which class based inequalities are contributed by caste differences, as documented in Gupta and Sudharsanan (2020) and (Vyas et al., 2021), may also explain the patterns observed here. Finally, working conditions, and class-based inequalities themselves may be a factor slow adult health improvements. The causes of slow adult mortality improvement, which were also noted by Saikia et al. (2011), deserve further examination.

3.6 Conclusion

This paper documents persistent disparities in mortality across the life course in India. Disparities are large in a comparative perspective as well as when considering temporal patterns in India. The life expectancy that wealthy Indians achieved in the 1990s have not yet been observed for the average Indian. The poor, on the other hand, have life expectancies that were observed a hundred years ago in high-income contexts, or are seen in contemporary Sub-Saharan Africa. In contrast to child mortality, adult mortality levels and differentials have not declined. The paper documents an urgent need for social and health policy to address disparities and improve health in the childhood and adult ages for the poorest Indians.

Figure 3–1: Wealth disparities in life expectancy at birth
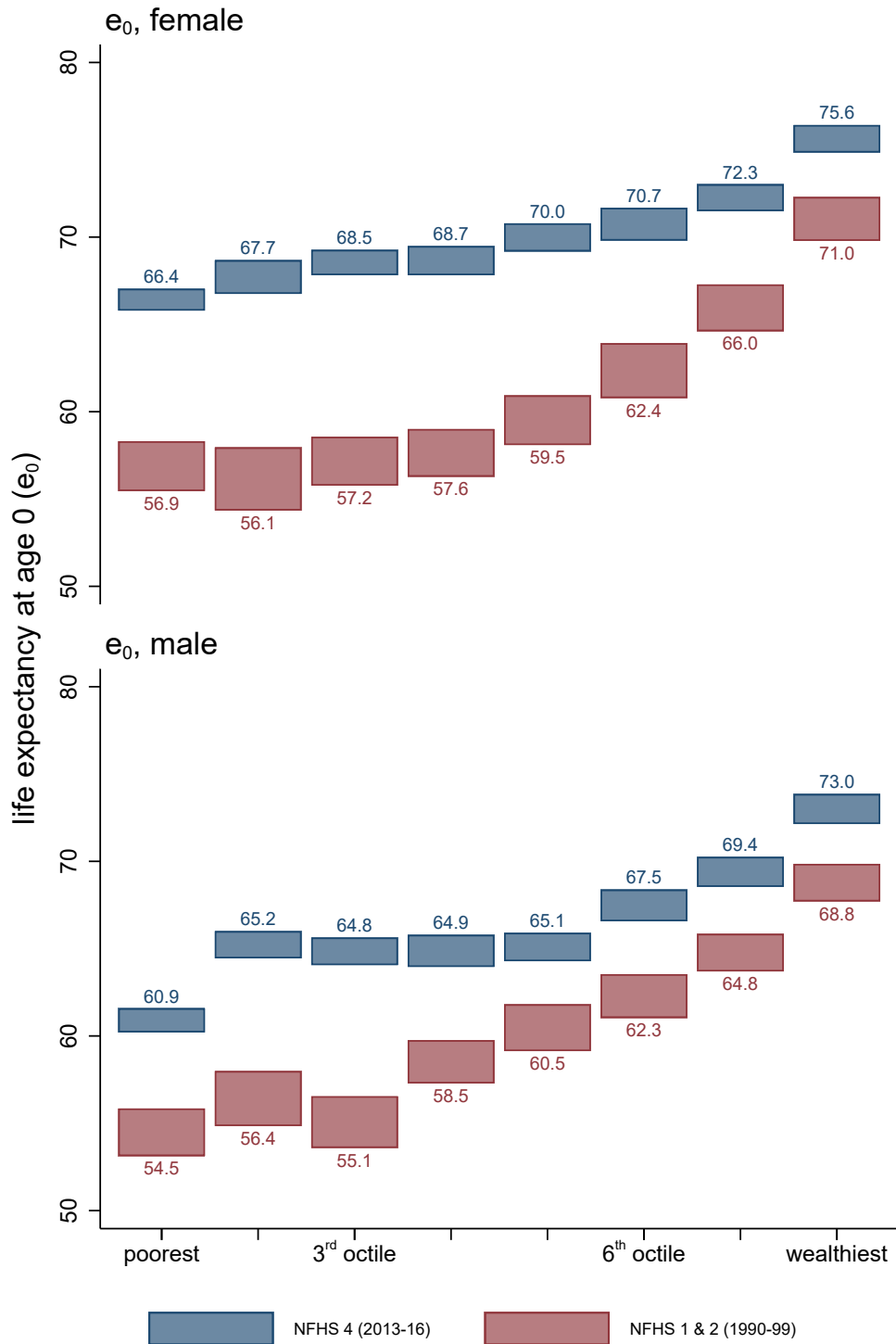
Figure 3–2: Disparities in child and adult mortality
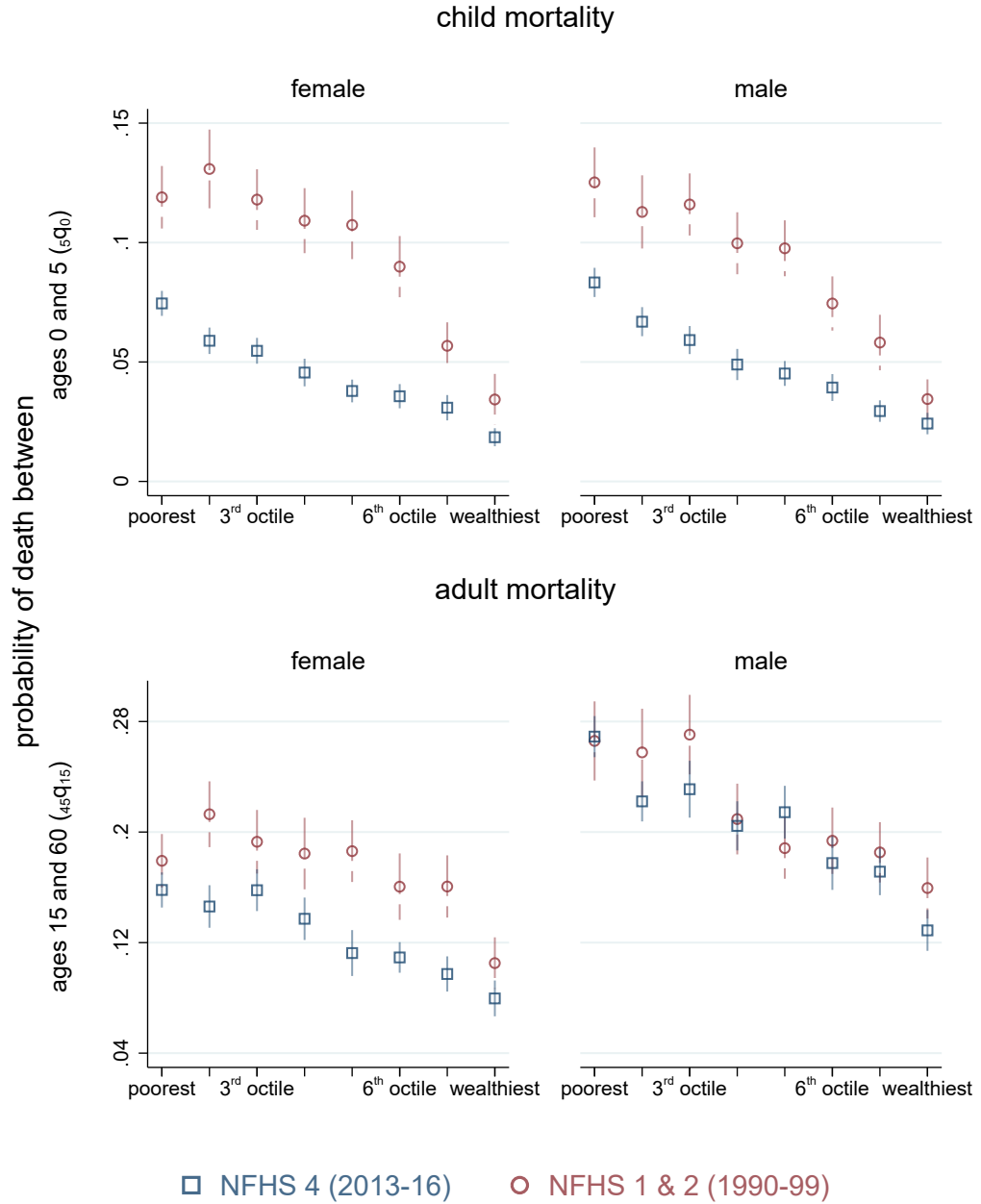
Note: 95% CIs estimated using a cluster-bootstrap approach in square brackets. The number of bootstrap samples drawn was 100.

Figure 3–3: Disparities in life expectancy at birth by residence
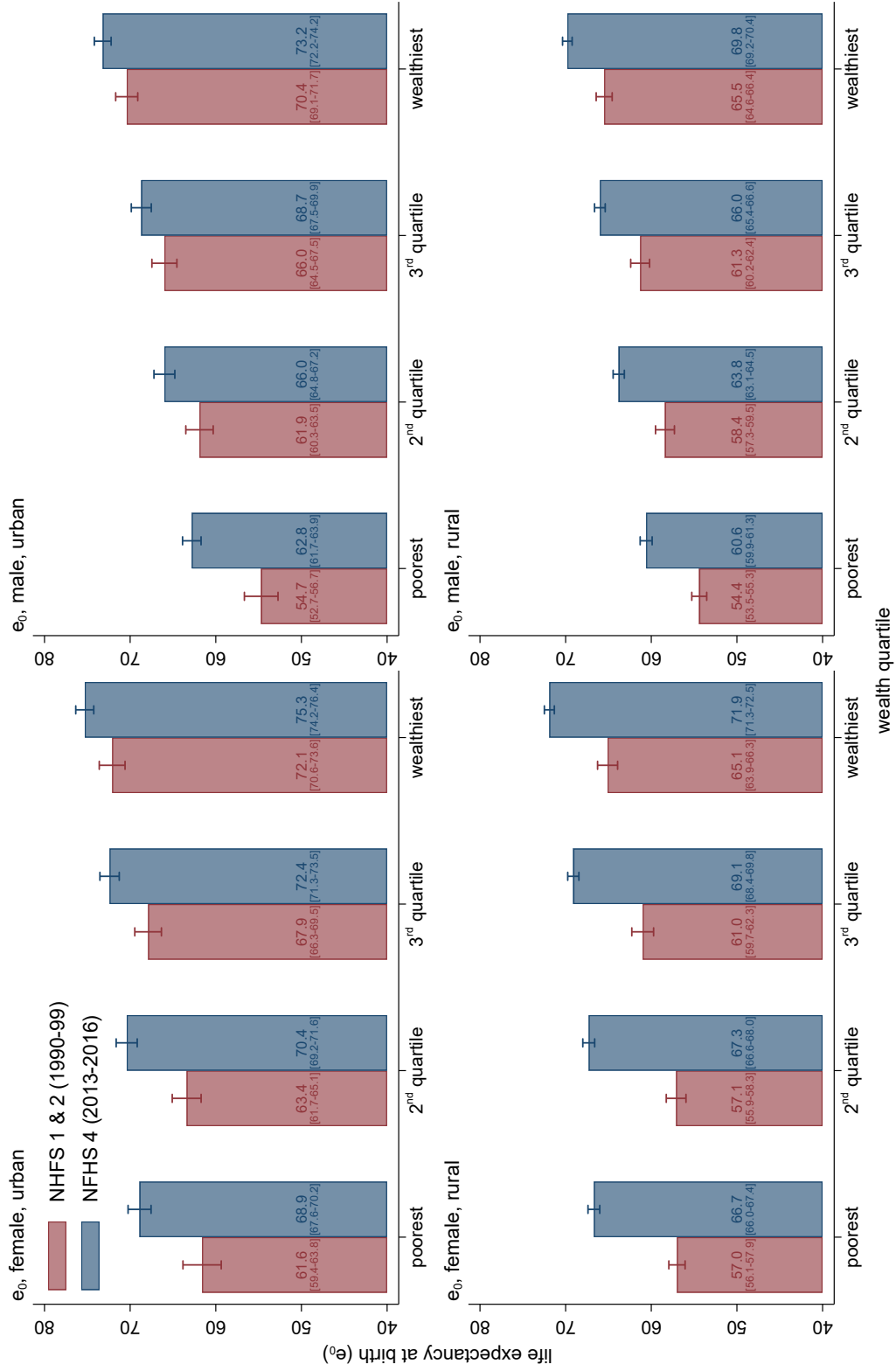
Note: 95% CIs estimated using a cluster-bootstrap approach in square brackets. The number of bootstrap samples drawn was 100.

Figure 3–4: Mortality disparities by region, 2013-2016
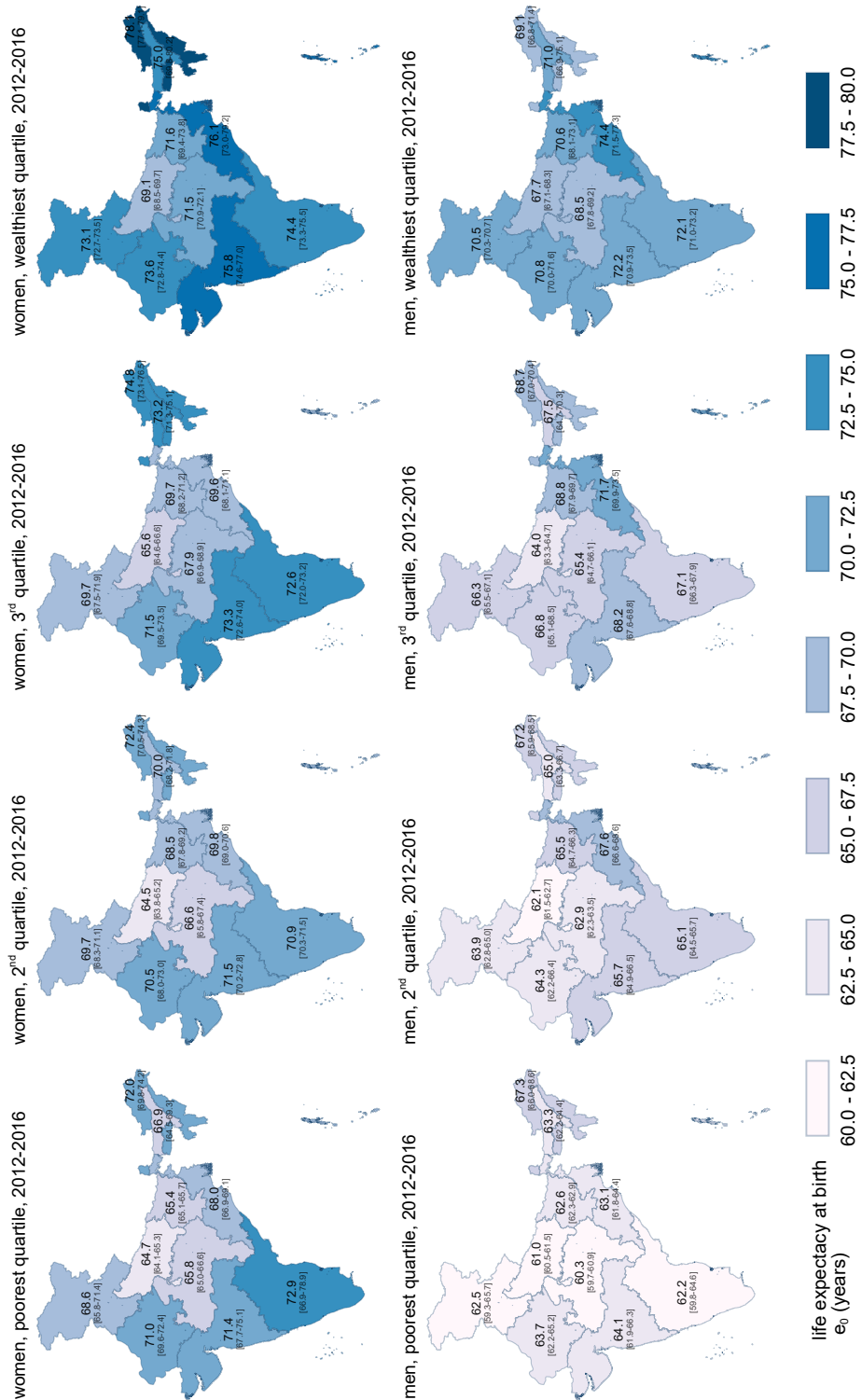
Note: Wealth quartiles are estimated for the all-India sample. Map labels are life expectancies at birth. 95% CIs estimated using a cluster-bootstrap approach are in square brackets. 100 bootstrap samples were drawn for each region.

111

Figure 3–5: Sex differences in mortality by wealth quartile
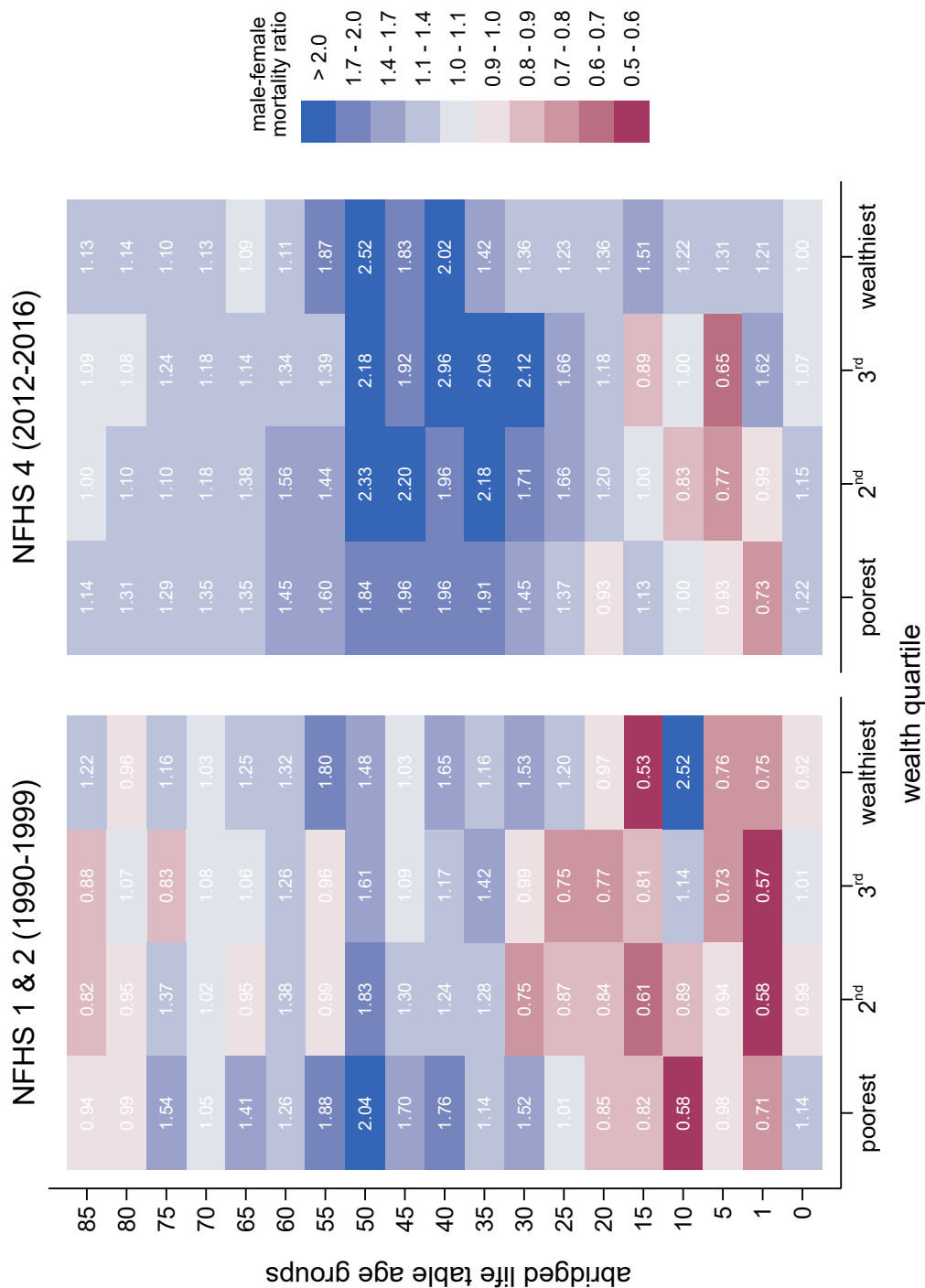
Note: Mortality ratios are calculated as male age-specific mortality rates in an age-group, wealth quartile, and round divided by corresponding female age-specific mortality rates. Ratios higher than 1 show male survival disadvantage. Ratios less than 1 denote female survival disadvantage.

3.7 Supplementary Appendix

Table 3A–1: Person-years and deaths, by sex, residence, and household wealth quartile

| Characteristics			NFHS-1 (1992-93)		NFHS-2 (1998-99)		NFHS-4 (2015-16)	
sex	sample	quartile	person-years	deaths	person-years	deaths	person-years	deaths
female	overall	poorest	150,116	1,670	150,062	1,658	1,105,286	8,492
female	overall	2nd quartile	147,375	1,698	164,296	1,570	1,205,110	8,481
female	overall	3rd quartile	146,904	1,355	165,984	1,334	973,585	6,258
female	overall	wealthiest	162,760	950	166,428	1,120	1,124,684	6,577
female	urban	poorest	36,697	339	37,434	294	331,269	2,121
female	urban	2nd quartile	38,210	285	41,713	331	358,336	2,234
female	urban	3rd quartile	42,951	256	44,169	293	385,670	2,111
female	urban	wealthiest	43,294	190	44,659	268	370,456	2,030
female	rural	poorest	136,564	1,529	104,651	1,164	685,865	5,316
female	rural	2nd quartile	108,604	1,269	123,098	1,311	737,098	5,521
female	rural	3rd quartile	113,076	1,135	123,584	1,071	746,855	5,213
female	rural	wealthiest	87,760	670	127,462	951	793,116	5,262
male	overall	poorest	153,370	2,012	153,918	1,951	1,098,174	11,219
male	overall	2nd quartile	159,472	1,747	175,634	1,769	1,217,446	11,017
male	overall	3rd quartile	159,140	1,390	178,056	1,521	1,008,503	8,048
male	overall	wealthiest	172,048	1,186	175,733	1,340	1,176,869	8,428
male	urban	poorest	39,921	484	41,234	439	360,041	3,140
male	urban	2nd quartile	43,246	372	45,076	343	372,186	2,826
male	urban	3rd quartile	45,943	290	47,772	350	400,726	2,937
male	urban	wealthiest	45,486	249	47,115	312	393,231	2,487
male	rural	poorest	139,506	1,814	105,834	1,390	664,529	7,367
male	rural	2nd quartile	116,496	1,253	130,574	1,373	738,516	6,815
male	rural	3rd quartile	120,588	1,129	131,666	1,239	756,654	6,629
male	rural	wealthiest	92,844	744	134,070	1,136	815,110	6,511
	exposure period		Jan1990-Aug1993		Jan1996-Dec1999		Jan2013-Nov2016	

All estimates weighted by national households weights. Wealth quartiles are estimated using a principal component analysis of assets owned by the household. Quartiles are estimated separately for rural, urban, and all households in each round. This implies, for example, that person years and deaths in the poorest quartile in urban areas in NFHS-4 are from households who are the poorest 25% *within* urban areas in NFHS-4.

Figure 3A–1: Comparison of household size in decennial censuses and NFHS surveys
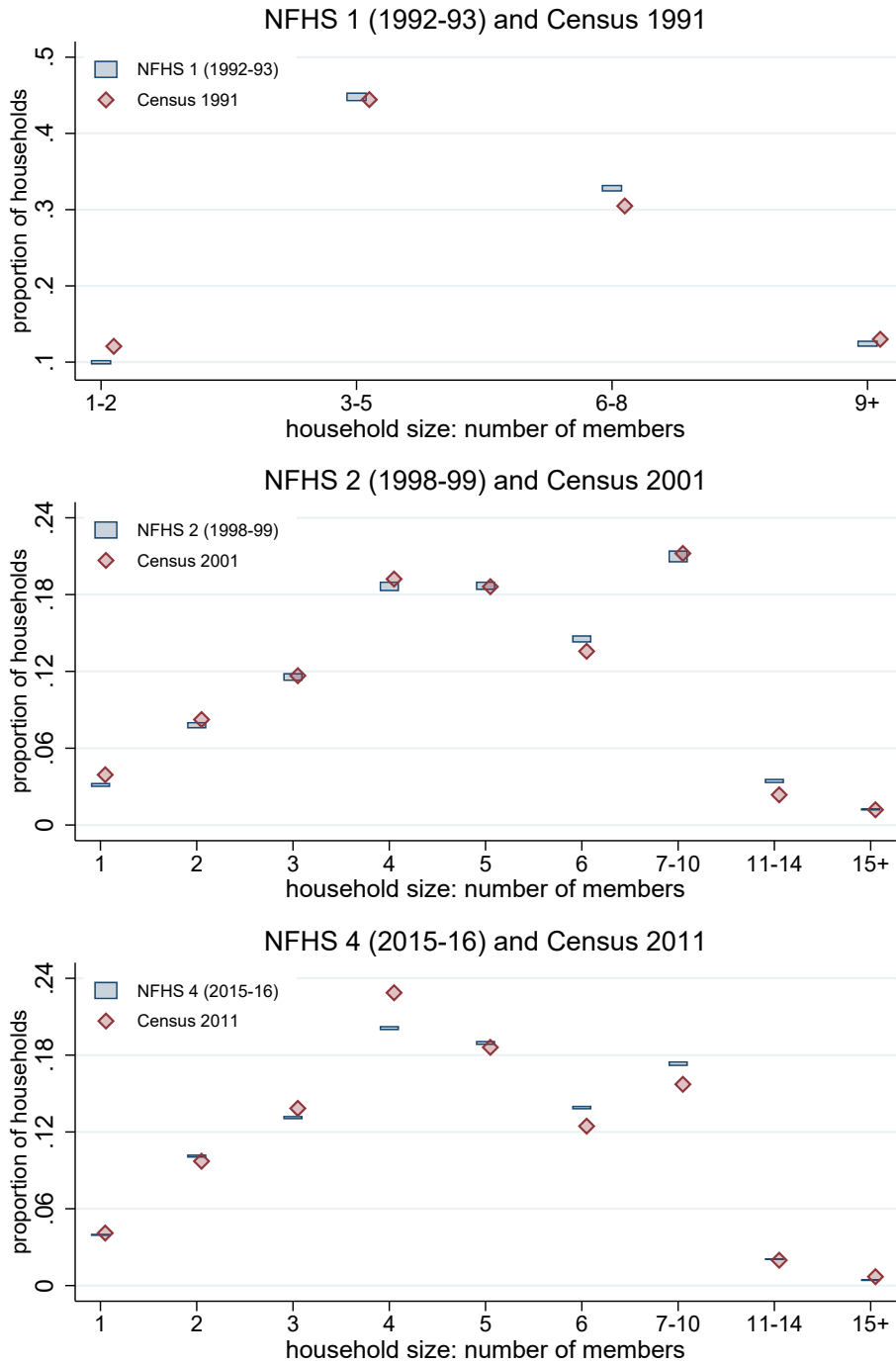
Note: The censuses combine household size categories. They do not release full household size distributions. NFHS household sizes are matched to household size categories available from the decennial censuses. The height of the blue rectangles represent 95% CIs, calculated to account for clustering of observations within primary sampling units.

Figure 3A–2: Comparison of age-specific mortality rates from NFHS and SRS

116
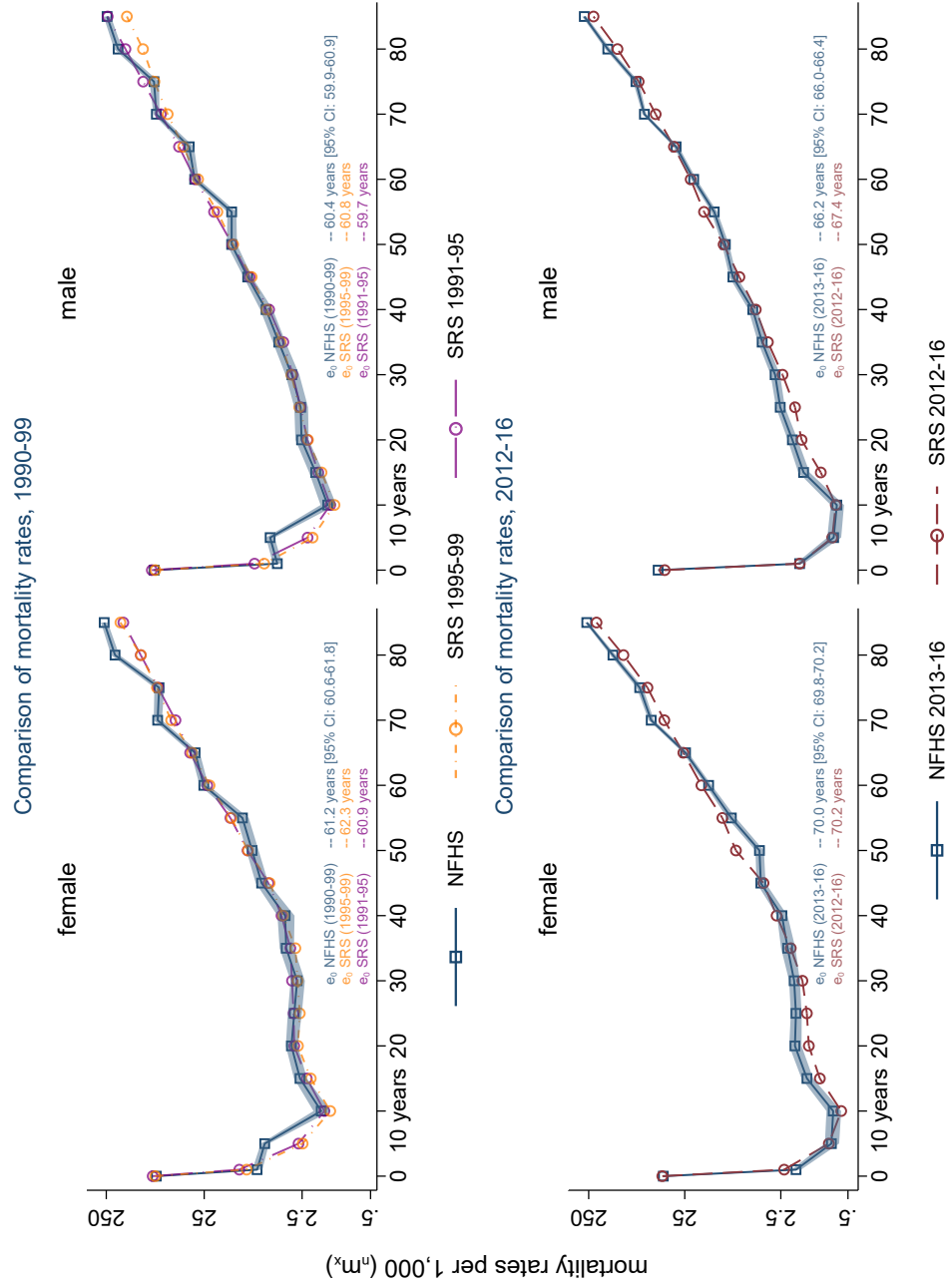
Figure 3A–3: Life expectancy at birth by household asset ventile, 2013-2016

The vertical height of each bar depicts the 95% confidence interval.
CIs were estimated using a cluster bootstrap approach with 100 replications.
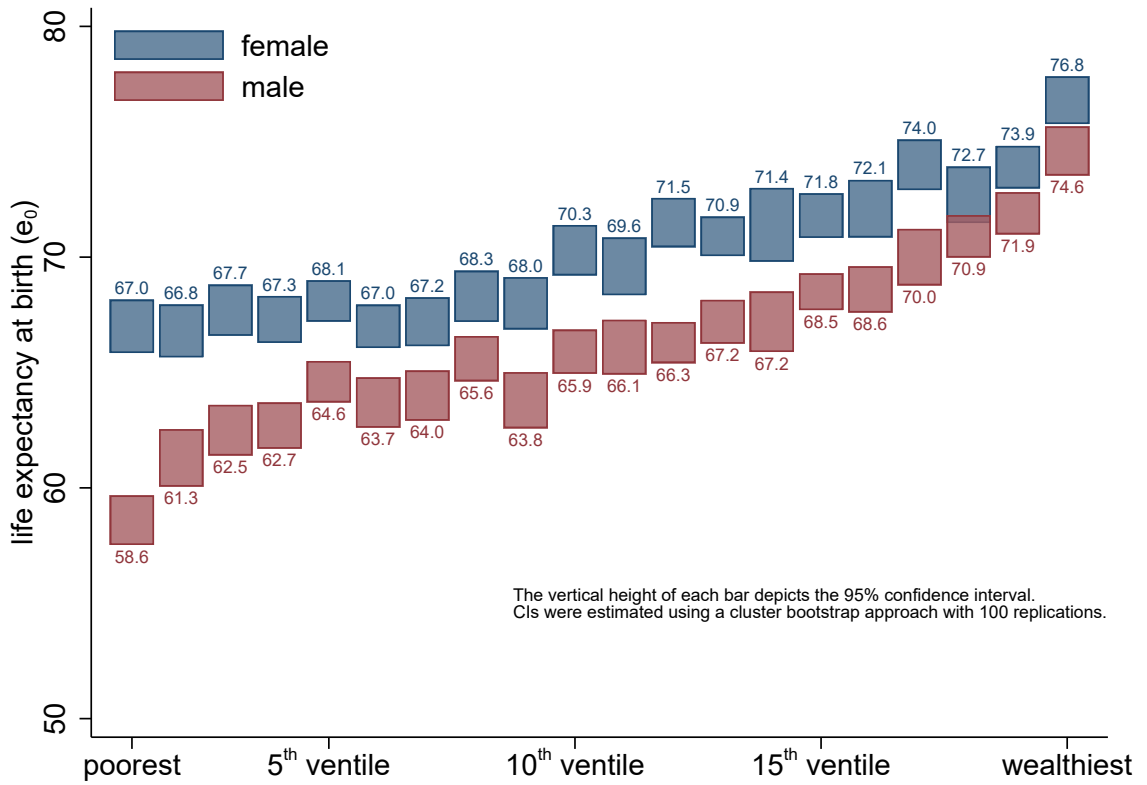
Figure 3A–4: Wealth disparities in life expectancy at birth
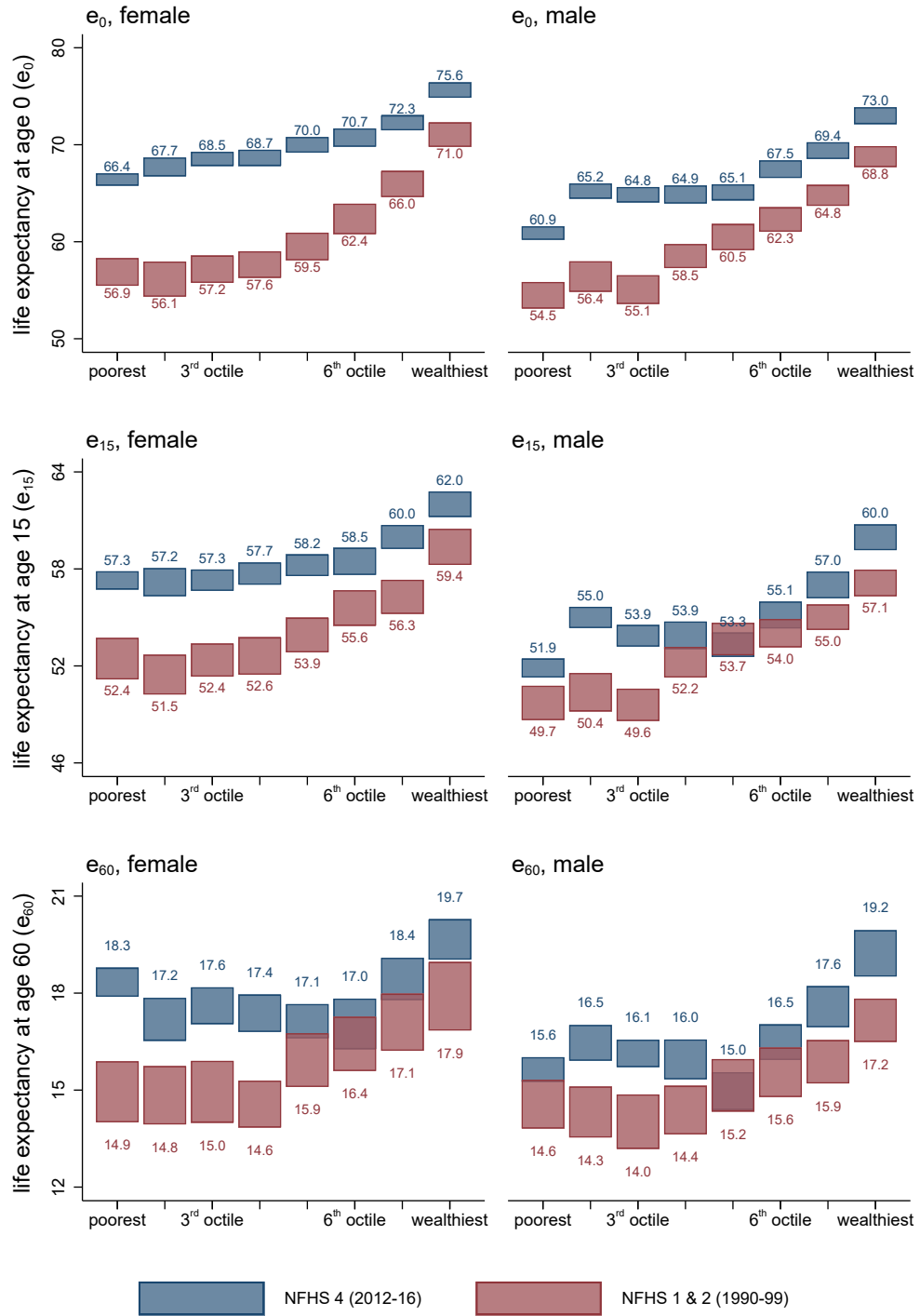
The vertical height of each bar depicts the 95% confidence interval.
CIs were estimated using a cluster bootstrap approach with 100 replications.

Figure 3A–5: Probabilities of death across the life course, 1990-2016
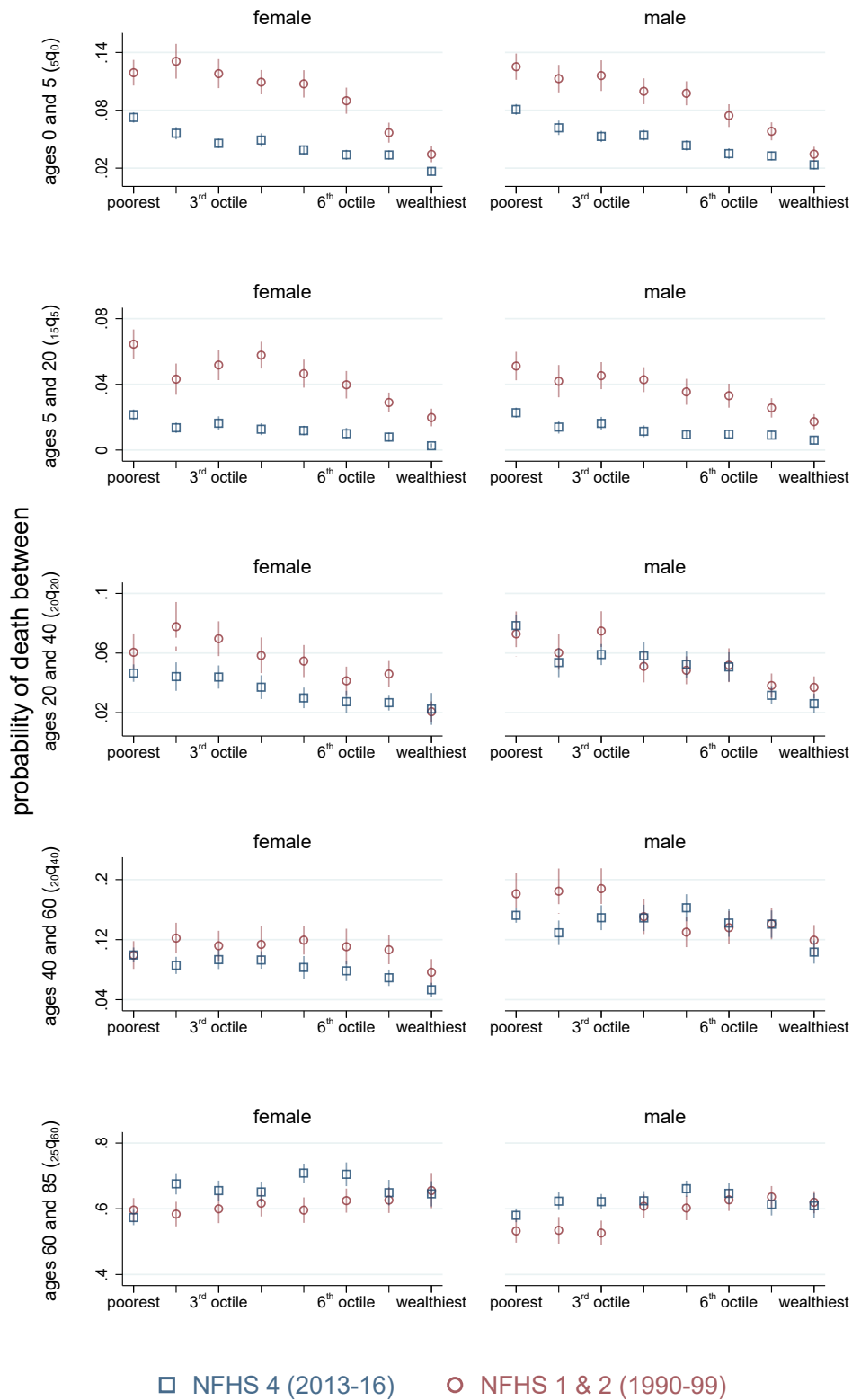
Figure 3A–6: Age-specific death rates by household wealth quartile
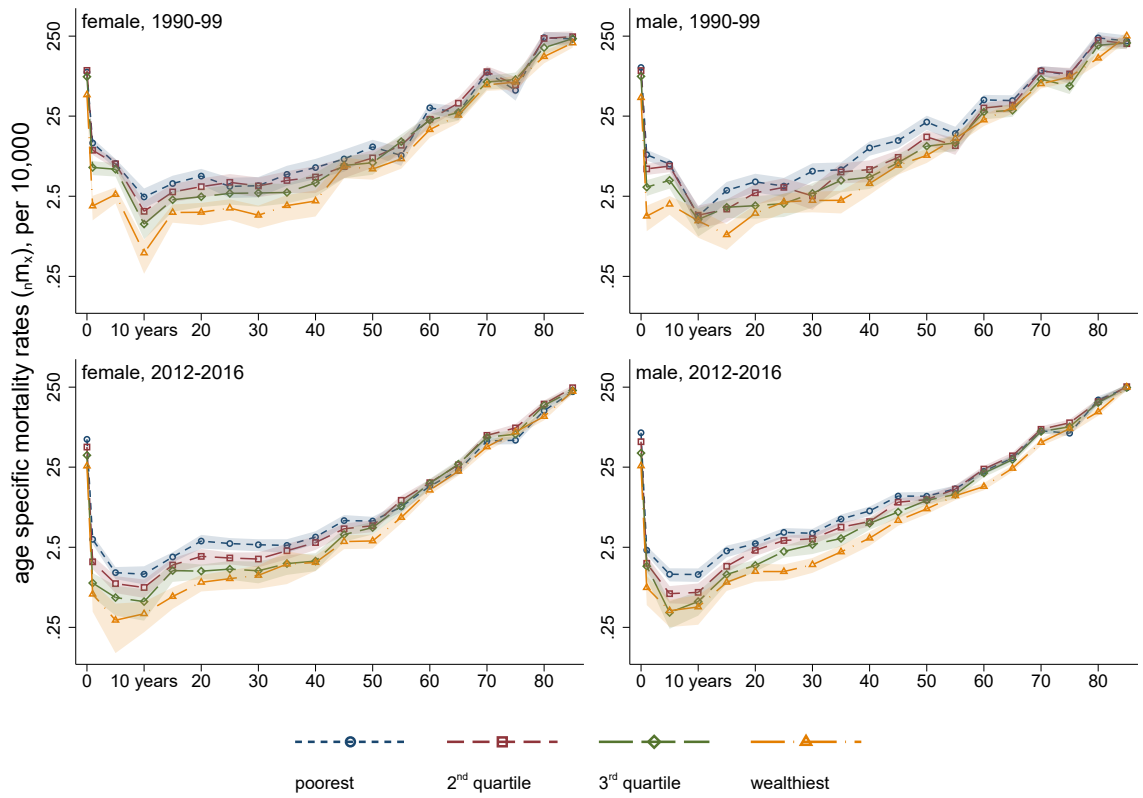
Estimates for 1990-99 are from NFHS 1 & 2. Estimates for 2013-2016 are from NFHS 4.

Figure 3A–7: Survivorship curves by household wealth quartile
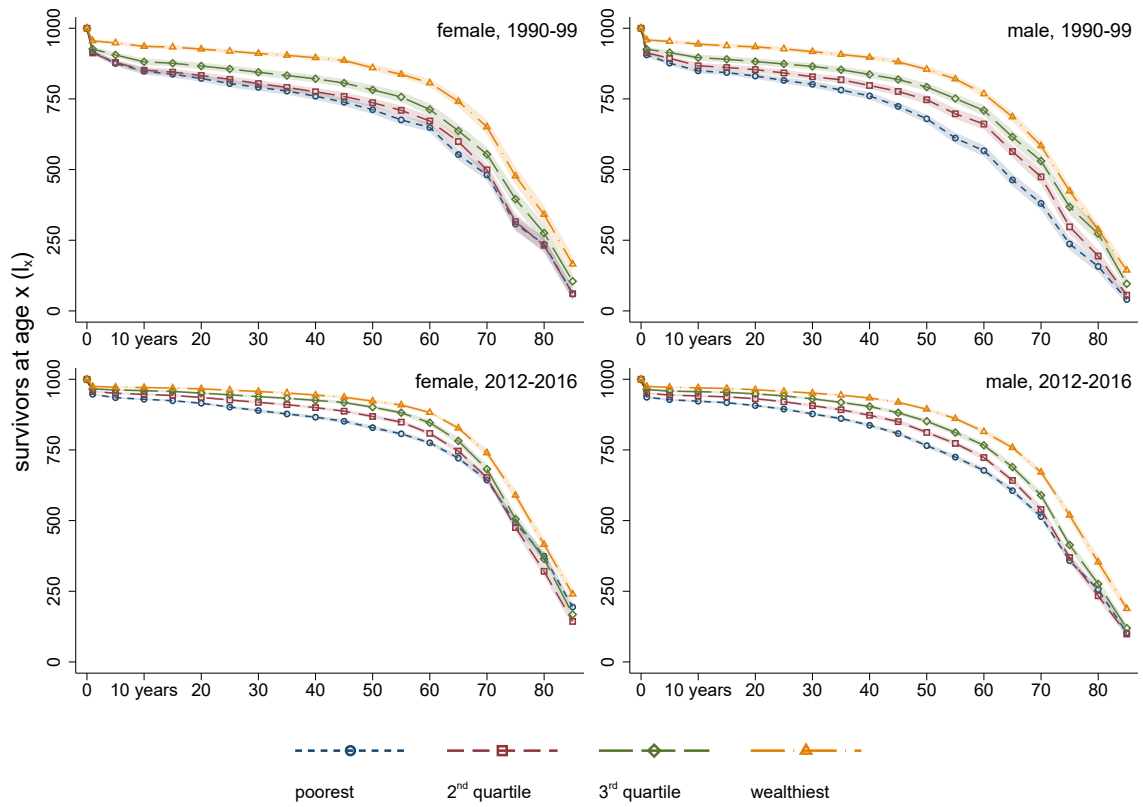
Figure 3A–8: Age heaping in deaths by wealth quartile
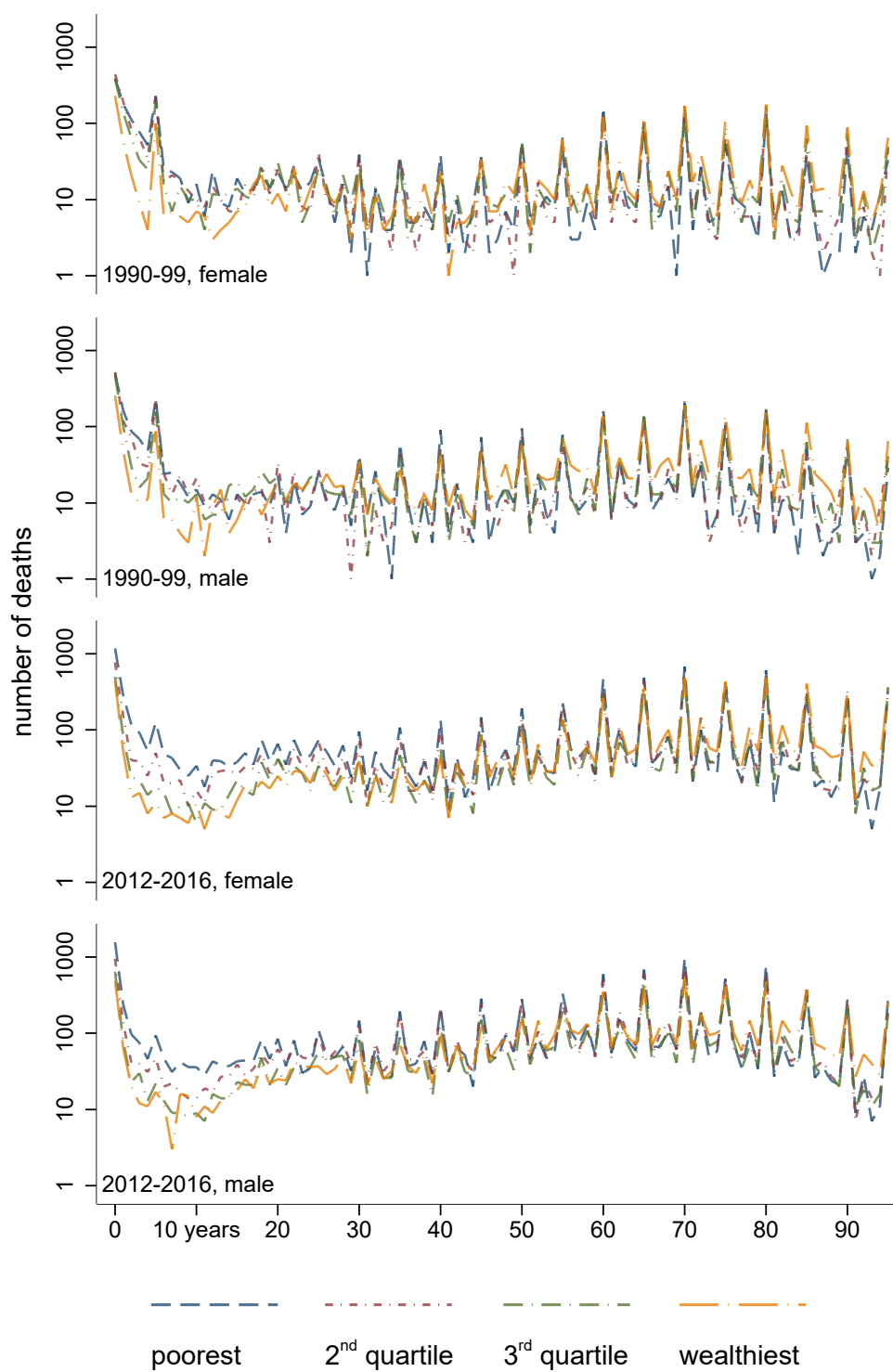
Figure 3A–9: Robustness: 95% CIs by number of bootstrap samples drawn, NFHS-4

The vertical height of each bar depicts the 95% confidence interval.
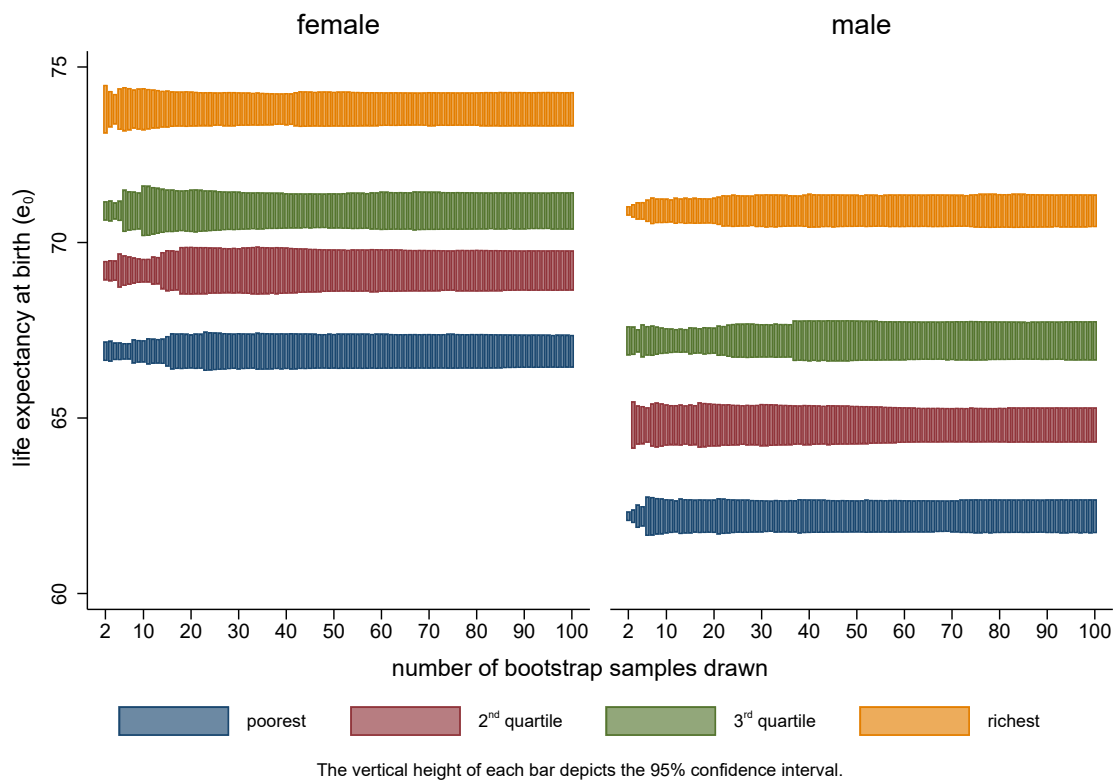
Figure 3A–10: Proportion of clusters which have multiple deaths in the same sex, age-group, and wealth octile
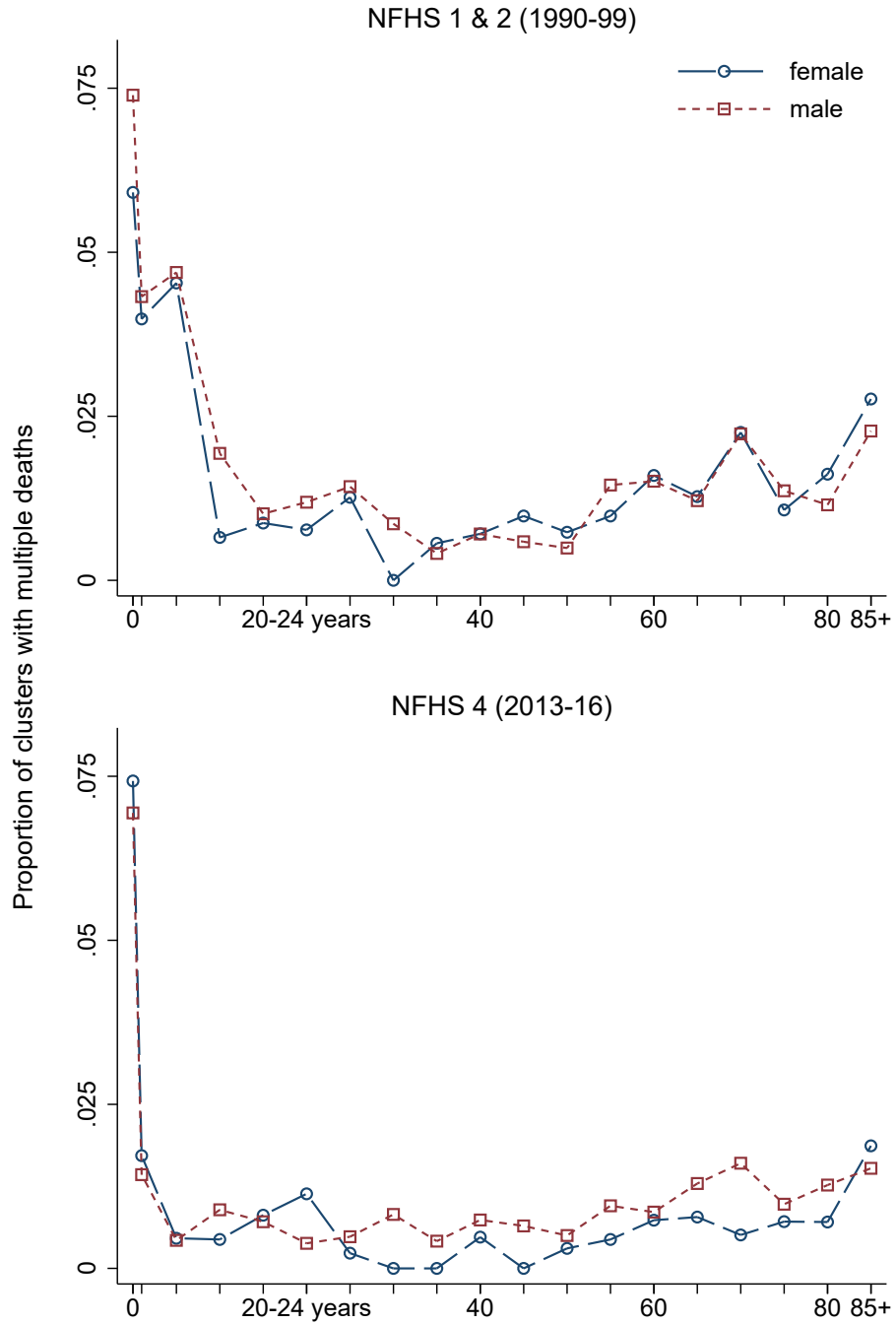
BIBLIOGRAPHY

Abdullah, S. et al. (2007). Patterns of age-specific mortality in children in endemic areas of sub-saharan africa. In *Defining and Defeating the Intolerable Burden of Malaria III: Progress and Perspectives: Supplement to Volume 77 (6) of American Journal of Tropical Medicine and Hygiene*. American Society of Tropical Medicine and Hygiene.

Adams, F. (1849). *The genuine works of Hippocrates: Volume I*. Sydenham society.

Agarwal, N., Aiyar, A., Bhattacharjee, A., Cummins, J., Gunadi, C., Singhania, D., Taylor, M., and Wigton-Jones, E. (2017). Month of birth and child height in 40 countries. *Economics Letters*, 157:10–13.

Alkire, S., Oldiges, C., and Kanagaratnam, U. (2021). Examining multidimensional poverty reduction in india 2005/6–2015/16: Insights and oversights of the headcount ratio. *World Development*, 142:105454.

Amaral, A. F. et al. (2017). Airflow obstruction and use of solid fuels for cooking or heating: Bold results. *American journal of respiratory and critical care medicine*, (ja).

Ambedkar, B. R. (1936). *Annihilation of Caste: With a reply to Mahatama Gandhi*. Government of Maharashtra: Writings and Speeches of BR Ambedkar (Compiled by Vasant Moon).

Ambedkar, B. R. (1989). *Untouchables or the Children of India's Ghetto*. Government of Maharashtra: Writings and Speeches of BR Ambedkar (Compiled by Vasant Moon).

Ambedkar, B. R. (1990). Dr. babasaheb ambedkar: Writings and speeches, vol. 14. *Government of Maharashtra, Bombay*.

Anand, K., Kumar, G., Kant, S., Kapoor, S., et al. (2000). Seasonality of births and possible factors influencing it in a rural area of haryana, india. *Indian pediatrics*, 37(3):306–311.

Apte, J. S. and Pant, P. (2019). Toward cleaner air for a billion indians. *Proceedings of the National Academy of Sciences*, 116(22):10614–10616.

Ariès, P. (1971). *Histoire des populations françaises*. Ed. du Seuil.

Arokiasamy, P., Bloom, D. E., Sekher, T., et al. (2021). Longitudinal ageing study in india (lasi): An investigation of health, economic, and social well-being of india's growing elderly population. Technical report.

Arokiasamy, P. et al. (2015). The impact of multimorbidity on adult physical and mental health in low-and middle-income countries: what does the study on global ageing and adult health (sage) reveal? *BMC medicine*, 13(1):178.

Aryal, S. et al. (2013). Copd and gender differences: an update. *Translational Research*, 162(4):208–218.

Asaria, M., Mazumdar, S., Chowdhury, S., Mazumdar, P., Mukhopadhyay, A., and Gupta, I. (2019). Socioeconomic inequality in life expectancy in india. *BMJ global health*, 4(3):e001445.

Attri, S. and Tyagi, A. (2010). Climate profile of india. *Environment Monitoring and Research Center, India Meteorology Department: New Delhi, India*.

Austad, S. (2011). Chapter 23-sex differences in longevity and aging. *Handbook of the Biology of Aging*, pages 479–495.

Austin, K. F. and Mejia, M. T. (2017). Household air pollution as a silent killer: women's status and solid fuel use in developing nations. *Population and Environment*, 39(1):1–25.

Balakrishnan, K. et al. (2014). Addressing the burden of disease attributable to air pollution in india: the need to integrate across household and ambient air pollution exposures. *Environmental health perspectives*, 122(1):A6.

Balarajan, Y., Selvaraj, S., and Subramanian, S. (2011). Health care and equity in india. *The Lancet*, 377(9764):505–515.

Barford, A., Dorling, D., Smith, G. D., and Shaw, M. (2006). Life expectancy: women now on top everywhere.

Barik, D., Desai, S., and Vanneman, R. (2018). Economic status and adult mortality in india: Is the relationship sensitive to choice of indicators? *World development*, 103:176–187.

Barker, D. J. (1995). Fetal origins of coronary heart disease. *Bmj*, 311(6998):171–174.

Baru, R., Acharya, A., Acharya, S., Kumar, A. S., and Nagaraj, K. (2010). Inequities in access to health services in india: caste, class and region. *Economic and political Weekly*, pages 49–58.

Bassani, D. G., Jha, P., Dhingra, N., and Kumar, R. (2010). Child mortality from solid-fuel use in india: a nationally-representative case-control study. *BMC public health*, 10(1):1–9.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific data*, 5(1):1–12.

Becker, S. (1981). Seasonality of deaths in matlab, bangladesh. *International journal of epidemiology*, 10(3):271–280.

Becker, S. and Weng, S. (1998). Seasonal patterns of deaths in matlab, bangladesh. *International journal of epidemiology*, 27(5):814–823.

Behrman, J. R. (1988a). Intrahousehold allocation of nutrients in rural india: Are boys favored? do parents exhibit inequality aversion? *Oxford Economic Papers*, 40(1):32–54.

Behrman, J. R. (1988b). Nutrition, health, birth order and seasonality: Intrahousehold allocation among children in rural india. *Journal of Development Economics*, 28(1):43–62.

Behrman, J. R. and Deolalikar, A. B. (1989). Agricultural wages in india: The role of health, nutrition, and seasonally. *Seasonal variability in third world agriculture: The consequences for food security*, pages 107–117.

Beltrán-Sánchez, H., Goldman, N., Pebley, A. R., and Morales, J. F. (2020). Calloused hands, shorter life? occupation and older-age survival in mexico. *Demographic Research*, 42:875–900.

Bengtsson, T., Dribe, M., and Helgertz, J. (2019). When did the health gradient emerge?: Social class and adult mortality in southern sweden, 1813-2015. *Demography*.

Bhalotra, S. R., Diaz-Cayeros, A., Miller, G., Miranda, A., and Venkataramani, A. S. Urban water disinfection and mortality decline in lower-income countries. *American Economic Journal: Economic Policy*.

Bhat, P. M., Preston, S. H., and Dyson, T. (1984). *Vital rates in India, 1961-1981*, volume 24. National Academies.

Bhatty, K. (2014). Review of elementary education policy in india: Has it upheld the constitutional objective of equality? *Economic and Political Weekly*, pages 100–107.

Bhishagratna, K. K. (1963). The sushruta samhita.[english translation based on the original sanskrit text.]. *Varanasi, Chowkhamba, Sanskrit Series*.

Boerma, T., Requejo, J., Victora, C. G., Amouzou, A., George, A., Agyepong, I., Barroso, C., Barros, A. J., Bhutta, Z. A., Black, R. E., et al. (2018). Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. *The Lancet*, 391(10129):1538–1548.

Bonjour, S. et al. (2013). Solid fuel use for household cooking: country and regional estimates for 1980-2010. *Environmental Health Perspectives (Online)*, 121(7):784.

Bosworth, B. (2018). Increasing disparities in mortality by socioeconomic status. *Annual review of public health*, 39:237–251.

Brown, P., Maher, Y. I., Balakrishnan, K., Fu, S. H., Kumar, R., Chakma, J., Menon, G., Dikshit, R., Dhaliwal, R., Rodriguez, P. S., et al. (2019). Mortality from particulate matter 2· 5 in india: National prospective proportional mortality study.

Burkart, K., Khan, M. H., Krämer, A., Breitner, S., Schneider, A., and Endlicher, W. R. (2011). Seasonal variations of all-cause and cause-specific mortality by age, gender, and socioeconomic condition in urban and rural areas of bangladesh. *International journal for equity in health*, 10(1):32.

Burnett, R. T. et al. (2014). An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental health perspectives*, 122(4):397.

Caldwell, J. C. (1993). Health transition: the cultural, social and behavioural determinants of health in the third world. *Social Science & Medicine*, 36(2):125–135.

Caldwell, J. C. (1994). How is greater maternal education translated into lower child mortality? *Health transition review*, 4(2):224–229.

Caldwell, J. C. and Caldwell, P. (1993). Roles of women, families, and communities in preventing illness and providing health services in developing countries. *The Epidemiological Transition: Policy and Planning Implications for Developing Countries. Edited by JN Gribble and SH Preston. National Academy Press, Washington, DC.*

Caldwell, J. C., Reddy, P. H., and Caldwell, P. (1988). *The causes of demographic change: Experimental research in South India*. University of Wisconsin Press.

Calverley, P. and Walker, P. (2003). Chronic obstructive pulmonary disease. *The Lancet*, 362(9389):1053 – 1061.

Canudas-Romo, V., Saikia, N., and Diamond-Smith, N. (2016). The contribution of age-specific mortality towards male and female life expectancy differentials in india and selected states, 1970-2013. *Asia-Pacific Population Journal*, 30(2):1–20.

Case, A. and Paxson, C. (2005). Sex differences in morbidity and mortality. *Demography*, 42(2):189–214.

Chafe, Z. A., Brauer, M., Klimont, Z., Van Dingenen, R., Mehta, S., Rao, S., Riahi, K., Dentener, F., and Smith, K. R. (2014). Household cooking with solid fuels contributes to ambient pm2. 5 air pollution and the burden of disease. *Environmental health perspectives*, 122(12):1314–1320.

Chalasani, S. and Rutstein, S. (2014). Household wealth and child health in india. *Population studies*, 68(1):15–41.

Chambers, R., Longhurst, R., Pacey, A., et al. (1981). *Seasonal dimensions to rural poverty*. Citeseer.

Chao, F., You, D., Pedersen, J., Hug, L., and Alkema, L. (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: a systematic assessment. *The Lancet Global Health*, 6(5):e535–e547.

Chapman, R. S. et al. (2005). Improvement in household stoves and risk of chronic obstructive pulmonary disease in xuanwei, china: retrospective cohort study. *Bmj*, 331(7524):1050.

Chapoto, A., Kirimi, L., and Kadiyala, S. (2012). Poverty and prime-age mortality in eastern and southern africa: Evidence from zambia and kenya. *World Development*, 40(9):1839–1853.

Chen, A., Oster, E., and Williams, H. (2016). Why is infant mortality higher in the united states than in europe? *American Economic Journal: Economic Policy*, 8(2):89–124.

Chen, L. C., Rahman, M., D'Souza, S., Chakraborty, J., and Sardar, A. (1983). Mortality impact of an mch-fp program in matlab, bangladesh. *Studies in family planning*, pages 199–209.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the united states, 2001-2014. *Jama*, 315(16):1750–1766.

Coffey, D. (2015a). Little women: Essays on maternal nutrition, social hierarchy, and health in india.

Coffey, D. (2015b). Prepregnancy body mass and weight gain during pregnancy in india and sub-saharan africa. *Proceedings of the National Academy of Sciences*, 112(11):3302–3307.

Coffey, D., Deshpande, A., Hammer, J., and Spears, D. (2019). Local social inequality, economic inequality, and disparities in child height in india. *Demography*, 56(4):1427–1452.

Coffey, D., Gupta, A., Hathi, P., Spears, D., Srivastav, N., and Vyas, S. (2017). Understanding open defecation in rural india: Untouchability, pollution, and latrine pits. *Economic and Political Weekly*, 52(1):59–66.

Committee on Population (2021). High and rising mortality rates among working-age adults.

Cotes, J. E. et al. (2009). *Lung function: physiology, measurement and application in medicine*. John Wiley and Sons.

Currie, G. P. (2010). *ABC of COPD*. John Wiley & Sons.

Currie, J. and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annu. Rev. Econ.*, 5(1):1–36.

Das Gupta, M. and Mari Bhat, P. (1997). Fertility decline and increased manifestation of sex bias in india. *Population studies*, 51(3):307–315.

Dave, M. et al. (2017). Household air pollution and lung function in indian adults: a cross-sectional study. *The International Journal of Tuberculosis and Lung Disease*, 21(6):702–704.

De Walque, D. and Filmer, D. (2011). *Trends and socioeconomic gradients in adult mortality around the developing world*. The World Bank.

Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health affairs*, 21(2):13–30.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Delaunay, V., Etard, J.-F., Préziosi, M.-P., Marra, A., and Simondon, F. (2001). Decline of infant and child mortality rates in rural senegal over a 37-year period (1963–1999). *International Journal of Epidemiology*, 30(6):1286–1293.

Devereux, S., Sabates-Wheeler, R., and Longhurst, R. (2013). *Seasonality, rural livelihoods and development*. Routledge.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Doblhammer, G. and Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences*, 98(5):2934–2939.

Dommaraju, P. (2015). One-person households in india. *Demographic Research*, 32:1239–1266.

Dorélien, A. (2019). The effects of in utero exposure to influenza on birth and infant outcomes in the us. *Population and development review*, 45(3):489.

Dorélien, A. M. (2016). Birth seasonality in sub-saharan africa. *Demographic Research*, 34:761–796.

Dowd, J. B., Albright, J., Raghunathan, T. E., Schoeni, R. F., LeClere, F., and Kaplan, G. A. (2011). Deeper and wider: income and mortality in the usa over three decades. *International journal of epidemiology*, 40(1):183–188.

Drevenstedt, G. L., Crimmins, E. M., Vasunilashorn, S., and Finch, C. E. (2008). The rise and fall of excess male infant mortality. *Proceedings of the National Academy of Sciences*, 105(13):5016–5021.

Drèze, J. (1990). Famine prevention in india. *The political economy of hunger*, 2:13–122.

Dreze, J. and Gazdar, H. (1997). Uttar pradesh: The burden of inertia. *Indian development: Selected regional perspectives*, page 33.

Drèze, J., Gupta, A., Parashar, S. A., and Sharma, K. (2021). Pauses and reversals of infant mortality decline in india in 2017 and 2018. *Forthcoming, Economic and Political Weekly*.

Drèze, J. and Khera, R. (2017). Recent social security initiatives in india. *World Development*, 98:555–572.

Drèze, J. and Sen, A. (2013). *An uncertain glory: India and its contradictions*. Princeton University Press.

Dyson, T. (1991). On the demography of south asian famines part ii. *Population studies*, 45(2):279–297.

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

El Arifeen, S., Hill, K., Ahsan, K. Z., Jamil, K., Nahar, Q., and Streatfield, P. K. (2014). Maternal mortality in bangladesh: a countdown to 2015 country case study. *The Lancet*, 384(9951):1366–1374.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual review of sociology*, 35:553–572.

Elo, I. T., Hendi, A. S., Ho, J. Y., Vierboom, Y. C., and Preston, S. H. (2019). Trends in non-hispanic white mortality in the united states by metropolitan-nonmetropolitan status and region, 1990–2016. *Population and development review*, 45(3):549.

Elo, I. T. and Preston, S. H. (1992). Effects of early-life conditions on adult mortality: a review. *Population index*, pages 186–212.

Elo, I. T. and Preston, S. H. (1996). Educational differentials in mortality: United states, 1979–1985. *Social science & medicine*, 42(1):47–57.

Etard, J.-F., Le Hesran, J.-Y., Diallo, A., Diallo, J.-P., Ndiaye, J.-L., and Delaunay, V. (2004). Childhood mortality and probable causes of death using verbal autopsy in niakhar, senegal, 1989–2000. *International Journal of Epidemiology*, 33(6):1286–1292.

Ezzati, M. and Kammen, D. M. (2002). The health impacts of exposure to indoor air pollution from solid fuels in developing countries: knowledge, gaps, and data needs. *Environmental health perspectives*, 110(11):1057.

Fadel, S. A., Rasaily, R., Awasthi, S., Begum, R., Black, R. E., Gelband, H., Gerland, P., Kumar, R., Liu, L., Mathers, C., et al. (2017). Changes in cause-specific neonatal and 1–59-month child mortality in india from 2000 to 2015: a nationally representative survey. *The Lancet*, 390(10106):1972–1980.

Fairbrother, M. (2016). Externalities: why environmental sociology should bring them in. *Environmental Sociology*, 2(4):375–384.

Farrar, D. S., Awasthi, S., Fadel, S. A., Kumar, R., Sinha, A., Fu, S. H., Wahl, B., Morris, S. K., and Jha, P. (2019). Seasonal variation and etiologic inferences of childhood pneumonia and diarrhea mortality in india. *eLife*, 8.

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure

data—or tears: an application to educational enrollments in states of india. *Demography*, 38(1):115–132.

Fishman, E. (2015). Variance estimation for a complex life table quantity: Disease-free life expectancy.

Fisman, D. N. (2007). Seasonality of infectious diseases. *Annu. Rev. Public Health*, 28:127–143.

Forey, B. A. et al. (2011). Systematic review with meta-analysis of the epidemiological evidence relating smoking to copd, chronic bronchitis and emphysema. *BMC pulmonary medicine*, 11(1):1.

Fu, S. H., Gasparrini, A., Rodriguez, P. S., and Jha, P. (2018). Mortality attributable to hot and cold ambient temperatures in india: a nationally representative case-crossover study. *PLoS medicine*, 15(7):e1002619.

GBD Causes of Death Collaborators (2017). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the global burden of disease study 2016. *Lancet*, 390(10100):1151–1210.

Geruso, M. (2012). Black-white disparities in life expectancy: how much can the standard ses variables explain? *Demography*, 49(2):553–574.

Geruso, M. and Spears, D. (2018a). Heat, humidity, and infant mortality in the developing world. Technical report, National Bureau of Economic Research.

Geruso, M. and Spears, D. (2018b). Neighborhood sanitation and infant mortality. *American Economic Journal: Applied Economics*, 10(2):125–62.

Godtfredsen, N. et al. (2008). Copd-related morbidity and mortality after smoking cessation: status of the evidence. *European Respiratory Journal*, 32(4):844–853.

GOLD (2016). Global strategy for the diagnosis, management and prevention of copd. Technical report, Global Initiative for Chronic Obstructive Lung Disease - GOLD.

Goldman, N., Pebley, A. R., and Lord, G. (1984). Calculation of life tables from survey data: A technical note. *Demography*, 21(4):647–653.

Gomes, M., Begum, R., Sati, P., Dikshit, R., Gupta, P. C., Kumar, R., Sheth, J., Habib, A., and Jha, P. (2017). Nationwide mortality studies to quantify causes of death: relevant lessons from india's million death study. *Health Affairs*, 36(11):1887–1895.

Government of India (1962). The national sample survey. seventh round - october 1953–march 1954. vital rates. Report np. 54, The Cabinet Secretariat, Government of India.

Government of India (2011a). Census of india 2011: Household listing operations. *Office of the Registrar General Census Commissioner, India*.

Government of India (2011b). Census of india 2011: Houselisting operations. *Housing and Houselisting Operations*.

Government of India (2020). Abridged Life Tables based on the Sample Registration System, 2014-2018. Technical report, Government of India Office of the Registrar General and Census Commissioner of India.

Gruner, O. C. (1984). A treatise on the canon of medicine of avicenna. *Incorporating a Translation of the First Book*.

Guillot, M. (2002). The dynamics of the population sex ratio in india, 1971–96. *Population studies*, 56(1):51–63.

Guillot, M., Gerland, P., Pelletier, F., and Saabneh, A. (2012). Child mortality estimation: a global overview of infant and child mortality age patterns in light of new empirical data. *PLoS medicine*, 9(8):e1001299.

Gupta, A. (2013). Old-age pension scheme in jharkhand and chhattisgarh. *Economic and Political Weekly*, pages 54–59.

Gupta, A. (2019). Where there is smoke: Solid fuel externalities, gender, and adult respiratory health in india. *Population and Environment*, 41(1):32–51.

Gupta, A. and Spears, D. (2017). Health externalities of india's expansion of coal plants: Evidence from a national panel of 40,000 households. *Journal of environmental economics and management*, 86:262–276.

Gupta, A. and Sudharsanan, N. (2020). Differences in mortality between india's social groups are large, exist across the life course, and persist over time. Technical report.

Gupta, A., Vyas, S., Hathi, P., Khalid, N., Srivastav, N., Spears, D., and Coffey, D. (2019). Persistence of solid fuel use despite increases in lpg ownership: New survey evidence from rural north india.

Gupta, A., Vyas, S., Hathi, P., Khalid, N., Srivastav, N., Spears, D., and Coffey, D. (2020). Persistence of solid fuel use in rural north india. *Economic & Political Weekly*, 55(3):55.

Guttikunda, S. (2017). It's about time we got smarter about monitoring our air pollution. *The Wire*.

Guttikunda, S. K. and Calori, G. (2013). A gis based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in delhi, india. *Atmospheric Environment*, 67:101–111.

Gwatkin, D. R. (1993). Distributional implications of alternative strategic responses to the demographic-epidemiological transition—an initial inquiry. *The Epidemiological Transition*.

Gwatkin, D. R., Rutstein, S., Johnson, K., Suliman, E., Wagstaff, A., and Amouzou, A. (2007). Socio-economic difference in health, nutrition and population within developing countries. *Washington, DC: World Bank.*

Heligman, L., Chen, N., and Babakol, O. (1993). Shifts in the structure of population and deaths in less developed regions. *The Epidemiological Transition: Policy and Planning Implications for Developing Countries*, page 9.

Heuveline, P., Guillot, M., and Gwatkin, D. R. (2002). The uneven tides of the health transition. *Social science & medicine*, 55(2):313–322.

Hill, K. (1991). Approaches to the measurement of childhood mortality: a comparative review. *Population Index*, pages 368–382.

Hill, K. (1995). Age patterns of child mortality in the developing world. *Population bulletin of the United Nations*, (39):112–132.

Hill, K., El Arifeen, S., Koenig, M., Al-Sabir, A., Jamil, K., and Raggers, H. (2006). How should we measure maternal mortality in the developing world? a comparison of household deaths and sibling history approaches. *Bulletin of the World Health Organization*, 84:173–180.

Hill, K., You, D., Inoue, M., and Oestergaard, M. Z. (2012). Child mortality estimation: accelerated progress in reducing global child mortality, 1990–2010. *PLoS medicine*, 9(8):e1001303.

Himanshu (2019). Inequality in india: A review of levels and trends.

Hirway, I. and Jose, S. (2011). Understanding women's work using time-use statistics: The case of india. *Feminist Economics*, 17(4):67–92.

Ho, T. and Noymer, A. (2017). Summertime, and the livin'is easy: Winter and summer pseudoseasonal life expectancy in the united states. *Demographic Research*, 37:1445–1476.

Holguin, F. et al. (2005). Comorbidity and mortality in copd-related hospitalizations in the united states, 1979 to 2001. *Chest Journal*, 128(4).

Horiuchi, S. (1999). Epidemiological transitions in human history. *Health and mortality: Issues of global concern*, pages 54–71.

Hu, G. et al. (2010). Risk of copd from exposure to biomass smoke: a metaanalysis. *Chest Journal*, 138(1):20–31.

IIPS (1995). National family health survey (nfhs-1), 1992–93. Survey report, International Institute for Population Sciences (IIPS), Mumbai, India.

IIPS and ICF (2014). National family health survey (nfhs-4), 2015–16. Interviewer manual, International Institute for Population Sciences (IIPS), Mumbai, India.

IIPS and ICF (2017). National family health survey (nfhs-4), 2015–16. Survey report, International Institute for Population Sciences (IIPS), Mumbai, India.

IIPS and ORC Macro (2001). National family health survey (nfhs-2), 1998–99. Survey report, International Institute for Population Sciences (IIPS), Mumbai, India.

Jagnoor, J., Suraweera, W., Keay, L., Ivers, R. Q., Thakur, J., Jha, P., Collaborators, M. D. S., et al. (2012). Unintentional injury mortality in india, 2005: nationally representative mortality survey of 1.1 million homes. *BMC public health*, 12(1):487.

Jeuland, M. et al. (2015). The economics of household air pollution. *Annu. Rev. Resour. Econ.*, 7(1):81–108.

Jha, P., Gajalakshmi, V., Gupta, P. C., Kumar, R., Mony, P., Dhingra, N., Peto, R., Collaborators, R.-C. P. S., et al. (2005). Prospective study of one million deaths in india: rationale, design, and validation results. *PLoS Med*, 3(2):e18.

Jha, P., Ranson, M. K., Nguyen, S. N., and Yach, D. (2002). Estimates of global and regional smoking prevalence in 1995, by age and sex. *American journal of public health*, 92(6):1002–1006.

Johnson, P. et al. (2011). Prevalence of chronic obstructive pulmonary disease in rural women of tamilnadu: implications for refining disease burden assessments attributable to household biomass combustion. *Global health action*, 4(1):7226.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., Alam, M., Sack, R. B., and Colwell, R. (2013). Environmental factors influencing epidemic cholera. *The American journal of tropical medicine and hygiene*, 89(3):597–607.

Kamble, B. (1990). Jina amucha. *Pune: Mansanman Prakashan. 2007 The Prisons We Broke*.

Kampe, E. O. i., Müller, O., Sie, A., and Becher, H. (2015). Seasonal and temporal trends in all-cause and malaria mortality in rural burkina faso, 1998–2007. *Malaria journal*, 14(1):300.

Kashyap, R. and Behrman, J. (2020). Gender discrimination and excess female under-5 mortality in india: A new perspective using mixed-sex twins. *Demography*, 57(6):2143–2167.

Kishore, A. and Spears, D. (2014). Having a son promotes clean cooking fuel use in urban india: Women's status and son preference. *Economic Development and Cultural Change*, 62(4):673–699.

Kitagawa, E. M. and Hauser, P. M. (1973). *Differential mortality in the United States*. Harvard University Press.

Knodel, J. (1983). Seasonal variation in infant mortality: an approach with applications. In *Annales de démographie historique*, pages 208–230. JSTOR.

Kowal, P. et al. (2012). Data resource profile: the world health organization study on global ageing and adult health (sage). *International journal of epidemiology*, 41(6):1639–1649.

Krieger, N. (2011). *Epidemiology and the people's health: theory and context*. Oxford University Press.

Krishnan, A. et al. (2011). Adult mortality surveillance by routine health workers using a short verbal autopsy tool in rural north india. *Journal of Epidemiology & Community Health*.

Krueger, P. M. and Burgard, S. A. (2011). Work, occupation, income, and mortality. In *International handbook of adult mortality*, pages 263–288. Springer.

Kumar, S., Molitor, R., and Vollmer, S. (2016). Drought and early child health in rural india. *Population and Development Review*, pages 53–68.

Kumar, V., Shearer, J., Kumar, A., and Darmstadt, G. (2009). Neonatal hypothermia in low resource settings: a review. *Journal of Perinatology*, 29(6):401.

Kunst, A., Looman, C., and Mackenbach, J. (1991). The decline in winter excess mortality in the netherlands. *International Journal of Epidemiology*, 20(4):971–977.

Kurmi, O. P. et al. (2010). Copd and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax*, 65(3):221–228.

Kurmi, O. P. et al. (2012). Indoor air pollution and the lung in low-and medium-income countries. *European Respiratory Journal*, 40(1):239–254.

Laaksonen, M., Tarkiainen, L., and Martikainen, P. (2009). Housing wealth and mortality: a register linkage study of the finnish population. *Social science & medicine*, 69(5):754–760.

Larsen, A. F., Headey, D., and Masters, W. A. (2019). Misreporting month of birth: Diagnosis and implications for research on nutrition and early childhood in developing countries. *Demography*, 56(2):707–728.

Link, B. G. and Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of health and social behavior*, pages 80–94.

Lu, C., Cuartas, J., Fink, G., McCoy, D., Liu, K., Li, Z., Daelmans, B., and Richter, L. (2020). Inequalities in early childhood care and development in low/middle-income countries: 2010–2018. *BMJ global health*, 5(2).

Luo, W. and Xie, Y. (2014). Socio-economic disparities in mortality among the elderly in china. *Population studies*, 68(3):305–320.

Luo, Y., Zhang, Z., and Gu, D. (2015). Education and mortality among older adults in china. *Social Science & Medicine*, 127:134–142.

Mackenbach, J. P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemström, Ö., Valkonen, T., and Kunst, A. E. (2003). Widening socioeconomic inequalities in mortality in six western european countries. *International journal of epidemiology*, 32(5):830–837.

Manesh, A. O., Sheldon, T. A., Pickett, K. E., and Carr-Hill, R. (2008). Accuracy of child morbidity data in demographic and health surveys. *International journal of epidemiology*, 37(1):194–200.

Mannino, D. M. and Buist, A. S. (2007). Global burden of copd: risk factors, prevalence, and future trends. *The Lancet*, 370(9589):765–773.

Marmot, M. (2015). *The Health Gap: The Challenge of an Unequal World*. Bloomsbury Publishing.

Marmot, M. G., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., and Smith, G. D. (1991). Health inequalities among british civil servants: the whitehall ii study. *The Lancet*, 337(8754):1387–1393.

Masquelier, B., Eaton, J. W., Gerland, P., Pelletier, F., and Mutai, K. K. (2017). Age patterns and sex ratios of adult mortality in countries with high hiv prevalence. *Aids*, 31(1):S77–S85.

McGovern, M. E. and Canning, D. (2015). Vaccination and all-cause child mortality from 1985 to 2011: global evidence from the demographic and health surveys. *American journal of epidemiology*, 182(9):791–798.

McKay, A. J. et al. (2012). Prevalence of copd in india: a systematic review. *Primary Care Respiratory Journal*, 21:313–321.

Meslé, F. and Vallin, J. (2011). Historical trends in mortality. In *International handbook of adult mortality*, pages 9–47. Springer.

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., De Savigny, D., Lozano, R., and Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, 386(10001):1395–1406.

Million Death Study Collaborators and others (2010). Causes of neonatal and child mortality in india: a nationally representative mortality survey. *The Lancet*, 376(9755):1853–1860.

Mishra, S. et al. (2016). Trends in bidi and cigarette smoking in india from 1998 to 2015, by age, gender and education. *BMJ global health*, 1(1).

Mishra, V., Smith, K. R., and Retherford, R. D. (2005). Effects of cooking smoke and environmental tobacco smoke on acute respiratory infections in young indian children. *Population and Environment*, 26(5):375–396.

Mobarak, M. et al. (2012). Low demand for nontraditional cookstove technologies. *Proceedings of the National Academy of Sciences*, 109(27):10815–10820.

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1):67–82.

Moschovis, P. P. et al. (2015). The association of household air pollution and smoking with risk of obstructive lung disease among adults in india. In *A105. COPD EPIDEMIOLOGY: GLOBAL PERSPECTIVE*, pages A6353–A6353. Am Thoracic Soc.

Mosley, W. H. and Gray, R. (1993). Childhood precursors of adult morbidity and mortality in developing countries: implications for health programs. *The epidemiological transition: Policy and planning implications for developing countries*, pages 69–100.

Moultrie, T. A., Dorrington, R., Hill, A., Hill, K., Timæus, I., and Zaba, B. (2013). *Tools for demographic estimation*. International Union for the Scientific Study of Population.

Muhuri, P. K. (1996). Estimating seasonality effects on child mortality in matlab, bangladesh. *Demography*, 33(1):98–110.

Murthi, M., Guio, A.-C., and Dreze, J. (1995). Mortality, fertility, and gender bias in india: A district-level analysis. *Population and development review*, pages 745–782.

Mutisya, M., Orindi, B., Emina, J., Zulu, E., and Ye, Y. (2010). Is mortality among under-five children in nairobi slums seasonal? *Tropical medicine & international health*, 15(1):132–139.

Narayan, K. V. and Ali, M. K. (2013). Commentary: Shielding against a future inferno: the not-so-problematic discourse on socioeconomic status and cardiovascular health in india. *International journal of epidemiology*, 42(5):1426–1429.

National Academies of Sciences (2015). *The growing gap in life expectancy by income: Implications for federal programs and policy responses*. National Academies Press.

Nehru, J. (1983). *Jawaharlal Nehru's speeches*, volume 2. Publications Division, Ministry of Information and Broadcasting, Government . . . .

Nikoi, C. A. and Odimegwu, C. (2013). The association between socioeconomic status and adult mortality in rural kwazulu-natal, south africa. *Oman medical journal*, 28(2):102.

Office of the Registrar General and Census Commissioner of India (2018). Abridged life tables based on the sample registration system, 2012-2016. Technical report, Government of India.

Onyeneho, N. G., Ozumba, B. C., and Subramanian, S. (2019). Determinants of childhood anemia in india. *Scientific reports*, 9(1):1–7.

OPHI and UNDP (2020). Global mpi 2020–charting pathways out of multidimensional poverty: Achieving the sdgs.

ORGI (2016). Compendium of India's Fertility and Mortality Indicators 1971-2013 based on the Sample Registration System (SRS). Report, Office of the Registrar Genegral and Census Commissioner of India, New Delhi, Government of India.

Orsini, F. et al. (2004). The oxford india premchand.

Pai, M., Correa, N., Mistry, N., and Jha, P. (2017). Reducing global tuberculosis deaths—time for india to step up. *The Lancet*, 389(10075):1174–1176.

Pandey, G. (2014). *Gorakh Pandey: Samagra Kavitayein*, volume 1. Jan Sanskriti Manch.

Parikh, J. et al. (1999). Indoor air pollution: a reflection on gender bias. *Economic and Political Weekly*, pages 539–544.

Pascual, M., Bouma, M. J., and Dobson, A. P. (2002). Cholera and climate: revisiting the quantitative evidence. *Microbes and Infection*, 4(2):237–245.

Patel, S. A., Cunningham, S. A., Tandon, N., and Narayan, K. V. (2019). Chronic diseases in india—ubiquitous across the socioeconomic spectrum. *JAMA network open*, 2(4):e190404–e190404.

Patel, V. et al. (2011). Chronic diseases and injuries in india. *The Lancet*, 377(9763):413–428.

Pauwels, R. A. and Rabe, K. F. (2004). Burden and clinical features of chronic obstructive pulmonary disease (copd). *The Lancet*, 364(9434):613–620.

Pedersen, J. and Liu, J. (2012). Child mortality estimation: appropriate time periods for child mortality estimates from full birth histories. *PLoS Medicine*, 9(8):e1001289.

Peel, J. L. et al. (2015). Are randomized trials necessary to advance epidemiologic research on household air pollution? *Current Epidemiology Reports*, 2(4):263–270.

Pellegrino, R. et al. (2005). Interpretative strategies for lung function tests. *European Respiratory Journal*, 26(5):948–968.

Pigou, A. (1924). *The economics of welfare*. Routledge.

Pitt, M. M. and Sigle, W. (1998). *Seasonality, weather shocks and the timing of births and child mortality in Senegal*. Brown University, Population Studies and Training Center Providence, RI.

Po, J. Y. et al. (2011). Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. *Thorax*, 66(3):232–239.

Po, J. Y. and Subramanian, S. (2011). Mortality burden and socioeconomic status in india. *PLoS One*, 6(2):e16844.

Pollitt, E., Gorman, K. S., Engle, P. L., Martorell, R., Rivera, J., Wachs, T. D., and Scrimshaw, N. S. (1993). Early supplementary feeding and cognition: effects over two decades. *Monographs of the Society for Research in Child Development*, pages i–118.

Pope III, C. A. et al. (2015). Health benefits of air pollution abatement policy: role of the shape of the concentration–response function. *Journal of the Air & Waste Management Association*, 65(5):516–522.

Prescott, E. et al. (1995). Chronic mucus hypersecretion in copd and death from pulmonary infection. *European Respiratory Journal*, 8(8):1333–1338.

Preston, S., Heuveline, P., and Guillot, M. (2000). Demography: Measure and modeling population processes.

Preston, S., Patrick, H., and Michel, G. (2001). Demography: measuring and modeling population processes. *MA: Blackwell Publishing*.

Preston, S. H., Elo, I. T., and Stewart, Q. (1999). Effects of age misreporting on mortality estimates at older ages. *Population studies*, 53(2):165–177.

Preston, S. H. and Haines, M. R. (2014). *Fatal years: Child mortality in late nineteenth-century America*. Princeton University Press.

Preston, S. H., Hill, M. E., and Drevenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among african–americans. *Social science & medicine*, 47(9):1231–1246.

Preston, S. H. and Taubman, P. (1994). Socioeconomic differences in adult mortality and health status. *Demography of aging*, 1:279–318.

Rai, S. K., Kant, S., Srivastava, R., Gupta, P., Misra, P., Pandav, C. S., and Singh, A. K. (2017). Causes of and contributors to infant mortality in a rural community of north india: evidence from verbal and social autopsy. *BMJ open*, 7(8):e012856.

Ramakrishnan, A., Kalkuhl, M., Ahmad, S., and Creutzig, F. (2020). Keeping up with the patels: Conspicuous consumption drives the adoption of cars and appliances in india. *Energy Research & Social Science*, 70:101742.

Rao, S., Kanade, A. N., Yajnik, C. S., and Fall, C. H. (2009). Seasonality in maternal intake and activity influence offspring's birth size among rural indian mothers—pune maternal nutrition study. *International Journal of Epidemiology*, 38(4):1094–1103.

Rau, R. (2006a). The impact of social factors on excess winter mortality in denmark. *Seasonality in Human Mortality: A Demographic Approach*, pages 125–168.

Rau, R. (2006b). *Seasonality in human mortality: A demographic approach*. Springer Science & Business Media.

Ravallion, M. (2008). Global poverty and inequality: a review of the evidence.

Rayco-Solon, P., Moore, S. E., Fulford, A. J., and Prentice, A. M. (2004). Fifty-year mortality trends in three rural african villages. *Tropical medicine & international health*, 9(11):1151–1160.

Reed, M. N. (2020). The labor force participation of indian women before and after widowhood. *Demographic Research*, 43:673–706.

Rosero-Bixby, L. and Dow, W. H. (2009). Surprising ses gradients in mortality, health, and biomarkers in a latin american population of adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 64(1):105–117.

Roy, A. (2002). *The algebra of infinite justice*. Penguin Books India.

Royston, P. et al. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, 25(1):127–141.

Rumisha, S., Smith, T., Abdulla, S., Masanja, H., and Vounatsou, P. (2013). Assessing seasonal variations and age patterns in mortality during the first year of life in tanzania. *Acta tropica*, 126(1):28–36.

Saikia, N., Jasilionis, D., Ram, F., and Shkolnikov, V. M. (2011). Trends and geographic differentials in mortality under age 60 in india. *Population studies*, 65(1):73–89.

Saikia, N. and Ram, F. (2010). Determinants of adult mortality in india. *Asian Population Studies*, 6(2):153–171.

Saikia, N., Singh, A., Jasilionis, D., and Ram, F. (2013). Explaining the rural–urban gap in infant mortality in india. *Demographic Research*, 29:473–506.

Sakamoto-Momiyama, M. (1978). Changes in the seasonality of human mortality: a medico-geographical study. *Social Science and Medicine*, 12(1D):29–42.

Salvi, S. S. and Barnes, P. J. (2009). Chronic obstructive pulmonary disease in non-smokers. *The lancet*, 374(9691):733–743.

Salvi, S. S. and Barnes, P. J. (2010). Is exposure to biomass smoke the biggest risk factor for copd globally? *CHEST Journal*, 138(1):3–6.

Sana, A. et al. (2018). Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. *BMJ Open Respiratory Research*, 5(1):e000246.

Sankar, M., Neogi, S., Sharma, J., Chauhan, M., Srivastava, R., Prabhakar, P., Khera, A., Kumar, R., Zodpey, S., and Paul, V. (2016). State of newborn health in india. *Journal of Perinatology*, 36(s3):S3.

Sawyer, C. C. (2012). Child mortality estimation: estimating sex differences in childhood mortality since the 1970s. *PLoS Med*, 9(8):e1001287.

Schlüter, B.-S., Masquelier, B., Metcalf, C. J. E., and Rasoanomenjanahary, A. (2020). Long-term trends in seasonality of mortality in urban madagascar: the role of the epidemiological transition. *Global health action*, 13(1):1717411.

Sen, A. (1992). Missing women. *BMJ: British Medical Journal*, 304(6827):587.

Sen, A. (2001). *Development as freedom*. Oxford Paperbacks.

Sethi, N. and Deep, A. (2018). Pm's plan for free gas connections is failing its objective – as government had been warned it would. *Scroll.in*.

Sharma, K. (2018). Living with pain: women's everyday lives and health in rural bihar. the hindu centre for public policy. *Policy Report*, (23).

Shkolnikov, V. M., Andreev, E. M., Jdanov, D. A., Jasilionis, D., Kravdal, Ø., Vågerö, D., and Valkonen, T. (2012). Increasing absolute mortality disparities by education in finland, norway and sweden, 1971–2000. *J Epidemiol Community Health*, 66(4):372–378.

Sin, D. D. et al. (2006). Mortality in copd: role of comorbidities. *European Respiratory Journal*, 28(6):1245–1257.

Singer, J. D. and Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Singh, A., Kumar, K., and Singh, A. (2019). What explains the decline in neonatal mortality in india in the last three decades? evidence from three rounds of nfhs surveys. *Studies in family planning*, 50(4):337–355.

Smith, K. R. (2002). Indoor air pollution in developing countries: recommendations for research. *Indoor air*, 12(3):198–207.

Smith-Sivertsen, T. et al. (2009). Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: the respire randomized trial, guatemala. *American journal of epidemiology*, 170(2):211–220.

Spears, D. (2018). Exposure to open defecation can account for the indian enigma of child height. *Journal of Development Economics*, page 102277.

Spears, D. (2019a). *Air: Pollution, Climate Change and India's Choice Between Policy and Pretence*. Harper Collins.

Spears, D. (2019b). *AIR: pollution, climate change and indias choice between policy and pretence*. Harper Collins India.

Subramanian, S. (2020). What is happening to rural welfare, poverty, and inequality in India? *Ideas for India*.

Subramanian, S., Corsi, D. J., Subramanyam, M. A., and Davey Smith, G. (2013). Jumping the gun: the problematic discourse on socioeconomic status and cardiovascular health in india. *International journal of epidemiology*, 42(5):1410–1426.

Subramanian, S. et al. (2006). The mortality divide in india: the differential contributions of gender, caste, and standard of living across the life course. *American Journal of Public Health*, 96(5):818–825.

Subramanyam, M. A., Kawachi, I., Berkman, L. F., and Subramanian, S. (2010). Socioeconomic inequalities in childhood undernutrition in india: analyzing trends between 1992 and 2005. *PloS one*, 5(6):e11392.

Sudharsanan, N. (2019). The association between socioeconomic status and adult mortality in a developing country: Evidence from a nationally representative longitudinal survey of indonesian adults. *The Journals of Gerontology: Series B*, 74(3):484–495.

Sudharsanan, N. and Ho, J. Y. (2020). Rural–urban differences in adult life expectancy in indonesia: A parametric g-formula–based decomposition approach. *Epidemiology (Cambridge, Mass.)*, 31(3):393.

Sudharsanan, N., Zhang, Y., Payne, C. F., Dow, W., and Crimmins, E. (2020). Education and adult mortality in middle-income countries: Surprising gradients in six nationally-representative longitudinal surveys. *SSM-population health*, 12:100649.

Tabutin, D., Masquelier, B., Grieve, M., and Reeve, P. (2017). Mortality inequalities and trends in low-and middle-income countries, 1990-2015. *Population*, 72(2):221–296.

Timæus, I. M. (1991). Measurement of adult mortality in less developed countries: a comparative review. *Population index*, pages 552–568.

Tiwari, S., Srivastava, A. K., Bisht, D. S., Parmita, P., Srivastava, M. K., and Attri, S. (2013). Diurnal and seasonal variations of black carbon and pm2. 5 over new delhi, india: Influence of meteorology. *Atmospheric Research*, 125:50–62.

Torres-Duque, C. et al. (2008). Biomass fuels and respiratory diseases: a review of the evidence. *Proceedings of the American Thoracic Society*, 5(5):577–590.

UN IGME (2019). Estimates developed by the un inter-agency-group for child mortality estimation.

UN-IGME (2020). Levels and trends in child mortality: Report 2020. Technical report, United Nations Interagency Group on Child Mortality Estimation.

UN inter-agency group for child mortality estimation (2019). Levels and trends in child

mortality 2019, estimates developed by the united nations inter-agency group for child mortality estimation. *United Nations Children's Fund*.

UNICEF (2019). The state of the world's children 2019. Un report, UNICEF.

United Nations (2015). World population prospects: The 2015 revision. *United Nations Econ Soc Aff*, 33(2):1–66.

United Nations General Assembly (2015). Sustainable development goals. *SDGs Transform Our World*, 2030.

United Nations Population Division (2011). Mortality estimates from major sample surveys: towards the design of a database for the monitoring of mortality levels and trends. Technical report.

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology*, 50(7):3762–3772.

Veith, I. (2015). *The yellow emperor's classic of internal medicine*. Univ of California Press.

Venkataramani, A. S. and Fried, B. J. (2011). Effect of worldwide oil price fluctuations on biomass fuel use and child respiratory health: evidence from guatemala. *American journal of public health*, 101(9):1668–1674.

Venkataramani, A. S., O'Brien, R., Whitehorn, G. L., and Tsai, A. C. (2020). Economic influences on population health in the united states: Toward policymaking driven by data and evidence. *Plos Medicine*, 17(9):e1003319.

Vyas, S., Gupta, A., and Khalid, N. (2020). Gender, government intervention, and slow lpg adoption in rural north india. *SocArXiv. September*, 21.

Vyas, S., Hathi, P., and Gupta, A. (2021). Social disdavantage and life expectancy in nine indian states. Technical report.

Weinberg, C. R. (1995). How bad is categorization? *Epidemiology*, pages 345–347.

World Health Organisation (2013). *Study on Global AGEing and Adult Health: India*. WHO & IIPS.

World Health Organisation (2015). World health organisation fact-sheet on chronic obstructive pulmonary disease. *Geneva: World Health Organization*.

Wrigley, E. A., Davies, R. S., Schofield, R., and Oeppen, J. (1997). *English population history from family reconstitution 1580-1837*. Number 32. Cambridge University Press.

Zajacova, A. and Hummer, R. A. (2009). Gender differences in education effects on all-cause mortality for white and black adults in the united states. *Social science & medicine*, 69(4):529–537.