Essays in Development and Environmental Economics

Jan von der Goltz

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Science

COLUMBIA UNIVERSITY
2016

Abstract

Essays in Development and Environmental Economics

Jan von der Goltz

This dissertation discusses three questions of development and environmental economics. First, it assesses the impact of mineral mining on the health and wealth of households in local communities across 44 developing countries, using micro data. Secondly, it presents evidence from a randomized controlled trial on the cost-shared provision of well-water tests for arsenic. Finally, it analyzes measurement error in a satellite night light data product widely used in development research, and investigates the scope for using the data in very high spatial resolution.

The first chapter compiles 104 rounds of household surveys from 44 countries to study health-wealth trade-offs arising due to mining activity. Households in mining communities enjoy substantial economic benefits. Yet, these are counterbalanced by a health burden specific to environmental contamination. Adult women experience a ten percentage point increase in the incidence of anemia, and young children, a five percentage point rise in the prevalence of stunting. Prior evidence links both of these health impacts to metal toxicity – and in particular, exposure to high levels of lead. We show that there are health impacts only near mines of a type where heavy metal pollution is to be expected, and find no systematic evidence that health is affected in ways that are not specific to exposure to such pollutants. Benefits and costs are strongly concentrated in the immediate vicinity ($\leq$ 5km) of a mine. Consistent results emerge from a range of distinct identification strategies, including fixed effects models, an instrumental variables strategy, and two difference-in-difference tests tailored to the known association of certain mine types with heavy metal pollution, and to the pathophysiology of lead toxicity. Our results add to the nascent literature on health impacts

near industrial operations in developing countries.

The second chapter reports results from a randomized controlled trial conducted in Bihar, India. It assesses the scope for cost-shared provision of well-water arsenic tests, and studies how households use the information revealed by testing. Groundwater contaminated with arsenic of natural origin threatens the health of tens of millions of villagers across South and Southeast Asia. Because contamination varies greatly even over small distances, water quality tests can allow households to form agreements to share water from safe wells. Tests have largely been provided through public blanket testing campaigns. However, these important campaigns are conducted infrequently, and have not kept up with high growth in the use of privately-owned tube wells. Cost-shared private provision might therefore be a useful complement. We find that demand is substantial, and a degree of cost-sharing is possible. However, in line with prior evidence on cost-sharing in preventive health goods, we show that cost-sharing comes at the price of strongly reduced take-up. Even at a small price of Rs. 10, uptake drops to 69% from the universal adoption found under free provision. It falls further, to 22% of households, over our price range (Rs. 10 to Rs. 50 – about equivalent to daily per capita income). Repeating the sales offer after a two-year hiatus raises overall uptake substantially, from 27% to 45%. About one-third of households with unsafe wells switch to a safer water source. Households that bought at higher prices are no more likely to switch, consistent with an absence of sunk cost or screening effects. Finally, we demonstrate that households selectively forget and remove evidence of adverse test outcomes.

The final chapter assesses whether night lights data from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) are observed precisely enough to measure wealth at high spatial disaggregation. Night lights are routinely used as proxies of ground-based activity at the level of countries, sub-national regions, or metropolitan areas. Given the data's resolution (about one square kilometer at the equator), they might also be useful in studying processes at much higher spatial disaggregation – for instance, at the level of towns or villages. Yet, DMSP-OLS data are subject to several sources of measurement error that could interfere with such uses. I assess one error component, namely error in geolocation, in a new data set of 185 calibration sites that are small, bright, and remote. The offset between the actual location of light sources and their recorded location in the most commonly used yearly night lights data product is small enough to be ignored, even in applications where the spatial scales of interest are on the order of a few kilometers. Root mean square error is a mere 0.52km in zonal and 0.67km in meridional direction. I then illustrate the potential and limits of very high-resolution applications by benchmarking light data on household asset wealth in all official localities in Mexico. Night lights are a strong proxy measure of cross-sectional wealth differences even within small administrative units, in particular in the poorest, least populous, and most dimly lit regions. However, the analysis of changes over time is more subtle: the relationship between changes in brightness and changes in wealth is non-monotonic, and noise compounds when the data is used as a panel.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 1

Acknowledgements

I owe this dissertation and the learning it embodies to the support of many people. I have been truly fortunate, and will work passionately to make this gift bear fruit.

Foremost, I thank my advisors Doug Almond and Eric Verhoogen for giving their time and care to teaching me how to do research. And for their unfailing understanding of the competing demands of being a father as well as a student. To Eric, I am profoundly grateful for a decade of unique, patient mentorship. I thank Stefan Dercon for first welcoming me to academic research, and his warm support along the path.

I am grateful to the many Columbia faculty members who have helped this work advance. Kiki Pop-Eleches for unrelenting cheerful support, Lex van Geen for inviting me into his field work, Joe Graziano and Chris Small for sharing their knowledge and enthusiasm with an economist 'moonlighting' in their fields, Supreet Kaur, Suresh Naidu, Matt Neidell, Serena Ng, Bernard Salanié, Wolfram Schlenker, and Miguel Urquiola.

I warmly thank Alan Gelb, Nancy Birdsall, Susan McAdams, and Grahame Dixie for letting me continue as a professional while in graduate school, and for being entirely supportive as I learned to balance research, work, and parenting. It has meant the world to me.

I thank my friends in the Sustainable Development program for a wealth of inspiration, and Prabhat Barnwal and my co-authors for the joy of being creative and entrepreneurial together.

My work would not have been possible without the ingenuity and kindness of Mona Khalidi and Tomara Aldrich in solving administrative challenges. Nor could I have done it without the support of the Morningside Heights community. I thank Althea, Ann, Alex, Maya, Elizabeth, and Jany for their loving care of my daughter. Philipp and his team for a friendly and serene place to think. And many neighbors and friends for making daily life a joy in this extraordinary community.

Finally, and most of all, my family for unconditional love and support, and Anora for the happiest times in my life.

Chapter 1

Mines: the local wealth and health effects of mineral mining in developing countries

Mines

The local wealth and health effects of mineral mining in developing countries *

Jan von der Goltz   Prabhat Barnwal

Abstract

Do residents of mining communities face health-wealth trade-offs? We conduct the first extensive assessment of this question using micro-data from communities near about 800 mineral mines in 44 developing countries. Households in mining communities enjoy a substantial medium-term gain in asset wealth (0.3 standard deviations), but experience a ten percentage point increase in the incidence of anemia among adult women, and a five percentage point rise in the prevalence of stunting in young children. Prior evidence links both of these health impacts to metal toxicity – and in particular, exposure to high levels of lead. We observe health impacts only near mines of a type where heavy metal pollution is to be expected, and find no systematic evidence that health is affected in ways that are not specific to exposure to such pollutants. Benefits and costs are strongly concentrated in the immediate vicinity ($\leq$ 5km) of a mine. Consistent results emerge from a range of distinct identification strategies. Baseline results come from a cross-sectional fixed effects model, and mine-level and mother-level panels. An instrumental variables approach serves as a robustness check. To demonstrate that the observed health impacts are due to pollution, we develop two difference-in-difference tests tailored to the known association of certain mine types with heavy metal pollution, and to the pathophysiology of lead toxicity. Our results add to the nascent literature on health impacts near industrial operations in developing countries.

*The authors particularly thank Douglas Almond, Joseph Graziano, and Eric Verhoogen for their guidance and advice. We are indebted to Jesse Anttila-Hughes for his generous support and helpful discussions in the early stages of the project. We thank Ram Fishman, Alexander van Geen, Kyle Meng, Suresh Naidu, Matthew Neidell, Cristian Pop-Eleches, Bernard Salanié, Wolfram Schlenker, Anna Tompsett, and seminar participants for comments. Nicole Dussault and Dheeraj Sharma provided excellent research assistance. We gratefully acknowledge financial support from the Centre on Global Economic Governance and the Center for International Business Education and Research at Columbia University. All errors are ours.

1.1 Introduction

This paper studies the local wealth and health impacts of mineral mining in 44 developing countries. We show that, while residents living close to mines enjoy greater wealth, there is a trade-off: life near mines exacts a price in terms of specific health burdens.

In any country and at any time, the decision to live near centers of industrial activity involves weighing the promise of economic opportunity against the risk of disamenity caused by pollution. Nowhere is this choice starker than in developing countries. More often than not, opportunities for making a good living are precious and few. At the same time, pollution tends to be poorly regulated, and information on health risks and on ways to avoid them, scarce. Poor infrastructure and inflexible housing markets commonly make commuting to avoid pollution impracticable. Yet, while "the literature on the health effects of pollution has advanced greatly in the last two decades, almost all of this research has been conducted in developed country settings." (Greenstone and Jack, 2013, p. 12)

In the following paper, we present the first systematic empirical assessment of the micro-level trade-offs between health and wealth posed by the mining industry in developing countries. We seek to add to the very limited number of broad micro data analyses of local health impacts near *any* kind of industrial operations in developing countries. For the study of industrial pollution in poor countries, the mining and mineral processing industry is an attractive test case in that it poses particularly sharp trade-offs. Single plants generate very high value – in some instances, in the hundreds of millions or billions of dollars per year. The location of ore deposits dictates where mines open, and because of transport costs, often also where smelters locate. Therefore, large operations are found in remote areas where they dwarf any other enterprises – and the economic opportunities generated by the latter. Mines and smelters therefore tend to play a conspicuous economic role. At the same time, however, they are very large polluters, and precisely because they are important sources of revenue, foreign exchange, and employment, there is a risk of weak environmental regulation and enforcement.

The importance of mining to development is reflected in a long tradition of research on the macroeconomic implications of mining and the optimal management of mineral resources. However, there is little empirical evidence on the local economic impact of mining, and on its effects on other dimensions of well-being. This particularly includes implications for the health of local communities: although there is an important literature on pollution near mines, and an extensive body of knowledge on the toxic properties of common pollutants, there is scant systematic evidence linking the two. No more than a handful of case studies have carefully assessed the actual clinical consequences of exposure to a mining environment. This paucity of empirical evidence on the local welfare effects of mining is in stark contrast to the strong passions that mining projects habitually evoke among the communities affected. In some places, projects have been supported vociferously, and people have fought over the right to work in mines. Yet, in other places, mining has been desperately opposed, as citizens feared damage to their health and environment. Our work shows that these political passions are grounded in a real trade-off. Across a broad range of settings, the local benefits of mining are real, but so are the costs.

We analyze the effect of mining activity on asset wealth, on general health, and on two specific health outcomes known to be linked to pollutants that may be present around mines in our sample, namely anemia in adults and children, and growth in young children. To study the interplay of health and wealth effects across the developing world, we compile 104 waves of Demographic and Health Surveys from 44 countries. The pooled data provide us with about 1.2m household-level records and several million individual-level observations, spanning a time period from 1986 until 2012; they also record the geo-location of each cluster of households sampled. We overlay this household data with information on the location of mining and smelting operations across the world. We use a large cross-sectional dataset of mines that records the types of minerals mined, and characteristics of the local geology; as well as two business intelligence datasets that document annual production at individual mines. Guided by prior evidence from the environmental sciences on the spatial extent of pollution around mines, we define households *within 5km* of a mine to be in its direct vicinity, and consider those to be treated. We regard households *within 5-20km* of a mine to be in its general vicinity, and rely on those in constructing control groups. Using production data, we create pseudo-panels that enable us to compare our treatment and control groups across time, namely between years when the mine was operational, and when it was dormant.

We then construct a broad range of complementary statistical tests that rely on different control groups, and offer extensive placebo tests. In our baseline models, we estimate the effect of closeness to mines and smelters in the cross-section, and the effect of closeness and operational status in the panel. (We prefer pooling mineral mining and processing facilities, but show robustness to excluding smelters from the sample.) We weaken cross-sectional identifying assumptions by defining our control group conservatively, and by allowing for a fixed effect common to all clusters observed near the same mine, and in the same survey year. Because of the possibility of residential sorting, we argue that our cross-sectional estimates are best read as the long-run effect of mining on *communities*, much like district or county-level studies assess impacts on those units of analysis. To assess the effect of exposure to mining on individuals, we create two sets of pseudo-panels: a mine-level panel compares households observed near the same mine in different years, and a mother-level panel compares among siblings born in different years. The panels allow for common effects shared by all households observed in the same country and survey round. An instrumental variables (IV) approach further reassures us that our results are not due to endogenous choice of mine location or periods of operation; to this end, we use the location of mineral deposits and world mineral prices to instrument for the location and operational status of mines.

Beyond these standard identification frameworks, we develop two difference-in-differences tests that are tailored to prior knowledge on the toxic properties of mining pollution. Our purpose in designing these tests is two-fold. Firstly, they help bolster our claims to observing a causal impact of mining on health. In particular, we believe that they are largely immune to residential sorting. Secondly, our tests provide evidence to suggest that the observed health effects are due to pollution, not other mechanisms. To devise the tests, we (i) leverage knowledge on the association of specific mine types with lead contamination – and by extension, health impacts specific to lead – to conduct falsification tests. We show both that we only observe those health impacts that are expected from exposure to lead pollution, and that we only observe them near mine types strongly associated with the release of lead. Furthermore, we (ii) exploit detailed information on the birth history of women to describe a pattern of impaired ability to recover from blood loss after pregnancy among women living in mining communities, as compared to those living in the general vicinity. We argue that this effect is consistent with a known pathophysiological pattern of lead toxicity in adults, but not easily consistent with other mechanisms.

Our results show that, at the global mean, long-run asset wealth in mining communities rises by about 0.1 standard deviations of an asset index computed for the country where the community is located and the year in which the survey was taken. The medium-term wealth of households living in the vicinity of an operating mine rises by about 0.3 standard deviations. We illustrate that these are considerable effects, given the high variation in asset ownership within survey rounds. Wealth effects are strongly concentrated in the direct vicinity of the mine; there are benefits across the wealth distribution, although in the long run, the wealthiest households benefit the most; across countries, wealth gains are greatest near mines where the overall economic environment is poor.

We find clear evidence of two health impacts that are known consequences of exposure to lead and other heavy metals that may be present near mines. Thus, women in mining communities show depressed blood hemoglobin, and increases in the incidence of anemia of three to ten percentage points. They also recover more slowly from blood loss during pregnancy and delivery, a pattern consistent with prior toxicological research. Children in mining communities suffer some important adverse growth outcomes from *in utero* exposure, with a five percentage point increase in the incidence of stunting – although there is very little evidence of lower birth weight. Growth impacts are weaker among older children, perhaps because of the greater wealth enjoyed by households in mining communities. We note particularly that, while our data contains no good measure of cognitive ability, lead exposure has previously been shown to cause cognitive deficits in children at exposure levels below those associated with growth retardation, and far below those associated with overt anemia. By way of contrast to these specific health impacts, we find no effects on health outcomes that are not linked to heavy metal pollution, nor are mining communities differentially affected by other known causes of anemia, or under-served by health care.

Because our paper shows reduced form impacts (that is, it allows for the effect of exposure to a mining environment on health to play out through any channel, including greater wealth), our health results should be interpreted as the *compensated* impact of mining. By implication, since living in mining communities goes hand in hand with economic benefits across the distribution, there is no indication that ill health is caused by deprivation. Rather, health impacts arise despite wealth gains.

This paper seeks to make three contributions to the literature. It is the first to demonstrate that residents of mining communities in developing countries face a trade-off between real economic benefits and specific health costs. Secondly, we add to the limited evidence on the consequences of industrial pollution in developing countries. Finally, we complement the toxicological and epidemiological literature by showing that the health effects of mining pollution are salient in a study of the general population near a large number of mines (rather than in local treatment effects found in case studies), and are robust to tests that require weak identifying assumptions.

The remainder of the paper is organized as follows. Section 1.2 describes what is known about welfare in mining communities. Section 1.3 introduces results from environmental science and toxicology that guide the way we develop hypotheses, measure impacts, and interpret results. Section 1.4 discusses data, and Section 1.5 summarizes econometric methods. Section 1.6 presents results; Section 1.7 concludes.

1.2 Mining, wealth, and health

This section discusses the state of knowledge on mining and wealth (Section 1.2.1), and on health in mining communities (Section 1.2.2). Section 1.3 provides additional background on individual links in the causal chain from mining to ultimate health impacts, namely (i) pollution near mines, (ii) the body burden of pollutants in residents of mining communities, and (iii) the toxic impacts of substances released near mines.

1.2.1 Mining and wealth

Economics has traditionally studied mineral mining in the context of optimal resource management, or in a macroeconomic context of growth and public finance. For a textbook-level overview of the former, see, e.g., Hartwick et al. (1986); for a survey of the latter, Frankel (2010).

The economic impacts of mining at the local level have only recently received some attention. As of the time of writing, we are aware of only two published papers that study mining at the kind of disaggregated scale we consider. In a pioneering paper, Aragón and Rud (2013) leverage a change in local hiring and procurement policies in a single very large gold mine in Peru to identify local economic impacts. Incomes in communities within 100km of the mine showed an elasticity of 0.3 to production at the mine, alongside significant increases in the price of housing and of locally produced agricultural output, and higher local public spending. Wilson (2012) shows that asset ownership increased among residents of copper mining communities in Zambia during a boom in the 2000s. A second paper by Aragón and Rud (2015) investigates the impacts of gold mining in twelve operations in Ghana on agricultural productivity. It finds stark decreases in productivity (40%) in the general vicinity (less than 20km) of mines, relative to control areas farther away. Productivity losses in the general vicinity are accompanied by large increases in the poverty headcount (18 percentage points), and decreases in consumption, all driven by dire developments for rural households. The latter two papers and a working paper by Kotsadam and Tolonen (2013) use sub-sets of the micro data from the Demographic and Health Surveys also used for the present study. Kotsadam and Tolonen (2013) argue that mining activity in a comprehensive sample of African mines fosters sectoral shifts in employment out of agriculture (among women, into services, and among men, into skilled manual labor) and increases cash employment among women, but is also associated with women leaving the labor force altogether.

Long-term welfare in mining communities was also brought to the attention of the research community by Dell's (2010) work on the *mita* forced labor policy in Peru, although the focus of the paper is on institutions and development, rather than the direct welfare impacts of mining per se. In other related work, Acemoglu et al. (2013), Dube and Vargas (2013), and Monteiro and Ferraz (2009) have recently leveraged resource revenue at a disaggregated scale as an instrument in the study of other objects of interest (health expenditure, conflict, and corruption, respectively).

1.2.2 Health effects of mining

Our paper asks how significant are the ultimate health effects in the general population of exposure to pollution from every-day mining and mineral processing operations. Few studies have attempted this before, and to the best of our knowledge, none considers the possible trade-off between wealth and health effects and assesses the issue across many mine sites in a manner that allows for a causal interpretation of results.

Prior work in economics on the issue is limited. Aragón and Rud (2013) find a significant decrease in general health problems among adults with an expansion of production in the Yanacocha mine, Peru, and no effect among children. In their recent working paper on Ghana, the same authors find evidence of an adverse effect of mining activity on weight-for-height ratios and the prevalence of cough in children living in the general vicinity of twelve gold mines (perhaps due to air pollution around the mines studied), but no impact on stunting and diarrhea (Aragón and Rud, 2015). Some attention has been given to behavioral correlates of mining activity. Wilson (2012) finds that sexual risk-taking tended to decrease in Zambian copper towns during a boom. Corno and De Walque (2012) argue that in mining communities in southern Africa, there was increased risk taking and HIV infection among migrant miners, but no such effect among non-migrants.

In the field of public health, some case studies directly analyze health impacts in communities near smelters. (Factor-Litvak et al., 1999, p. 14) find impacts on "intelligence, physical growth, preschool behavior problems, renal function, blood pressure and hematopoiesis," among children of up to 7.5 years of age living in a smelter town in Kosovo. Among school-age children living near a lead smelter in Belgium, Roels et al. (1976) find changes in sensitive biomarkers that indicate an incipient disruption of the process of blood formation, but not overt anemia. Both papers show comparisons to a matched control group in addition to dose-response relationships. Dose-response relationships alone have also been reported between blood lead (PbB) and lower blood hemoglobin (Hgb)[1], as well as reduced nerve conductivity, among children living near a lead smelter in Idaho, U.S. (Landrigan and Baker, 1981; Schwartz et al., 1990). Baghurst et al. (1992) show a dose-response of IQ to PbB in children living near a lead smelter in Port Pirie, Australia. A range of papers by Hendryx and various co-authors (see for instance Hendryx and Ahern, 2008) shows cross-sectional correlations between county-level health outcomes and Appalachian coal mining, without clear

[1]The papers report hematocrit, not hemoglobin levels, but the two measures are closely correlated, and are both used to define anemia.

causal claims.

Pollution due to mining is a special case of industrial pollution, and the latter has been analyzed in large and well-identified studies. Yet, most of these investigate developed countries (see Currie et al. (2013) for a major recent contribution); studies of developing countries – especially using large samples – remain rare. Chen et al. (2013) study reduced life expectancy from air pollution due to power generation in China; Ebenstein (2012) assesses the effect of water pollution on gastro-intestinal cancer rates in China; and Rau et al. (2013) show cognitive losses from lead exposure near an abandoned toxic waste site in Chile. Hanna and Oliva (2011) describe reductions in air pollution from the closure of a large refinery in Mexico city, and an associated increase in labor demand. Studies of overall urban pollution (Arceo et al., 2015; Greenstone and Hanna, 2011) are related, but not specific to industry, while studies of air pollution from urban traffic (e.g. Gallego et al., 2013) are less closely related. We seek to contribute to this nascent literature by presenting a multi-site micro-data study of the comparative health and wealth impacts of an important industry, across many developing countries.

1.3 Scientific background

This section first discusses environmental pollution near mines, and its relationship to the body burden of toxicants (Section 1.3.1). We then establish that metals, and in particular lead, are of most interest as pollutants in our sample, and discuss the toxic effects of lead (Section 1.3.2).

1.3.1 Environmental pollution due to mining and its relationship to the body burden of toxicants

A voluminous literature in environmental science has catalogued the pollutants emitted in the course of normal operations near mines and smelters of different types. We base the following discussion on Alloway (2013), Ripley et al. (1996), and Wright and Welbourn (2002).

Local communities can be exposed to pollution through a multitude of channels. These include dust from mining, handling and processing; mine waste water; direct exposure to abandoned mine spoil and tailings; metals leached from tailings into soil and water; and particulate and gaseous emissions from roasting and smelting. Sometimes, the material extracted is itself of concern, such as in lead, uranium, or asbestos mining. At other times, pollutants are used in processing, such as in the case of cyanide leaching of gold, or gold and silver extraction by mercury amalgamation. Finally, sometimes the concern is with toxicants co-located with the mineral mined and released either in processing or weathering of mine spoils, such as in the case of heavy metals in non-ferrous metal mining.

Two stylized facts on pollution near mines are essential to the way we analyze the health impacts of mining.

(i) The kinds of pollutants near a given mine can be predicted well from the ore mined.

Table 1.1 summarizes pollutants associated with common (and non-exclusively defined) mine types in our sample. The mapping is far from exact, but serves as a useful first-order approximation. We leverage the association between target minerals and toxicants to compare health effects across mine types, and to show that we find predicted health impacts only near mine types where pollutants specific to the health impact in question are found.

Of particular interest to us is the association of "non-ferrous metalliferous mining and smelting industries ... with very high levels of heavy metal(loid) contamination of the environment." (Alloway, 2013, p. 43) Thus, 'polymetallic' mines, where any combination of copper, gold, lead, silver, and zinc are extracted, are linked with a characteristic suite of highly toxic pollutants that includes most prominently lead, but also arsenic, cadmium, and chromium. (We will refer to these metals and metalloids as 'heavy metals' – a term that is imprecise in that it does not refer to a well-defined group of chemical elements, but has the advantage of being in everyday semantics associated with the pollutants we have in mind. See Section 1.4.2.1 for coding notes.) Pollution near polymetallic mines is of particular concern both because heavy metals are important toxicants, but also because the minerals mined are often nested in sulfide rock. When exposed to air and water, the latter will tend to generate sulphuric acid, which in turn leaches metals from the mine's tailings; the resulting acid mine drainage can pose severe health and environmental concerns (Salomons, 1995).

(ii) The area in which highly polluted sites are found is typically small, and extends to at most a few kilometers around the mine.

Thus, for lead and in the case of smelters, high exposure ranges have been associated in the literature with distances from the point source of emissions of 0.5 to 4km. Mean blood lead levels (PbB) among children in the highly exposed communities ranged from 13 μg/dL to more than 40 μg/dL. (Table 1.2) All mean PbB values far exceed the reference value of 5 μg/dL (the 97.5$^{\text{th}}$ percentile of blood lead levels found in the U.S.) set by the Centers for Disease Control to "trigger lead education, environmental investigations, and additional medical monitoring," (CDC, 2012) as well as the laxer and more dated 'level of concern' of 10 μg/dL. (Roper et al., 1991)

In this paper, we do not directly observe environmental pollution or the body burden of toxicants. Rather, we use distance to the nearest mine as a proxy. The choice of a distance cutoff to define the treated group is therefore crucial. In line with the empirical evidence reviewed above, we look for health effects in a tightly defined treatment group, and consider only households within no more than five kilometers of a mine to have been exposed. This choice also corresponds to the extent of high-exposure buffer zones around mines in van Geen et al. (2012). It is considerably tighter than in other current studies of mining in economics, as is appropriate for our focus on health impacts.[2] A key benefit of working with our large multi-country dataset is that it allows us to restrict our treatment group in this manner, while retaining sufficient statistical power.[3]

1.3.2 Pathophysiological and clinical effects of lead and other metal exposure

As noted, the mines in our sample are associated with characteristic sets of pollutants. Because the latter are known to cause specific health effects, we can develop predictions for expected health impacts that are well-grounded in scientific knowledge. To the degree that we find expected health impacts, but not others, we strengthen our case that impacts are likely due to environmental pollution, rather than any other mechanism.

In our baseline investigation of health impacts, we do not distinguish between different types of mines. Yet, our main concern is with the health consequences of environmental contamination with heavy metals, and in particular, with lead. We focus on heavy metal contamination, first, because the health impacts of exposure are important and observable in our data, and second, because a large share of mines in our sample is associated with this type of pollution (40% of mines in the cross-section, and 70-90% in the panel, depending on definitions). Among heavy metal pollutants, lead takes a central role, because it is known

[2]Wilson (2012) uses a cutoff of 10km, while Aragón and Rud (2015, 2013) and Kotsadam and Tolonen (2013) use a baseline cutoff of 20km, with sensitivity analysis for other choices.

[3]With perfect data, we might define closeness even more restrictively. In the context of available data, a tighter cut-off would risk introducing noise, both because of the practice of jittering cluster geolocations in our socio-economic data, and because of the fact that we work with (imperfectly recorded) mine point locations, while mining operations can measure several kilometers across.

that the lead body burdens previously measured near smelters (reported above) are high enough to cause health problems that we observe in our data.

1.3.2.1 Sequalae of lead exposure observed in our data

The toxic properties of lead have long been studied, and are well understood.[4] The wide-ranging effects on adults include reduced blood hemoglobin (Hgb) and overt anemia, cognitive defects, hypertension, and impaired renal function. In our data, we observe only one of these conditions, namely low blood Hgb and anemia. We adduce two additional unspecific health outcomes as falsification tests, namely miscarriage and general grave illness.

For children under five years of age, we analyze two health outcomes that have previously been linked with lead exposure *in utero* and among young children: anemia and growth retardation. We use for falsification tests some health outcomes that have not been linked to lead (cough, fever), or linked only weakly (all-cause mortality), or at very high exposure (gastro-intestinal problems). Regrettably, we do not have a good measure of impaired cognitive performance and behavioral problems due to neurological damage in children. However, while the health impacts we do observe – anemia and growth deficits – are known to require high blood lead, “there is no evidence of a threshold for the adverse consequences of lead exposure” for intellectual development (Lanphear et al., 2005, p. 899). Hence, demonstrating overt anemia or growth deficits implies a strong likelihood that the affected individuals – and presumably others with lower PbB – also suffer some cognitive and behavioral impairment. (Appendix Table 1.E summarizes how the health consequences we observe affect the well-being of those exposed, and what their economic cost might be.)

[4] See ATSDR (2007) for a full discussion.

1.3.2.2 Hematologic toxicity of lead

Lead depresses blood Hgb levels both by shortening red blood cell life spans, and by interfering with enzymes essential to the synthesis of the heme component in hemoglobin. Enzyme activity begins to be disrupted at very low PbB, but is not measured in our data. Effects on Hgb – which we can observe – have previously been reported at high PbB levels: in excess of 40μg/dL in children, and 50μg/dL in adults (ATSDR, 2007, pp. 69, 71f). That is, we expect the hurdle to finding impacts on Hgb to be quite high.

Therefore, we devise an additional, more sensitive test of hematotoxic effects. We build upon the insight in Grandjean et al. (1989) that, even when lead exposure is too low to reduce Hgb *levels* in adults, "increased demand on the formation of blood following blood loss could result in a delayed blood *regeneration* in individuals exposed to lead" (p. 1385 - our emphasis). Grandjean et al. demonstrate this effect by comparing Hgb recovery after blood donation in lead factory workers and a control group. In our study, we show that analogously, Hgb recovery is similarly impaired among women in mining communities after another kind of blood loss, namely pregnancy and delivery

The effect of lead on children is of particular concern, since children are both more sensitive in their reaction to body burdens of lead, and (in the case of lead ingested with food) absorb far larger portions of lead. In the case of anemia, however, we expect effects to be *harder* to demonstrate in children than in adults. This is because, by contrast to adults, children are able to compensate for erythrocyte loss by increasing production of the hormone erythropoietin (EPO), and thus boosting the generation of red blood cells. (Factor-Litvak et al., 1998)

In summary, based on the state of scientific knowledge, we expect Hgb in residents of mining communities to be measurably affected only if there is substantial exposure to environmental lead. An effect should be detected most easily in the recovery of Hgb after blood loss, followed by Hgb levels in adult women, and least readily in Hgb levels in children.

1.3.2.3 Effects of lead on child growth

While there is an epidemiological link between lead and anemia, and several hematotoxic mechanisms are known, studies are in less agreement on the effect of lead on growth in children, and "the mechanism by which lead may reduce a newborn's size is unknown." (Hernandez-Avila et al., 2002, p. 486)

Correlations have been observed – including at moderate PbB on the order of 10μg/dL – between maternal or child blood lead and gestational age, as well as a wide range of measures of height and weight from birth to adolescence. (ATSDR, 2007; Bellinger et al., 1991; Hernandez-Avila et al., 2002; SanÍÌĄn et al., 2001; Zhu et al., 2010) However, other studies have failed to show such correlations; indeed, it is common for a study to find impacts on some dimension of growth, but not on others, with no conclusive pattern of which indices are sensitive.

In this paper, we seek to exclude both endogeneity and small-sample variation as potential sources of ambiguous results. However, while we are able to show that *in utero* exposure affects one dimension of growth (height for age), our results mirror the existing evidence in that we find no clear effects on another key measure of growth (birth weight). In addition, in our study sites, growth effects are concentrated among infants, but abate in older children. As context for this finding, we note that, while, as a stylized fact, "blood lead levels [peak] in the age range of 1 to 3 years" (Bellinger, 2004, p. 1017), there is an important earlier path of exposure, through transfer of lead from the mother's body through cord blood and breast milk. Indeed, "infants are born with a lead body burden that reflects the burden of the mother," (ATSDR, 2007, p. 223) with correlations as high as 0.8 between maternal and infant PbB (Lauwerys et al., 1978, p. 280). Finding health impacts among infants is therefore particularly plausible if there is evidence of significant maternal lead burdens.

1.4 Data

1.4.1 Socio-economic and health data

We obtain socio-economic and health data by pooling all 104 available geo-coded Demographic and Health Surveys (DHS) from countries for which we have mining data. This yields a dataset of repeated cross-sections covering 44 countries, with a total of 1.2m households, and several million individual records. About 170,000 households are within no more than 20km of a mine recorded in our data, and enter our analysis. (Table 1.3) Their location is shown in Figure 1.1.

The DHS data has some notable strengths: it covers a very broad range of developing countries; surveys have been conducted for nearly 30 years; individual surveys are fairly comparable; sampling cluster geocodes are available for many survey rounds; and there is strong data on maternal and under-five health, including anthropometrics and specifically, hemoglobin (Hgb). These features currently make DHS an obvious choice to study health and development at the micro level across multiple countries.[5]

However, the data also has some important limitations with implications for our work. (i) There is relatively little data on socio-economic status, no information on wages, and little information on employment. We therefore work with an asset index, rather than more direct measures of wealth or of income, and discuss employment outcomes only in passing. (ii) Because the surveys have kept changing and improving, very few indicators of interest to us were collected in all surveys. Indeed, working with the largest set of observations for which all indicators are available is impractical, because the number of observations is very small. On the other hand, estimating results on pair-wise common sets would lead to

[5]Other data with high coverage that include both health and socio-economics are either less rich (IPUMS), or less harmonized (LSMS).

tedious repetition. We seek to strike a balance, and present side-by-side comparisons for core results. (iii) Finally, we stress again that the data are cross-sectional. Therefore, while our difference-in-difference tests are designed to yield evidence of causal effects, they always compare across different individuals.

Our core measure of wealth is a standard asset index computed over household durables and housing characteristics. (Filmer and Pritchett (2001); see Appendix 1.B for details.) We base it on the largest set of wealth proxies available within each survey round, but do not include slow-moving or immutable traits of the household head, such as gender, marital status, or education.

We obtain from the DHS detailed data on health among children below five years of age, and among women aged 15-49 years. There is little information on older children, men of any age, and women aged 50 years and over. Our core health indicators are blood Hgb levels and an age-adjusted height index. Hgb is adjusted for altitude, and expressed either as a continuous measure in units of grams of hemoglobin per deciliter of blood (g/dL), or as a binary indicator for the clinical condition of anemia, associated with blood Hgb below 12 g/dL in non-pregnant women and 11 g/dL in pregnant women, and in children (World Health Organization, 2011). Following standard practice, height is expressed as the difference between a respondent's height and the age-group median, normalized to standard deviations. We normalize using the median and standard deviation provided by DHS (alternative normalizations make no empirical difference). We consider the continuous height measure, as well as the clinical outcome of stunting (severe stunting), defined as a height of at least two (three) standard deviations below the median.

In addition to our core outcomes, we collect data on a range of general adult and child health outcomes, on health care, sexual risk taking, nutrition, and employment and occupation. Finally, we construct infant and under-five mortality data for all children whose births were recorded in any survey module.[6]

[6]Because we construct these variables from birth records of all children ever born to the women in sample,

1.4.2 Mining data

We obtain data on the location and characteristics of mines and mineral deposits from four data sources. These include a very large cross-sectional dataset that allows us to make meaningful claims about the mean effect of mining across many developing countries; two datasets of mine output that permit us to estimate mine-level panels; and an additional dataset of mine locations that serves to ensure robustness of our findings to measurement error in geo-location. In total, we observe communities near 838 mines in the cross-section, and 515 mines in the panel, though the set of mines that enters our estimating samples is generally smaller.[7] (Table 1.3)

1.4.2.1 Cross-sectional data on mine location and characteristics

In the cross-section, we work with the United States Geological Survey's Mineral Resource Database (United States Geological Survey, 2005). It contains the point locations of a very large set of mines, legacies, deposits, and smelters (about 25,000 locations in total) across developing countries. The data records geological information and some basic description of the nature of the mine for a substantial subset of entries. However, there is no data on production, and start dates and status of operation are only available for very few mines.

In our baseline cross-sectional sample, we include all active mines, legacies (that is, former mines that are now dormant), and smelters. We include smelters because they are often located close to mines, and it is intuitive to think of a single mineral extraction and

---

the mortality variables must be interpreted as being conditional on the mother's survival until the time the survey was taken.

[7]Nearly all of those mines enter into our model when we use state-level effects (see Part 4). The number of mines near which we observe at least one community within 5km (treatment) and one within 5-20km (control) is lower, with 226 mines in the cross-section, and 175 in the panel. These are the mines that enter into our mine-effects models.

processing chain from mining to smelting.[8] We include legacies, because the cross-sectional data gives us little guidance in defining whether a mine was operational during a given survey round. The resulting treatment definition should be thought of as yielding 'the effect of living in a location ever exposed to mineral mining or processing'.

We extensively parse information on the types of minerals present in a given location to sort mines into larger groups that share the same expected pollutants and hence, the same health effects. We remove from our baseline sample all quarries (see Appendix 1.A for a definition). We do so because we seek to study the welfare impacts of mining as an industry that generates very high value added, but is potentially severely polluting. Quarries differ from mineral mines in both respects, at least as a matter of degrees. As we have argued above, we are particularly interested in polymetallic mines near which we expect pollution with heavy metals, and particularly with lead. For the purposes of the present paper, we define a mine to be a 'heavy metal' mine if (i) lead is being mined or smelted, or (ii) lead, though not targeted for extraction, is known to be present in significant amounts, or (iii) any two of the metals copper, gold, silver, and zinc are being mined or processed. This definition is necessarily imprecise, but gives due recognition to the special role of lead, and seeks to exclude metal mines with different pollutant characteristics. For instance, among gold-producing mines, it would aim to exclude alluvial gold deposits, where gold is typically the only metal of interested, and we expect mercury contamination from processing to be the primary concern, rather than lead pollution.

1.4.2.2 Mine-level production data

Since the USGS data provides virtually no time variation, we draw additional information from two business intelligence firms: Infomine (2013), and IntierraRMG (2013) – for whose

[8]In Appendix Table 1.G, we show that our core results are nearly fully robust to excluding smelters. In one case (panel results on women's hemoglobin), the effect is not significant, although it is consistent in sign and approximate size.

product we henceforth write 'RMD', for 'Raw Materials Data'. Both sources record dates of operation and annual production, alongside diverse additional characteristics of the mines. Most mines included in the Infomine data are also available in the RMD data, but not vice versa. We therefore work with RMD as our basic data, and add those Infomine entries that are not also contained in the RMD data. RMD mines are more homogenous than those in the USGS sample: most of them are large mines, and most of those close to DHS clusters are metal mines. While the set of mines included is far smaller than for the USGS data, coverage of large mines is quite comprehensive, and the mines recorded in the dataset account for a very large share of global metal production. For instance, they account for around 80% of global gold production and 80-90% of global iron ore production in the most recent decade for which data is available.

Because there is some question as to the precision of geolocations recorded in the RMD data, we use mine geolocations from an additional dataset, Mining Atlas (2014), for three purposes. First, we add geolocations for RMD mines wherever location is missing in the original data. Secondly, we use company records and Google Earth images to investigate the small number of cases where there are very large discrepancies in location between the two sources; we discard a few records where location is plainly not recorded with any precision in either dataset. Thirdly, we use the two independent but noisy measures of location to check robustness of our results to measurement error in geolocation (see Appendix 1.D).

1.4.3 Other data

For the purpose of constructing a time-varying instrumental variable, we retrieve data on mineral prices from various sources, summarized in Appendix 1.A. In order to describe how the wealth effects of mining vary with the economic environment, we obtain country-level data on GDP and governance from the World Development Indicators; data on the efforts a given country made toward compliance with the Extractive Industries Transparency Initiative (EITI) from the Initiative's website (www.eiti.org); and state-level data on governance, geography, infrastructure, and education from Gennaioli et al. (2013).

1.5 Econometric Specification

1.5.1 Baseline treatment definition

We define exposure to mining as being geographically close to a mine in the cross-section, and as closeness interacted with the mine being active in the panel. This choice is immediate for the study of economic impacts: with transport and search cost, distance is the treatment of interest. For the purpose of studying health impacts, distance acts as a proxy for pollution – which we do not observe.

We define a cluster as being 'close', and hence, 'treated', when it is within five kilometers of the nearest mine. We will also refer to this as the 'direct vicinity' of the mine. We define a cluster as being in the control group when it is within 5-20km of the nearest mine. We will refer to this as the 'general vicinity' of the mine. As noted above, we bound our treatment group tightly, to enable us to detect health impacts within the region in which pollution is likely to occur. Bounding our control group conservatively greatly eases the stringency of identifying assumptions required for a causal interpretation of our results. The cost of working with these definitions is that we can only achieve reasonable sample size by allowing our panels to be unbalanced. We argue that this is a reasonable price to pay for the sake of working with a treatment definition that is in line with prior scientific knowledge, and a control group definition that promises to provide a credible counterfactual.[9]

[9]For the study of wealth benefits alone, a natural alternative would be to study effects of mine density in (hopefully quite balanced) panels of administrative units. This would, however, vitiate the purpose of studying health effects.

In the panel, we define mining activity as a dummy variable taking value one when the mine had non-zero output, and value zero when the mine was known to have had zero output. (We conservatively impute inactivity – see Appendix 1.A.) That is, we consider only extensive margin impacts of production. We do so because year-on-year variation in output is likely to be more weakly associated with health outcomes. In this, mines differ from sources of pollution studied elsewhere. Extracting minerals from the ground, breaking them up, and processing them generates a flow of pollution. At the same time, however, the stock of tailings dumped after processing will in many cases continue to pollute. The exact time pattern of pollution is thus hard to predict, but is bound to lie somewhere between a pure flow and a pure stock problem. We hope to do it justice by studying extensive margin variation alongside the cross-sectional 'once on, always on' measure.

1.5.2 Cross-sectional model

Identification in the cross-section rests on a conservative choice of control group, and restrictive group effects. Because they cannot decisively address the possibility of residential sorting, the correct way to read our cross-sectional results is to view them as the long-run effect of mining on 'mining *communities*', much as a district or county-level study estimates effects on those units. As such, we believe they can be interpreted as causal; and to the degree that regional disparities matter, they are of policy interest. Our difference-in-differences models then provide evidence that impacts are unlikely to be driven by sorting, and allow us to make stronger claims about the well-being of 'people exposed to mining'.

In our baseline specification, we consider outcomes y for individuals or households i in sampling cluster j within no more than 20km of a mine, conditional on whether the cluster is *close* (within 5km) to a mine, and conditional on other covariates X. Because distance is measured between mines and sampling clusters, the treatment varies at the cluster level, not the individual level. Covariates always include an indicator for whether the cluster is in an urban or rural setting, and some appropriate measure of the age of the respondent, the respondent's mother, or the household head. Because DHS conducts repeated cross-sections, our model allows for repeated measurements of effects near the same mine, while accounting for year-specific effects in each round of measurements. We therefore use common effects γ for all observations near the same mine surveyed in the same year (mine-year effects), and account for residual correlations by clustering error terms at the mine level (not the mine-year level). Wherever the outcome of interest is binary, we model it using a linear probability model.

$$y_i = \beta_1 close_j + \beta_2 X_i + \gamma_{mine-year} + \epsilon_i \tag{1}$$

Identifying assumptions would be violated if mining towns differed from neighboring communities in geography, institutions or other characteristics in ways that correlate with potential outcomes. However, differences would have to arise even compared to locations very close by, because we restrict control locations to those no more than 20km away from the nearest mine. Identification is also only affected by such differences if they are not in some way due to the presence of the mine in long-run equilibrium (for instance, through infrastructure construction, or the emergence of institutions).

1.5.3 Pseudo-panel model

We have argued that our cross-sectional setup offers valid estimates of the long-run impact of mining on communities. Still, it says less than is desirable about mechanisms of treatment transmission, and due to the possibility of sorting, it does not allow us to make claims about the impact of mining on individuals. An immediate way of addressing both challenges is to construct pseudo-panels from the repeated cross-sectional DHS surveys. We construct these in two ways. Firstly, we compare observations from households surveyed at different times, but near the same mine ('mine-level panel'). Secondly, we compare children born to the same mother at different times ('mother-level panel'). Plainly, comparisons in each case are across different individuals.

Equations 2 and 3 describe the mine-level and mother-level models. We analyze outcomes for individuals i in cluster j at time t.

$$\begin{aligned} y_{i(t)} = \beta_1 close_j + \beta_2 operating_{j(t-\tau)} + \beta_3 close_j * operating_{j(t-\tau)} \\ + \beta_4 X_{i(t-\tau)} + \gamma_{mine} + f(t) + \epsilon_{i(t)} \end{aligned} \tag{2}$$

$$\begin{aligned} y_{i(t)} = \beta_1 operating_{j(t-\tau)} + \beta_3 close_j * operating_{j(t-\tau)} \\ + \beta_4 X_{i(t-\tau)} + \gamma_{mother} + f(t) + \epsilon_{i(t)} \end{aligned} \tag{3}$$

In Equation 2, we allow for time-invariant effects γ_{mine} for each mine, and model outcomes at time t as being conditional on whether the respondent lived in a community *close* to a mine during the time period relevant for treatment, $t - \tau$, and whether the mine was *operating* at time $t-\tau$.[10] The time periods of interest t and $t-\tau$ depend on the outcome being investigated. For instance, where we analyze height-for-age in children, the outcome is measured in the survey year t, and may be modeled conditional on exposure to mining operations during the survey year ($\tau = 0$), the birth year ($\tau =$ age), or while the child was *in utero* ($\tau =$ age $+$ 1). The model also includes time-specific effects $f(t)$. We believe country-year dummies are sufficiently flexible and appropriate for sample size. We use these in our baseline models, and show robustness to using different time effects. Modifications in the mother-level model

[10]For each respondent in our sample, we only observe current residence, and how long the household has been resident there. We have no information on previous residence. Therefore, the panel is inherently restricted to respondents who have lived in the location where they were surveyed for at least τ years. (Although they may have moved to their present location at a time before $t - \tau$.)

are immediate (Equation 3); because of the much smaller sample sizes, we include country linear trends $f(t)$ in our baseline model.[11]

1.5.4 Difference-in-differences tests tailored to the health conditions studied

For some indicators, our sample is small near mines where there is production information, so that the pseudo-panel tends to be highly unbalanced. We therefore leverage the scientific understanding of the health conditions of interest to our study to construct additional difference in differences tests. Like the pseudo-panel, they compare the impact of mining across groups that are and are not expected to show effects. However, unlike the pseudo-panel, they do not rely on the use of time-varying production data, and hence, tend to preserve sample size better. Because they each build upon a different insight into the likely nature of exposure and the organism's reaction to it, they generate distinct control groups, and hence, further "reduce the importance of biases or random variation in a single comparison group" (Meyer, 1995, p.157).

Mine types: Firstly, we make use of the fact that, as discussed above, distinct mine types are associated with specific pollutants and health effects. This allows us to contrast differences across distance groups near mines where an effect is expected, and near mines where none is expected, as in Equation 4. (The effect of *heavy metal mine* alone is collinear with mine-year effects.)

$$y_i = \beta_1 close_j + \beta_2 heavy\ metal\ mine_j + \beta_3 close_j * heavy\ metal\ mine_j + \beta_4 X_i + \gamma_{mine-year} + \epsilon_i \tag{4}$$

[11] Notice that, because we do not observe location of prior residence for migrants, no coefficient on *close* can be estimated in the mother-level panel.

Identification rests on the assumption that potential outcomes vary among those close and not close to the mine in similar ways near mines of different types. Most obviously, if wealth effects varied systematically among mine types, health results might be confounded. With respect to preference-based sorting, the assumption would be violated if respondents were aware of how mine types differ in health outcomes, and sorted accordingly. We address the issue in two ways. Firstly, we compare DiD results on health to those on wealth, and show that differences arise for health outcomes, but not wealth. Secondly, we show that there are DiD effects only on specific expected health outcomes, not general health.

Maternal Hgb recovery: Secondly, we develop a DiD test based on the observation that in lead-exposed adults, the recovery of Hgb after blood loss is even more readily affected than the steady-state level of Hgb. As discussed above, this result was previously proven by studying Hgb recovery after donating blood. Of course, we cannot identify blood donors in our sample. We do, however, observe one population group that experiences dramatic drops in Hgb: women who are pregnant, or have recently given birth. This allows us to formulate a test that asks whether differences in Hgb between women i in mining and control communities j are particularly stark during pregnancy and postpartum. In our preferred specification, we estimate the model with state-year effects, since the number of women we observe within the time period of interest is borderline too small for allowing for mine-year effects. (We discuss identifying assumptions and extensive robustness checks below, in Section 1.6.2.)

$$\begin{aligned} y_i = \beta_1 close_j + \beta_2 pregnant\ or\ postpartum_i + \beta_3 close_j * pregnant\ or\ postpartum_i \\ + \beta_4 X_i + \gamma_{state-year} + \epsilon_i \end{aligned} \tag{5}$$

1.5.5 IV models

Finally, we use both cross-sectional and panel IV strategies to study wealth effects. Our purpose for the IV estimates is somewhat narrow: they provide reassurance against endogenous choice of location (even within 20km) in the cross-section, and endogenous decisions to produce in the panel. However, because they do not help address residential sorting, we discuss results relatively briefly, and for wealth only – for health impacts, we instead rely on the additional DiD tests described above.

1.5.5.1 Cross-sectional IV

In the cross-section, to instrument for whether a cluster is within 5km of a mine, we use the dummy (Wald) instrument *deposit* that simply indicates whether there is a mineral deposit within 5km of a given cluster (Equation 6).[12] The sample is restricted to clusters within no more than 20km of a deposit.

$$\begin{cases} y_i = \beta_1 close_j + \beta_2 X_i + \gamma_{state-year} + \epsilon_i \\ close_j = \phi deposit_j + \delta_{state-year} + \eta_j \end{cases} \tag{6}$$

Because coverage of deposit locations in the cross-sectional data is very broad, we can think of our IV estimates as general population effects. Because there can be no mine without a mineral deposit, there are neither 'defiers' nor 'always-takers', and we can interpret IV estimates as the effect of treatment on the treated. (Imbens and Wooldridge, 2009) Unsurprisingly, the dummy instrument is exceedingly strong. Since the true global distribution of mineral deposits is exogenous to human activity, the instrument is also exogenous, as long as there is no preferential *prospecting* for minerals. We believe this is likely the case, since

[12] This is similar in spirit to the geographic instrument in Duflo and Pande (2007).

all anecdotal evidence suggests that mining companies will seek out promising deposits in virtually any location, regardless of geographic or political obstacles. We also believe that the instrument satisfies the exclusion restriction. The most likely violations would be due to topographical features such as land quality, gradient, or water availability. Because we work at small spatial scales and across many countries, potential violations are hard to test directly. Yet, since we strongly restrict our analysis in space, characteristics would have to vary systematically over small scales to cause any problems.

1.5.5.2 Panel IV

Our cross-sectional IV strategy extends very naturally to the panel setting, by interacting the presence of mineral deposits with world minerals prices. Our panel data does not have very high coverage of mineral deposits, but it does include some deposits that are being explored or prepared for exploitation. We adjust the panel IV sample to include such deposits. Hence, in Equation 7, we treat the variable *deposit* that records whether cluster j was within 5km of any deposit as exogenous.

$$\begin{aligned} y_{i(t)} = \beta_1 deposit_j + \beta_2 operating_{j(t-\tau)} + \beta_3 deposit_j * operating_{j(t-\tau)} \\ + \beta_4 X_{i(t-\tau)} + \gamma_{mine} + f(t) + \epsilon_{i(t)} \end{aligned} \tag{7}$$

We then instrument for whether the mine was *operating*, and for the interaction of closeness and operating status, using world mineral prices *price*, and their interaction with *deposit*. (See Appendix 1.A for a full description of the instrument.)

1.6 Results

1.6.1 Effects on wealth

Mining towns are wealthier than neighboring communities, both in the long run and the medium term

Households in mining communities are at the mean considerably wealthier in terms of asset ownership than control households. The magnitude of the cross-sectional effect at the global average is on the order of 0.11 standard deviations of the asset index. (Table 1.4, Column 1) In the mine-level panel, the DiD coefficient on the effect of living close to a mine in a year when it is operating is 0.26 standard deviations of the asset index in our preferred specification. (Column 3) Since survey rounds are typically about five years apart, we interpret this as a medium-term effect.

The effect size is appreciable, given that in the countries in our sample, there is generally great within-country variation in asset ownership. In the linear index, the magnitude of the cross-sectional effect is comparable to that of owning a car or motorbike in the case of Peru in the year 2000, and to the effect of owning a radio or a watch in the case of Burkina Faso in the year 2010. The panel effect is comparable to the impact on the index of having an electricity connection or living in a dwelling with finished flooring in the case of Peru in the year 2000, and to the effect of owning a motorbike or mobile phone in the case of Burkina Faso, in the year 2010. (See Appendix 1.B for a description of the index and for examples of factor loadings.)[13]

[13]Regrettably, the DHS surveys have no wage data, and limited coverage of employment. The sample of men living near mines in our sample for whom employment data was collected is small. In consequence, an in-depth analysis of effects on these core dimensions of welfare is not possible. In the cross-section, unemployment among men is virtually unaffected, consistent with long-run general equilibrium. As is intuitive, the sectoral share of agriculture decreases alongside ownership of agricultural land. In the panel, employment effects tend to be adverse in sign – consistent with queuing – but we caution that the estimates are noisy and not stable. (Results available upon request.) We refer the reader to Kotsadam and Tolonen (2013) for a detailed discussion of effects on women and sectoral shifts in sub-Saharan Africa.

We argue below that, because of the spatial pattern of long-run wealth effects, the cross-sectional baseline estimate should be interpreted as a lower bound. In Appendix 1.C, we show that our unweighted baseline estimates are smaller than estimates obtained by (i) weighting each mine equally, or (ii) weighting by estimates of the mine-year population. In Appendix 1.D, we use two independent measures of the geolocation of mines to instrument with one distance measure for the other, and show that our baseline results likely carry substantial attenuation bias – in our preferred specification, some 18% of the estimate. Cross-sectional IV estimates yield results that are close to and not statistically different from both our baseline results, and the OLS benchmark estimated on the IV sample. Panel IV estimates are somewhat larger than the benchmark, but not significantly different. (Table 1.5)

We have argued that, if the object of interest is the effect of mining on household welfare, rather than on the spatial distribution of wealth, the most salient identification concern in the cross-section is residential sorting. Panel results can be presumed to be more robust, but with about five years between survey rounds, there is still the possibility that sufficiently rapid sorting could influence results. We therefore separately study results for households that report never having moved from their current location. Effects are somewhat smaller and weaker (if not significantly different) among never-movers in both the cross-section and the panel. (Table 1.4, Columns 2 and 4) We interpret this as limited evidence of sorting of migrants with better potential socio-economic outcomes into mining communities, or sorting of previous residents with better potential outcomes out of mining communities.[14]

Spatial extent of the wealth effect

Wealth effects decay steeply with distance to the nearest mine. In the panel, effects are limited to those living with 5km; in the cross-section, there is a gradient in wealth up to

[14] For background, we note that there is only weakly more migration in mining communities than in neighboring communities. However, in both mining and control communities, the share of migrant households is very high: around 60% of households migrated at some time, and about 23% migrated within the five years preceding the survey. Sorting could therefore easily explain cross-sectional differences, if the characteristics of migrants (including those unobserved households who left the communities) are sufficiently different.

a distance of 15-20km. (Figures 1.2 and 1.3) That is, in the long-run, communities in the general vicinity are economically affected to some degree, although less so than those in the direct vicinity. Hence, the cross-sectional treatment effect in our baseline model is smaller than the wealth effect on the direct vicinity of mines, as compared to those living *outside* of the general vicinity, within 20-40km of a mine (0.4σ – results not shown). Conversely, it is larger than the *average* effect of living either in the direct or general vicinity of the mine, as opposed to living at 20-40km (0.05σ).

The difference in spatial patterns between the cross-section and the panel allows for a number of explanations. If both patterns are well-identified, one would argue that the discrepancy reflects the contrast between medium-term and long-run impacts, with further diffusion of wealth effects over time. If we were not convinced of identification in the cross-section, we might feel that the pattern suggests that mines tend to locate in places that are already wealthier than their surroundings. We note that, even in the cross-section, the estimated spatial extent of treatment effects is smaller than in the case study analyzed in Aragón and Rud (2013, p. 26), who find "positive and significant [income effects] for households located within 100km of Cajamarca city," the community closest to the mine studied. The discrepancy could be due to the fact that Aragón and Rud study a policy change that can be presumed to be very favorable for local welfare; or the fact that they consider the case of a very large mine in a region with reportedly high transport cost. In addition, Aragón and Rud have income data available; presumably, a more sensitive measure of well-being than our asset index.

Effects on the distribution of asset wealth

Mining is associated with wealth benefits across the distribution, though in the long run, there are much higher gains for the top quantiles, and a mild increase in wealth inequality. Benefits are more evenly distributed among never-movers. The distributional pattern might, for instance, reflect slow sorting of high-income households into mining communities, or the gradual emergence of economic opportunities that are open only to a select few.

We obtain quantile regression estimates using the two-step procedure described in Canay (2011). The results suggest that closeness to mines raises long-run asset wealth quite evenly across the distribution, with effect sizes for most quantiles close to the mean effect. (Figure 1.4) That said, the top 5-10% benefit the most, with gains about three times as large as those at the median. Gains at the top are more limited among never-movers. In the panel, if anything, benefits are progressive, and the top quantiles gain less than others (Figure 1.5); this pattern is comparable to the distribution of income effects found in Aragón and Rud (2013).

Secondly, we directly consider effects on a simple measure of within-cluster inequality, namely the absolute deviation of a household's asset index value from the cluster mean.[15] In the cross-section, the mean absolute deviation increases moderately among all households, by 0.03 standard deviations of the asset index, or one-fourth of the cross-sectional wealth effect. (Table 1.4, Column 5) There is no effect among never-movers, nor in the panel. (Columns 6-8)

Correlates of long-run effects across countries

Long-run wealth effects vary greatly across mining communities. Table 1.6 shows correlations of mine-level wealth effects with measures of the larger economic, geographic and policy environment. Gains are greatest where the economic environment is weak, across a range of indicators – GDP, education, access to infrastructure, some dimensions of remoteness, and (directionally only) measures of institutional quality.[16] While these correlations cannot be

[15]This simple index seems more appropriate than more familiar inequality indices both due to the small number of households in many clusters, and to the nature of the mean-zero standardized asset index.

[16]Appendix 1.L shows the distribution of treatment effects across world regions and countries; correlations with measures of overall development empirically supersede regional patterns.

interpreted as causal relationships, they raise the question whether the local economic effect of mining might be driven not by the interaction of mining with other economic activity, but by the opportunities mining provides in areas where there is a paucity of other options.[17] With the same caveat regarding causal interpretation, we also note that we do observe stronger wealth effects in surveys conducted in countries at a time when the country had completed a report for the Extractive Industries Transparency Initiative,[18] or (weakly) when it had participated in the EITI in any way.

1.6.2 Evidence of hematologic toxic effects

We have argued above that exposure to lead among residents of mining communities may affect the hematopoietic system and reduce red blood cell survival. In the DHS data, we observe only a single indicator of potential hematologic toxicity – blood Hgb concentrations. As argued in Section 1.3.2.2, we would expect most strongly to see a reduced ability to *recover* from blood loss in adults, perhaps alongside depressed Hgb levels. In children, we might expect to see reduced blood Hgb levels, though in the age group we observe, children are likely able to compensate for lead exposure. Our results confirm this expectation: we find strong evidence of lower Hgb levels and slower Hgb recovery after blood loss in adult women, and weaker evidence of lower Hgb levels in children.

Hemoglobin levels in adult women are strongly depressed in mining communities In the cross-section, blood hemoglobin (Hgb) levels are depressed among women living in mining communities by about 0.09 g/dL. The effect among never-movers is larger (0.13

[17]We emphasize that, because we study effects purely at the local level, the correlation between local benefits and a weak economic environment cannot be read to contradict findings from the resource curse literature. Our findings have no implications for whether, beyond the local level, resource revenue creates corrupt structures or drives Dutch disease.

[18]See www.eiti.org. The EITI describes itself as "a global coalition of governments, companies and civil society working together to improve openness and accountable management of revenues from natural resources.'

g/dL), consistent with longer exposure to environmental lead, although (on this smaller sub-sample) it is just below significance (t = 1.56). Considering directly the clinical outcome of anemia, we find that prevalence is significantly elevated by three percentage points among all households, and by five percentage points among never-movers. (Table 1.7) Appendix 1.L shows the distribution of mine-level effects across countries.

Panel results confirm these patterns. Point estimates are larger, with DiD coefficients of a 0.33 g/dL decrease in blood Hgb, and a ten percentage point increase in the incidence of anemia in our preferred specification. (Table 1.7, Columns 3 and 6) A number of causes could account for the larger point estimate in the panel; notably, the share of metal mines associated with lead pollution is high in the panel sample (and, as we show below, the treatment effect is concentrated near such mines). In the long-run, there might also be more adaptation to avoid pollution.

The size of the effect on Hgb levels can be compared, for instance, to changes in Hgb on the order of 1g/dL associated with treating anemic pregnant women with a course of iron supplementation (Sloan et al., 2002). That is, we obtain a general population effect estimate on the order of one-tenth to one-third of the effect of a targeted intervention in a highly susceptible population. Another point of comparison is the drop in Hgb during pregnancy and the first year post-partum, estimated in our sample to be on the order of 0.44 g/dL (compared to women who gave birth two or three years ago, and among women living at least 20km away from any mine). The increase in the incidence of anemia is a large effect in absolute terms, though it must be seen in the context of a baseline proportion of anemic women of 36% in control locations. That is, the cross-sectional effect amounts to an 7% relative increase in incidence, and the panel effect, to a 27% relative increase.

We note that the single difference coefficient in distance suggests that when the mine is not operational, residents of mining communities have higher Hgb levels than the control group. This is perhaps surprising, given that our wealth results showed a zero or weak negative effect in mining communities when the mine is not operational. (Table 1.4) However, it further reassures us against any concerns that geographic features, for instance altitude, might be driving cross-sectional results.

We adduce two additional tests, both to further bolster identification, and to help establish that pollution, rather than other possible causes, is the likely cause of depressed blood hemoglobin. (i) Firstly, we show that Hgb effects are only observed near mines where the combination of minerals mined suggests that lead contamination is likely to be present. (ii) Secondly, we provide direct evidence of reduced ability to recover Hgb after blood loss — an effect that is hard to reconcile with any cause other than lead toxicity.

We observe effects on hemoglobin levels only near mines where we expect heavy metal pollution

Table 1.8 shows that the effect on Hgb levels of living in mining communities are statistically zero (and mildly negative) in women living near mines where there is less reason to expect heavy metal contamination. However, in mines where there is a high likelihood of such contamination, Hgb levels are strongly and significantly depressed – by about 0.22 g/dL relative to women living farther away from the same mines, and by 0.19 g/dL compared to women living near non-heavy metal mines. (Column 3) Correspondingly, the incidence of anemia is five percentage points higher compared to women living near non-heavy metal mines (compared to women living further away from the same mines, it is six percentage points higher). (Column 4) The size of the cross-sectional effect near heavy metal mines is far closer to the panel effect than the average effect in the cross-section.[19] As noted (in Section 1.4.2.1), our definition of heavy metal mines is best thought of as a meaningful but far from perfect proxy of the presence of lead and other toxic metals. In consequence, DiD estimates are likely attenuated.

[19]A similar test is hard to construct for the panel, since mines that are potentially associated with heavy metal contamination make up a large part of the sample.

The DiD effect is robust to including interactions of the treatment dummy with region indicators (hence allaying any concerns over geographical clustering of heavy metal mines), as well as to including an interaction of the treatment with a pregnancy dummy. (Columns 5-6) We note that there is a significant negative effect of living near *any* mine in Latin America (the base category for the region interaction), perhaps due to the imperfect nature of our definition of heavy metal mines. The effect near any mine is statistically zero for the other regions.[20] We further estimate the DiD model for the asset index, and confirm that there is no differential wealth impact of living close to a heavy metal mine, as opposed to any mine. (Column 7) Finally, we do not observe similar differential effects of living near a mine associated with heavy metal contamination on two general indicators of ill health among women, namely miscarriage, and grave sickness (Columns 8-9).

The trajectory of maternal Hgb recovery after birth in mining communities corresponds with known pathophysiological patterns

The left panel in Figure 1.6 shows the pattern of recovery from blood loss during pregnancy and delivery among women living close to heavy metal mines, and those living in adjacent areas. Hgb levels conspicuously diverge during pregnancy, and stay apart during the first one and one-half years of the child's life. However, thereafter, they converge to an apparent noise pattern about a common mean. (The right panel shows the same data, with effects smoothed out for the nine months from conception to birth, and each year of the newborn's life, thereafter.) The pattern is characteristic of a pollution-induced decrease in the ability to recover Hgb after blood loss, as described in Grandjean et al. (1989) and discussed above (in Section 1.3.2.2), but not of other causes of anemia.

While the pattern is visually striking, given limited sample size, it is too strong a test to assess the difference between coefficients for the two distance groups in each individual

[20] As a further robustness check, Appendix 1.L demonstrates that the median difference between heavy metal mines and non-heavy metal mines is always at least weakly negative in each individual country for which sufficient mine-level estimates can be computed.

trimester. Instead, we test for the difference in differences between the groups across two time periods: pregnancy and the first year of the infant's life (when there is the clear impression of divergence), and the second and third years of the child's life (when there is not). The results presented in Table 1.9 show that the DiD coefficient is negative, large (0.21 g/dL), and significant. (Column 1) That is, the difference in Hgb levels between women exposed to mining and other women is far greater during and after blood loss due to pregnancy and delivery, than after some time has passed since delivery. The single difference in distance is negative, but not stable on the small sub-sample of women in the model. As expected, Hgb is dramatically lower in all women during pregnancy and in the first year post-partum.

The pattern is similar when we estimate the model with mine-level fixed effects, as shown in Column (2). Mine-level results do not always reach significance, but are as stable as the state-level results when we include controls, vary the treatment definition, or conduct placebo tests. Because of the small sample size and strong identification from the DiD setup, we prefer the state-level model. In our baseline model, we consider a postpartum period of three years. This seems more appropriate than shorter periods because the detailed time pattern of Hgb recovery shown in Figure 1.6 suggests that differences even out only in the second year of the child's life. It seems more appropriate than longer periods because the more we extend the time window, the stronger are the identifying assumptions required. Results are robust to extending the post-partum control period to four or five years; they are directionally consistent but insignificant when we shorten it to just two years. (Results not shown.)

Alternative explanations for the pattern of Hgb recovery are harder to come by than those for cross-sectional differences in Hgb levels. Because the test uses as a counterfactual women whose most recent birth lies at most three years in the past, identification requires only that the precise timing of pregnancies is ignorable within a limited time window. However, somewhat complex behavior patterns could generate the observed effect. Perhaps most simply, wealth could be associated with different child bearing choices in mining communities and control locations. For instance, it might be that wealthier women (with higher baseline Hgb levels) tend to have fewer children or space out births more in mining communities than in communities farther afield – perhaps because of better earnings opportunities. The DiD effect could then be due to comparing (relatively) poorer women in mining towns to richer controls in the pregnancy and post-partum group, and (relatively) wealthier women in mining towns to poorer controls for the following years.

To conclusively assess this concern, we first (i) note that Column (7) shows that there are no significant DiD effects on wealth. Secondly, (ii) the DiD effect is robust to controlling directly for the woman's height as a slow-moving wealth proxy, or for whether she gave birth in an 'improved' setting. (Columns 3 and 4). Finally, we (iii) show a placebo regression to test whether a similar recovery pattern emerges when we compare mothers in households in the bottom wealth quintile (placebo treatment) to those in the top quintile (placebo control). We generate two samples: a small sample designed to match the baseline sample particularly tightly, and a larger sample designed to allow for more power. Both placebo samples include women who are pregnant or have given birth within the past three years, and reside at least 20km away from the nearest mine. The small sample is restricted to observations in the same state-year pairs as those observed in the main model, and the large sample, to observations within the same survey rounds. As expected, Columns (5) and (6) show that women in poor households always have lower Hgb levels than those in wealthy households – but there is no indication of an adverse time pattern around pregnancy and postpartum, with placebo DiD coefficients either near zero, or with an opposite sign.

In summary, we obtain two DiD tests by disaggregating effects, first among mine types, and then with respect to recent pregnancy. The results are instructive both regarding mechanisms of treatment transmission and regarding identification. In terms of mechanisms, they offer strong evidence that the observed health effect is caused by pollution, not other facets of life near mines. For instance, if the observed effect on Hgb were due to iron deficiency or malaria infection, then nutritional behavior and infection rates would have to vary across distance groups in systematically different ways near metal and non-metal mines, and among pregnant and non-pregnant women – despite the fact that socio-economic outcomes do not vary in such ways. The results also provide reassurance on identification, most importantly because they are very hard to explain with sorting. Because mine types differ in health impacts, but not in wealth and non-specific health impacts, one would have to hypothesize that in their migration decisions, people not only take mine type into account, but also differentially sort on their potential health and wealth outcomes. (We have discussed above the corollary for Hgb recovery.) This would require an extraordinary level of sophistication.

Residents of mining communities are not differentially affected by causes of anemia other than lead exposure, do not bear a higher burden of disease unrelated to pollution, and are not under-served by health care

The high dimensionality of the DHS data allows for diverse falsification tests that could yield evidence against our contention that the observed hematologic effects are due to pollution, not other mechanisms. Across a range of tests, we find no such evidence.

Firstly, we show in Appendix 1.F that here is no conclusive pattern in mining communities in the leading causes of anemia other than lead toxicity (nutritional iron deficiency, malaria, and intestinal worm infections). Secondly, we test whether residents of mining communities suffer ill health that is unlikely to be attributable to pollution. Significance patterns are very sparse in the cross-section, and there are no significant adverse health impacts at all in the panel (Table 1.10). Appendix Tables 1.I and 1.J show additional specifications with similarly sparse patterns.[21]

Finally, we note that residents of mining communities are at least as well off in terms of health care as those living farther afield. As appendix 1.K demonstrates, in the long run, women are more likely to have health insurance coverage, and to give birth with some level

[21] We find no indication of greater alcohol abuse among men or women, and at most a mild indication of increased sexual risk taking, consistent with Wilson (2012). (Results available upon request.)

of skilled assistance. Panel results suggest that such benefits, along with access to health care, may extend beyond the immediate vicinity of the mine. The one potential exception to this pattern is that in our mother-level panel, we find significant decreases in the share of women who gave birth in an improved setting in mining communities when the mine was operational. The cross-sectional and mine-level panel evidence contradicts this finding. However, we mention it here because it is at odds with our otherwise consistent evidence on wealth. We note that our discussion of maternal Hgb recovery explicitly sought to exclude the potential effect of differences in maternal health care.

Patterns of anemia among children mirror those among women, but are less conclusive

Our data shows patterns of anemia among children in mining communities that resemble those found among adult women. However, significant results are hard to come by. This may be because the true treatment effect is weaker – we have noted above (Section 1.3.2.2) that children can effectively compensate for the hematologic toxicity of lead by increasing production of EPO and red blood cell production. It may also be due to small sample size (for children, we only have about half the number of observations in the women's sample). In the cross-section, we observe insignificant decreases in Hgb on the order of 0.07 g/dL (Table 1.11, Column 1); the effect is strongly concentrated near heavy metal mines, but the DiD coefficient is again not significant. (Column 2) The panel shows statistically insignificant losses from current exposure to mining, but is highly sensitive to changes in the treatment definition. (Results omitted.)

Next, we ask whether infants might be more strongly affected by pollution than older children. There are two reasons to expect this pattern. Firstly, we have attributed anemia among women – and particularly, pregnant women – in mining communities to lead exposure, and it is known that children are born with a lead burden mirroring that of their mothers. Secondly, it has been previously shown that compensatory over-production of EPO and Hgb in lead-exposed children does not quite start at birth, but at some point during infancy. (Wasserman et al., 1992) When we consider impacts on infants only, we find a larger but insignificant effect on Hgb (0.13 g/dL), relative to infants living farther away from the mine. (Column 4) However, the differential impact on infants near heavy metal mines (Column 5) is both significant and large. The triple-difference coefficient shows a 0.60 g/dL difference in Hgb levels, with a nearly identical and significant difference in differences between the effect on infants near heavy metal mines and infants near other mines. Falsification results show that infants near these mines are not indiscriminately less healthy. (Columns 7-9) However, we caution that infants born in the direct vicinity of heavy metal mines tend to live in poorer households. (Column 6) As shown above, we did not find such a correlation between mine type and wealth in our analysis of hematologic toxic effects among women living near heavy metal mines. The fact that we do find it here makes it somewhat less compelling to interpret the difference among mine types as evidence that the health impacts are due to pollution.

1.6.3 Evidence of adverse growth outcomes

As noted, exposure to environmental lead has previously been linked to decreased growth early in life. However, the evidence is mixed. In the following, we consider impacts on height for age and the incidence of stunting and severe stunting (height more than two or three standard deviations below the age-appropriate median, respectively). We find strong evidence of lower height among children exposed to a mining environment *in utero*, but also evidence of a compensatory positive growth effect of living in mining communities after birth. Appendix 1.H reports that we observe an effect on birth weight in the mother-level panel, but lack corroborating evidence from our other models.[22]

[22]Prior studies have observed that adverse conditions *in utero* can impair long-run well-being without being reflected in birth weight. (Schulz, 2010)

Without regard to time patterns of exposure, children in mining communities grow taller than their peers

In the simple cross-section, we observe *better* outcomes for height among children of less than five years of age in mining communities than in the controls. (Table 1.12, Column 1) This may not be surprising: growth is strongly linked with nutrition (both the mother's and the child's), and with greater wealth in mining communities, there may also be better diets. There is also no differential impact near 'heavy metal' mines. (Column 3)

However, the evidence is somewhat more subtle. Firstly, as Column 2 makes obvious, there is no indication of a positive effect among never-movers. Secondly, although infants are not more affected than older children when we consider *all* types of mines (Column 4), there is at least some indication of an adverse effect on infants of living near a 'heavy metal' mine. The triple-difference effects are adverse, and significant for stunting. The DiD comparing the treatment effect of closeness on infants near metal mines and other mines amounts to a loss of 0.1 standard deviations in the height measure, and a four and two percentage point increase in the incidence of stunting and severe stunting, respectively, although none of these effects reach significance. (Columns 5-7) There are no significant differences between mine types in the economic status of families with infants. (Column 8)

The cross-sectional evidence alone is thus not easy to read. There clearly are growth benefits to be had for children in mining communities, and it seems obvious to connect these to the wealth increases enjoyed by residents. However, not all children appear to benefit. The question is whether this is because some children are simply left out from economic gains, or whether they suffer countervailing direct health damage. The absence of a DiD effect between mine types and of a differential effect on infants near all mines may suggest the former. Yet, the appreciable effect on infants near mines associated with heavy metals points toward the latter. Similarly, the difference between never-movers and the general population is consistent with lower economic benefits among never-movers. (See Table 1.4.) However, since the differential in wealth effects is not very large, it is reasonable to note that children born to never-movers are also more likely to have been exposed to pollution, particularly *in utero*, through the maternal body burden of lead. We look to the panel for more conclusive evidence of the impacts of different exposure patterns.

Panel evidence shows that *in utero* exposure to mining increases the incidence of stunting

Results from the mine-level panel suggest that there is an effect of mining activity on height, that the effect is chiefly due to exposure *in utero*, and that it attenuates with age. It also allows us to at least suggest that there are genuinely positive effects of life in mining communities on growth in older children, so that children do not simply 'out-grow' *in utero* effects without further exposure, as earlier reported by Shukla et al. (1991).

The DiD effect of *in utero* exposure among all children under five years of age shows a loss of 0.14 standard deviations in the height index, and a five percentage point increase in the incidence of stunting and severe stunting. (Table 1.13, Columns 1-3) The effect on the discrete outcomes is significant; the one on the continuous measure not significant (t = 1.39), but stable. With a baseline incidence of 23% and 8%, respectively, the impact on stunting is appreciable, and the impact on severe stunting dramatic.

We next note that, in the case of the continuous index and of stunting, the effect of *in utero* exposure is larger and stronger when we estimate it for infants only. (Columns 4-6) This points either to a balancing effect – perhaps due to household wealth – in older children, or a spontaneous attenuation of *in utero* impacts with time. We shed some further light on this question by studying the effect of different exposure patterns. Thus, the estimated effect of exposure during the first year of life alone is centered near zero. (Column 7) Results when estimating *in utero* and birth-year effects jointly are more instructive. We find robust and large adverse effects of *in utero* exposure on the continuous index (0.5σ), alongside beneficial effects of birth-year exposure. (Column 8) This is at least consistent with exposure to maternal lead loads *in utero*, alongside positive effects from the socio-economic benefits of mining, once the child is born. It points less toward a mere attenuation of impacts. While it is attractive to allow *in utero* and birth-year effects to jointly enter into the model, the sample of children born just before and just after a mine opened or closed is small (conversely, operational status is highly serially correlated).[23] To further solidify the result, we therefore show that a similar pattern emerges when we first estimate separately the effect of the mine operating during the survey year (Column 9), and then compare this estimate to the one obtained when we include also the effect of the mine operating during gestation. (Column 10)

Finally, when we estimate the effects of *in utero* and birth-year exposure with mother-level effects, the results match the pattern in the mine-level panel, but do not reach significance. (Columns 11-13) This is perhaps to be expected: although we observe more than 2,000 women near mines in our sample for whom our data records child growth outcomes for at least two children born within five years of each other, there are few mothers with recorded births both while the mine was operational and while it was not operational.

1.7 Conclusion

We present the first systematic empirical assessment of the health-wealth trade-off facing mining communities, using micro-data from 44 developing countries. In communities in the vicinity of mines, we find important economic benefits, alongside serious health impacts, namely increases in the incidence of anemia in adult women, and of stunting in young children. These health impacts are consistent with exposure to lead contamination, and have previously been observed at body burdens of lead that are known also to cause cognitive deficits in children.

[23]The DHS surveys record only health data from children born no more than five years before the survey time. This helps identification, but limits sample size, in particular where we use mother-level effects.

We obtain estimates of long-run effects from a cross-sectional fixed effects model; medium-term estimates come from mine-level and mother-level panels. We confirm our wealth results with an IV approach that uses deposit location and world mineral prices to instrument for mine locations and operating times. We then develop additional difference-in-difference tests that exploit (i) the association of certain mine types with lead pollution, and (ii) known pathological patterns of Hgb recovery in adults exposed to lead. These additional tests are intended both to allow for weaker identifying assumptions, and to demonstrate that the observed health impacts are due to pollution, rather than other mechanisms.

The economic benefits to mining communities in the long run are on the order of 0.1 standard deviations of a country and year-specific asset index. Medium-term benefits to households in communities near operating mines are larger, on the order of 0.3σ. Benefits are strongly concentrated within the immediate vicinity (5km) of mines, and we find no asset wealth effects at all beyond some 15-20km. Wealth rises quite evenly across the distribution, with modest increases in inequality in the long run. Benefits in terms of health care may extend beyond the most direct vicinity of mines, although mining communities do at least as well as communities farther afield.

The evidence conclusively reveals that the real economic benefits generated in mining communities go hand in hand with increases in the incidence of anemia, by three to ten percentage points in adult women. The ability to recover hemoglobin levels after blood loss due to pregnancy and delivery is particularly impaired. There is weaker but consistent evidence of hematologic toxic effects in children. Children in mining communities are not disadvantaged in all aspects of physical growth. Yet, young children exposed to a mining environment *in utero* are more likely to be stunted or severely stunted than those born in control groups, with an increase in incidence of five percentage points. There is very limited evidence of reduced birth weight, and increases in stunting are clearly strongest among infants, and may not persist. By way of contrast to these specific health impacts, there is no general pattern of ill health in mining communities.

We conclude by highlighting some conceptual and policy implications of our results.

Firstly, the presence of adverse compensated health impacts in a generally wealthier population poses an important question. The most straightforward explanation might be to suggest that the cost of avoiding exposure to pollution is high, perhaps due to the structure of settlements and the quality of public transport. We can speculate whether the decision on living in mining towns in developing countries might resemble less the choice of an optimal distance along a continuum, and more a discrete choice between two stark options – namely living either in relatively unpolluted communities outside of a reasonable commuting distance to the mine, or in a highly polluted but bustling community adjacent to the mine. The high spatial concentration of medium-term economic benefits is certainly consistent with such a situation, as is the fact that we observe the greatest wealth effects near mines in environments that are economically less active. An alternative explanation might point to limited information. Pollutant levels near mines vary greatly, even over small distances (van Geen et al., 2012). Hence, contamination may not always be easily observed. In addition, the health impacts of pollution may not be widely known. The fact that we find strongly raised wealth levels, but only weakly better health care among households in the direct vicinity of mines at least suggests that residents are not making very decisive health investments to compensate for exposure to pollution. We also note that we find no differences in wealth across mine types, and hence, no *prima facie* indication of the kind of compensating wage differential one might expect if residents were widely aware of health risks.

Secondly, while our estimates of health cost and wealth benefits are not directly comparable without strong assumptions, we can offer some observations. Thus, (i) we have argued above that the effects of mining on asset wealth reflect meaningful differences in household welfare. Similarly, however, (ii) the cost to affected individuals of the health consequences we observe is very significant. The contemporaneous productivity loss due to anemia in adults has been estimated to be on the order of 5-17% (Horton and Ross, 2003), while the persistent economic impact of stunting can be dramatic (if childhood stunting persists through adulthood) – perhaps as large as an annual 53% loss in adult wages (Hoddinott et al., 2011). The permanent annual productivity loss due to lead-induced cognitive deficits expected at levels of PbB associated with overt anemia or stunting may be on the order of 1.6-13%.[24] (See Appendix 1.E.) At the same time, (iii) it is also clear that the health burden imposed by mining pollution is very unequally distributed: at least in our compensated reduced-form estimates, relatively small population groups are affected. In consequence, the *expected* cost of health impacts is far more modest than the steep individual cost on those afflicted. However, (iv) the cost-benefit balance tilts dramatically toward costs if economic gains are less than permanent, or if legacy effects of pollution after operations cease outlast economic benefits. This is because the health cost due to cognitive losses and stunting is permanent (and the cost due to anemia may be persistent if there are adverse legacy effects of pollution after operations cease).

In consequence, we can conclude that the decision to live in mining communities is a risky choice. Whether it is rational depends on whether economic benefits are sufficiently persistent. Furthermore, while we have shown that economic gains are quite equally distributed, the *net* benefits of mining look to be very unequally distributed. Thus, mining makes winners and losers not only between communities that benefit and communities that suffer consequences, but also *within* mining communities.

From a policy perspective, our evidence suggests that – on the global average – residents of mining communities can expect wealth benefits from the industry. (This is of course not to say that there are not instances of egregious local environmental damage and gross wealth decline.) Still, the presence of a health externality due to normal operations at mines in our sample that is observable in compensated health outcomes suggests that the management of mining pollution deserves renewed scrutiny. Our results yield three leads

[24]While we do not find strong evidence of an effect of mining on the prevalence of other health conditions recorded in our data, mining communities may obviously suffer health impacts – or enjoy health benefits – that we do not observe.

as to what effective interventions might look like. One, health concerns are most acute in the immediate vicinity of mines. Proven but expensive engineering solutions to contain and remediate pollution therefore might deserve a second look. Similarly, policy approaches need not be too broad in spatial scope to allow residents to live away from the worst pollution, while still working in or near the mine. At least for some countries in our sample, there may be a case for experimentation with programs to improve public transport, road infrastructure, or flexibility in local housing markets. Secondly, the highly uneven distribution of damages may imply that there is a premium on interventions that reduce risk. We note that the uneven distribution of costs mirrors the great spatial variation in pollution around mines described in van Geen et al. (2012), and it is tempting to posit that it might be causally related. If so, then *testing* of pollution levels in residential areas might enable residents to avoid the most dangerous sites, at a comparatively low cost. Finally, we have pointed to evidence that residents may not have full information on pollution and health risks; interventions could remedy this.

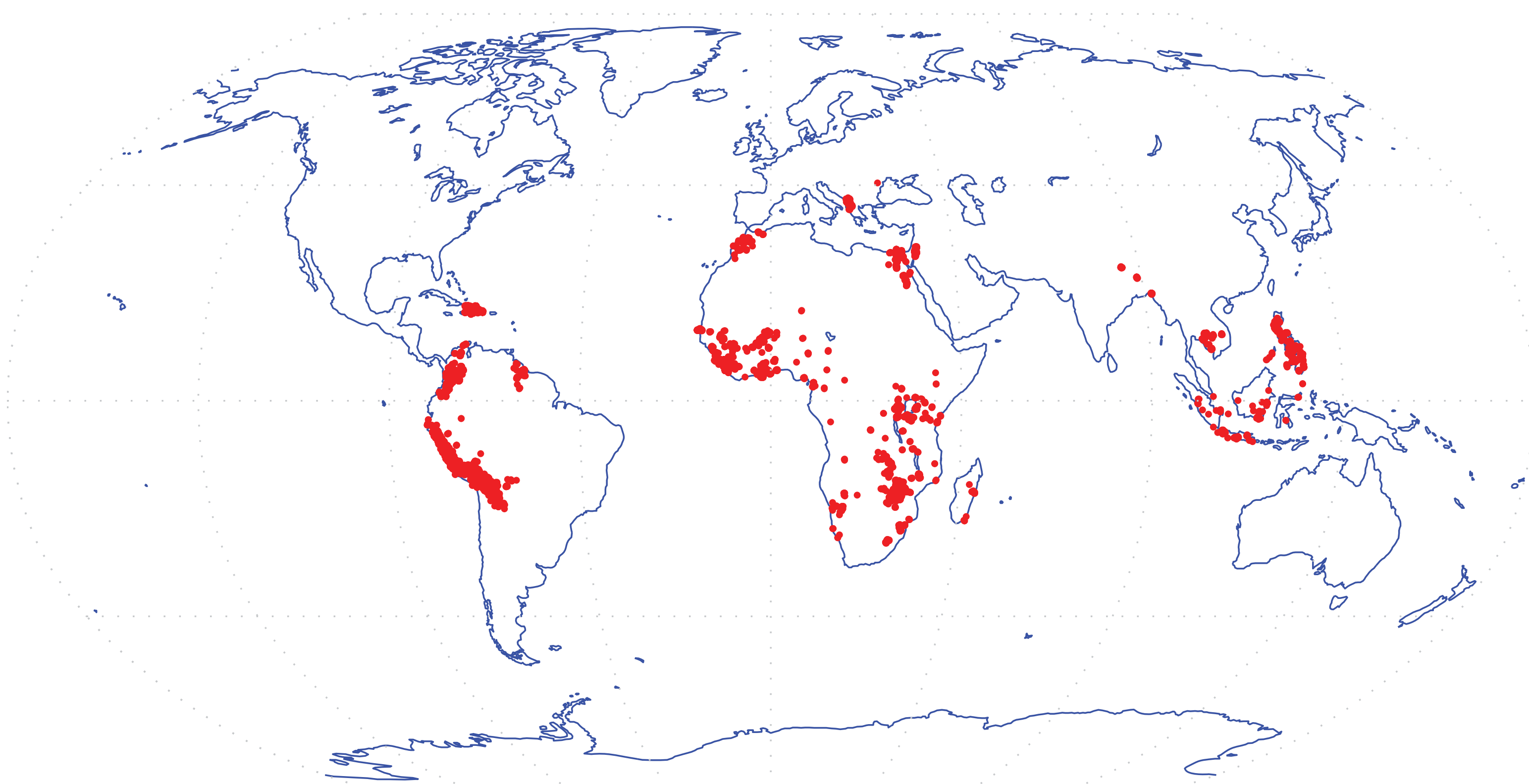


Figure 1.1: DHS clusters within no more than 20km of a mine in the sample

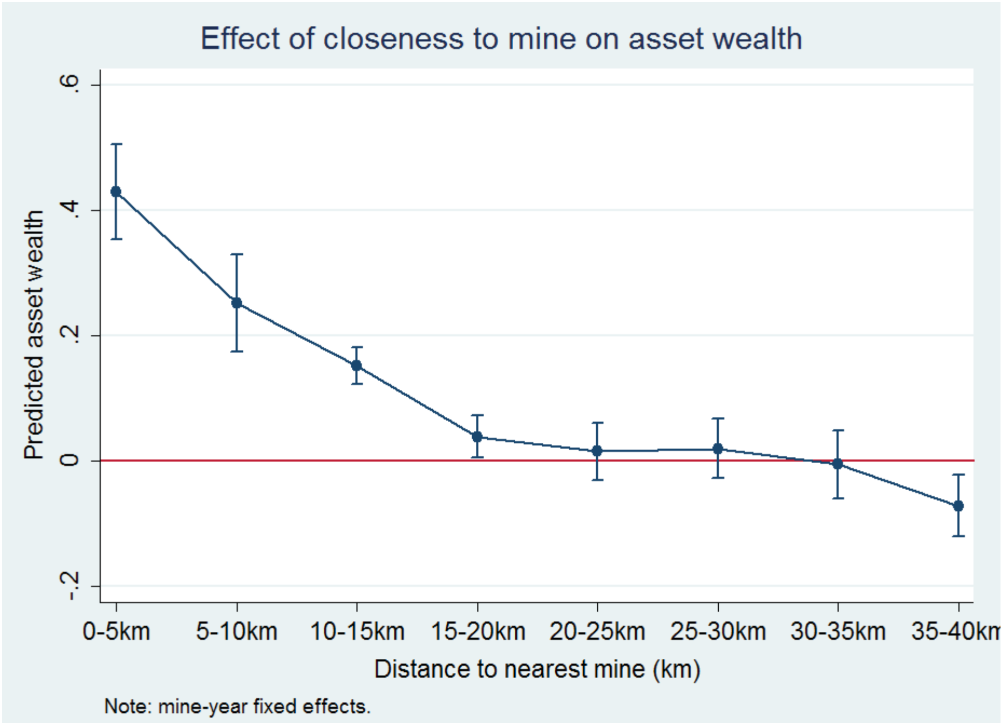


Figure 1.2: Effect of closeness to mine on asset wealth in the cross-section

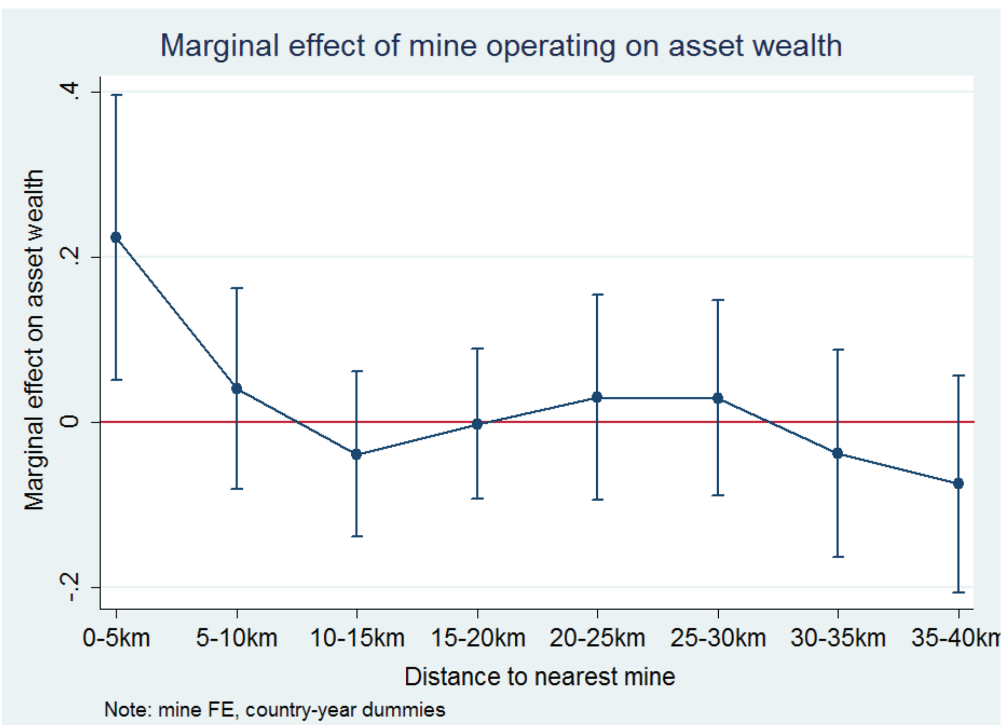


Figure 1.3: Marginal effect of mine operating on asset wealth in the panel

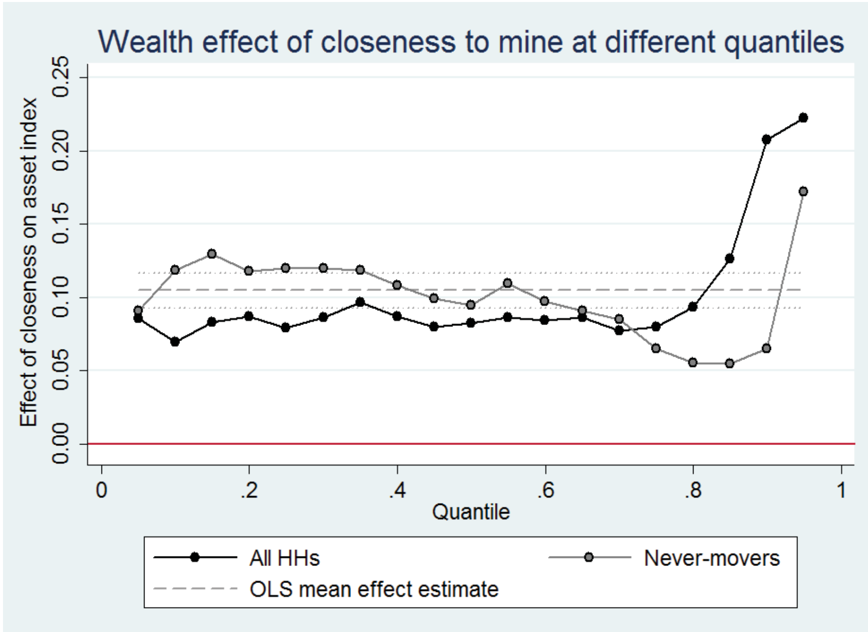


Figure 1.4: Cross-sectional effect of closeness to mine on asset wealth at different quantiles of the wealth distribution

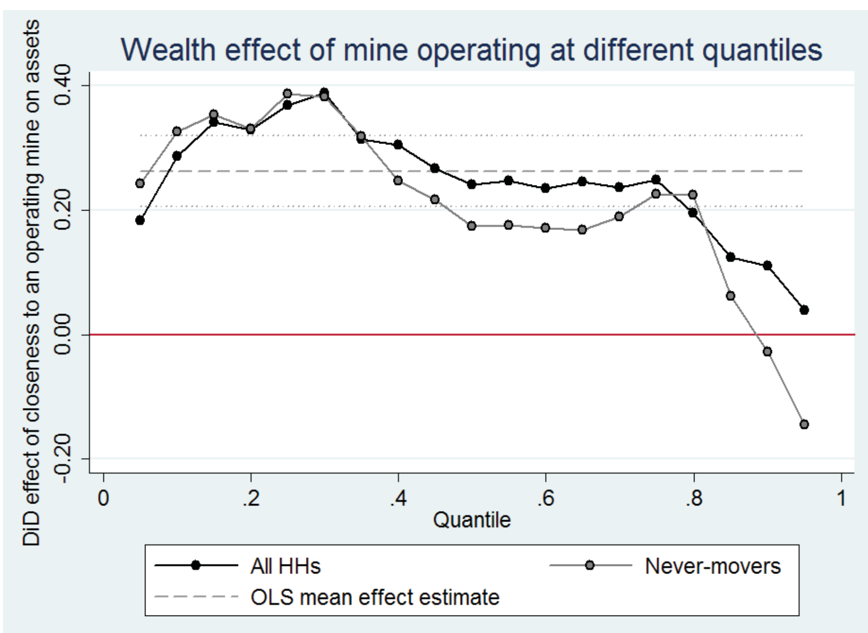


Figure 1.5: Panel effect of mine operating on asset wealth at different quantiles of the wealth distribution

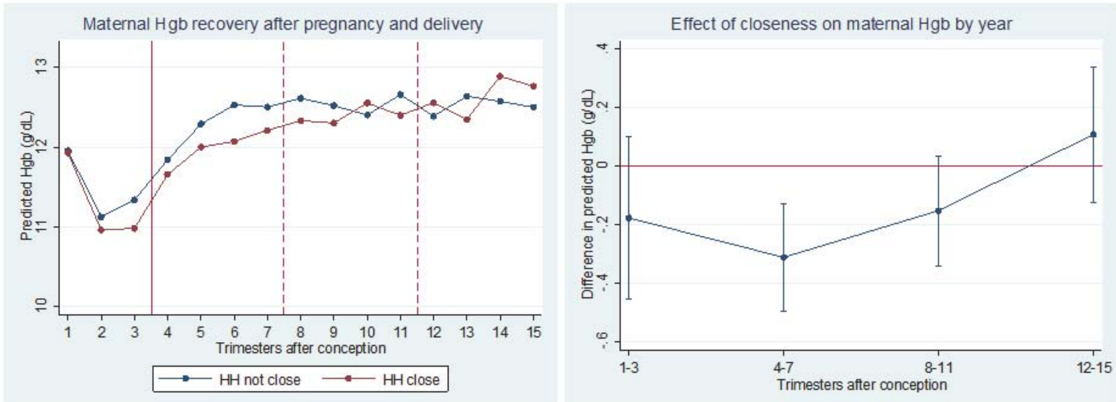


Figure 1.6: Effect of closeness to heavy metal mine on maternal Hgb recovery

Table 1.1

Mine types, associated pollutants, and health effects

Mine type	Pollutants of concern	Health effects
Polymetallic mines	Heavy metals, especially lead	Neurodevelopmental damage, anemia, growth deficits, renal disease
Small-scale gold and silver mining	Mercury	Renal disease, neurological conditions
Large-scale gold mining	Cyanide	Heart irregularities, thyroid problems
Bulk metal mines, gemstone mines	Particulates	Respiratory disease, GI problems associated with turbid water
Phosphate rock	Radionuclides	Lung cancer and non-malignant respiratory disease
Coal	Particulates, radionuclides	Respiratory disease, GI problems, lung cancer
Metal smelters	*Heavy metals,* SO_2	*As shown for polymetallic mines, and respiratory disease*

Notes. Not all mine types are mutually exclusive. Mapping based on ATSDR Toxicological Profiles for the respective pollutants, Alloway (2013), Ripley (1996), and Wright and Welbourn (2002). Health effects as reported from chronic low-level environmental exposure.

Table 1.2
Prior literature on blood lead levels in communities near smelters

	Distance to smelter	Mean PbB
Fontúrbel et al., 2011	0.5-1.8km	n/a
Roels et al., 1980	1-2.5km	13-30 μg/dL
Recio-Vega et al., 2012	2km	14-19 μg/dL
Factor-Litvak et al., 1999	2-4km	28-39 μg/dL
Benin et al., 1999	3km	20-40 μg/dL
Landrigan and Baker, 1981	4km	*≥ 40 μg/dL in 87% of subjects*

Notes. The table summarizes prior studies of lead levels in communities near smelters. It shows the maximum distance between the smelter and the communities considered highly exposed, alongside mean blood lead in highly exposed communities. Ranges of mean PbB refer to means for population groups that differ in age, gender, and other characteristics. Incidence for Landrigan and Baker summarized by the authors. In the case of Benin et al. (1999), PbB was predicted from observed environmental pollution; in all other studies, PbB was measured directly.

Table 1.3

Sample size

	Surveys with observations within 20km of a mine
Survey rounds	104
Countries	44
Interview years	25

	Number of households		
	Full sample	Within 5km of a mine	Within 5-20km of a mine
Households	1,192,492	37,608	132,797
% of total		*3.2%*	*11.1%*
Children under five years of age	1,364,156	31,964	121,519
Women aged 15 and over	2,877,024	87,234	310,096
Men aged 15 and over	2,717,928	82,973	294,723

	Mines and smelters near DHS sampling clusters		
	USGS data	RMD data	Infomine data
DHS cluster within 20km	838	508	7
DHS cluster within 0-5km	339	225	4
DHS cluster within 5-20km	687	455	6
DHS cluster in both distance categories	226	172	3

Notes. Sample size based on all types of mines, smelters and legacies, excluding quarries. Not all variables used in this study are available for the entire sample. The count of locations from Infomine includes only those mines not covered in the RMD data.

Table 1.4

Effects on mean asset wealth and wealth disparities

	Mean asset wealth				Mean absolute wealth deviation			
	All HHs	Never-movers	All HHs	Never-movers	All HHs	Never-movers	All HHs	Never-movers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HH close to mine	**0.105*****	**0.0784***	-0.113	-0.0352	**0.0274***	**0.00746**	-0.0451	-0.0492
	(0.035)	**(0.0423)**	(0.089)	(0.0892)	**(0.0156)**	**(0.0151)**	(0.0362)	(0.0493)
Mine operating			-0.0296	-0.0585*			-0.00611	-0.0303
			(0.0348)	(0.0354)			(0.0300)	(0.0255)
Mine operating * HH close (DiD)			**0.262*****	**0.173***			**0.0384**	**0.0223**
			(0.0958)	**(0.0897)**			**(0.0380)**	**(0.0505)**
Number of households	90,319	31,079	22,579	9,459	90,319	31,079	22,579	9,459
Number of groups	1,562	1,371	218	205	1,562	1,371	218	205
R-squared	0.094	0.081	0.13	0.141	0.010	0.004	0.029	0.034

Notes. The table reports estimates of equations (1) and (2). Cross-sectional estimates in columns (1-2) and (5-6) use indicator variables for each mine-year pair as fixed effects. Panel estimates in columns (3-4) and (7-8) use mine-level indicators for area fixed effects, and country-year indicators for time effects. The dependent variable in Columns (1-4) is the asset factor index, with units expressed in standard deviations. In Columns (5-8), it is the absolute deviation of a household's asset factor index from the sampling cluster mean, in units of standard deviations. Controls include a quadratic in the household head's age and an indicator for urban/rural status. Columns marked 'Never-movers' restrict the sample to households with at least one respondent who had always been resident in the current location at survey time. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.5
Instrumental variables estimates of the effect on asset wealth in the cross-section and in the panel

	Cross-section		Panel				
					IV robustness checks		
	OLS benchmark	IV	FE benchmark	IV	Full sample	Baseline with all deposits	Baseline with smelters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HH close	0.0758**	0.0925**	-0.191**	-0.288*	-0.242	-0.284	-0.530**
	(0.0322)	(0.0448)	(0.0962)	(0.153)	(0.167)	(0.182)	(0.257)
Mine operating in survey year			-0.0514	0.115	0.267	0.101	-0.00301
			(0.0442)	(0.150)	(0.251)	(0.146)	(0.172)
HH close * mine operating in survey year			**0.370***	**0.524**	**0.465**	**0.653***	**0.728***
			(0.110)	**(0.208)**	**(0.232)**	**(0.396)**	**(0.281)**
Number of households	126,434	126,434	14,671	14,671	14,735	20,593	19,780
Number of groups	618	618	187	187	188	253	200
R-squared	0.181		0.157				
First-stage F statistic		77.45		10.5	2.54	7.33	3.28
Mine-level first stage relationship		0.686***		0.215***	0.174***	0.217***	0.189***
		(0.0265)		(0.0447)	(0.0361)	(0.0445)	(0.0422)
Number of observations		1616		487	491	490	523
R-squared		0.64		0.041	0.042	0.042	0.034

Notes. The table shows cross-sectional and panel IV estimates of wealth effects, alongside fixed effects benchmark estimates computed for the IV sample, and first-stage relationships. Column (2) shows cross-sectional IV estimates following Equation (6); the instruments is an indicator recording whether there is a mineral deposit within 5km of a sampling cluster. Column (1) shows results from the cross-sectional baseline model in Equation (1), estimated on the same sample. For the panel, Columns (4-7) show IV results as in Equation (7). The instrument is an index of the world market price of minerals produced at a given mine, weighted by the preceding year's production. Column (3) shows the benchmark. The dependent variable is the asset index, expressed in units of standard deviations. The baseline panel IV sample in column (4) includes all mines and deposits that are in a stage of exploration or development. It excludes one outlier with an extremely high value of the instrument. Columns (5-7) show IV estimates using modified samples. Column (5) includes the outlier; Column (6) includes a broader range of un-mined deposits; Column (7) includes smelters, despite the poor first-stage fit. All regressions include a quadratic in the household head's age, and an indicator for urban/rural status. The cross-sectional regressions include state-year indicators, and the panel, mine fixed effects and country-year indicators. In addition to first-stage F statistics, we show simple mine-level (as opposed to household-level) OLS results on the first stage relationship between the presence of a deposit and the presence of a mine in the cross-section, and between the price index and operational status in the panel. Standard errors are clustered at the state level in the cross-section, and at the mine level in the panel. Significant at * 10%, ** 5%, *** 1%.

Table 1.6

Correlation of mine-level wealth effects with measures of development

Country log GDP	-0.0928*	State average years of education	-0.0794**
	(0.0493)		(0.0328)
Number of mines	228	Number of mines	135
R-squared	0.015	R-squared	0.043
State inverse distance to coast	-1.004**	State power line density (log)	-0.127**
	(0.483)		(0.0622)
Number of mines	137	Number of mines	137
R-squared	0.031	R-squared	0.030
Travel time to nearest city	0.0885	Access to land is an obstacle	0.504
	(0.0599)		(0.488)
Number of mines	137	Number of mines	66
R-squared	0.016	R-squared	0.033
State institutional quality	-0.771	Country completed an EITI report	1.500**
	(0.781)		(0.644)
Number of mines	70	Number of mines	228
R-squared	0.023	R-squared	0.039
Country ever participated in EITI	0.122		
	(0.0898)		
Number of mines	228		
R-squared	0.023		

Notes. The dependent variable in the table is composed of mine-level estimates of the cross-sectional effect of closeness to a mine on asset wealth, obtained by estimating equation (1) for each mine separately. (See Appendix E for details.) The regressors either record characteristics of the country in which the mine is located, during the year in which the survey was taken, or of the state in which the mine is located, during the year closest to the survey time for which data was available. All regressions also include log GDP. OLS estimates with conventional standard errors are shown throughout. We find no correlations with the World Bank's CPIA, and omit results for conciseness. Conventional standard errors. Significant at * 10%, ** 5%, *** 1%.

Table 1.7
Hematologic toxic effects on women

	Altitude-adjusted hemoglobin (g/dL)			Anemia		
	All HHs	Never-movers	All HHs	All HHs	Never-movers	All HHs
	(1)	(2)	(3)	(4)	(5)	(6)
HH close to mine	**-0.0863****	**-0.131**	0.396***	**0.0262****	**0.0495***	-0.107***
	(0.0438)	**(0.0838)**	(0.146)	**(0.0126)**	**(0.0268)**	(0.0292)
Mine operating			0.0852			-0.0277
			(0.136)			(0.0309)
Mine operating * HH close (DiD)			**-0.330***			**0.0966****
			(0.173)			**(0.0390)**
Number of women	38,217	13,506	9,845	36,225	13,204	9,845
Number of groups	934	785	122	934	784	122
R-squared	0.0001	0.001	0.007	0.0003	0.001	0.006

Notes. The table reports estimates of equations (1) and (2). Cross-sectional estimates in columns (1-2) and (4-5) use mine-year fixed effects, while panel estimates in columns (3) and (6) use mine-level indicators as area fixed effects, and country-year indicators as time effects. Columns (1-3) show effects on hemoglobin levels at survey time among adult women; columns (4-6) show results for the incidence of anemia (defined as Hgb below 12 g/dL in non-pregnant women, and Hgb below 11 g/dL in pregnant women). Controls include a quadratic in the respondent's age, and an indicator for urban/rural status. Columns (2) and (5) restrict the sample to respondents who had always been resident in the current location at survey time. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.8

Hematologic toxic effects on women near different mine types

	Benchmark		Effect near 'heavy metal' mines		Additional interactions		Falsification tests		
	Hgb (g/dL)	Asset index	Hgb (g/dL)	Anemia	Hgb (g/dL)		Asset index	Miscarriage	Grave illness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
HH close to mine	-0.0863**	0.145**	-0.0317	0.0123	-0.161*	-0.0285	0.140*	0.00347	0.00366
	(0.0438)	(0.0576)	(0.0533)	(0.0155)	(0.0886)	(0.0507)	(0.0756)	(0.00519)	(0.00595)
HHs close to a 'heavy metal' mine (DiD)			**-0.192****	**0.0466***	**-0.253*****	**-0.192****	**0.0176**	**-0.00377**	**-0.00286**
			(0.0944)	**(0.0247)**	**(0.0876)**	**(0.0902)**	**(0.101)**	**(0.0109)**	**(0.00898)**
Additional interactions					Region	Pregnancy			
Number of women	38,217	25,695	38,217	36,225	38,217	36,225	25,695	117,118	11,022
Number of groups	934	932	934	934	934	934	932	1,469	151
R-squared	0.0001	0.111	0.001	0.0004	0.001	0.027	0.111	0.061	0.011

Notes. Columns (1) and (2) show estimates of equation (1), and columns (3-9) estimates of equation (4), using indicators for each mine-year pair as fixed effects. In columns (3-9), the treatment variable is interacted with an indicator recording whether there is a high expectation of environmental contamination with heavy metals at the nearest mine. Columns (1), (3), (5), and (6) show effects on hemoglobin levels at survey time among adult women. Columns (2) and (7) show effects on asset wealth in households in which respondents in the health sample live. Column (4) shows results for the incidence of anemia (defined as Hgb below 12 g/dL in non-pregnant women, and Hgb below 11 g/dL in pregnant women). Columns (8) and (9) show effects on the incidence of two health conditions not specific to lead exposure. The dependent variable in column (8) is an indicator for whether a woman of reprductive age has ever suffered a miscarriage; in column (9), it is an indicator for whether the respondent was gravely ill for three months or more in the year preceding the survey. Where the dependent variable is a health condition, controls include a quadratic in the respondent's age at survey time and an indicator for urban/rural status. Where the dependent variable is the asset index, a quadratic in the household head's age replaces the respondent's age. Columns (5) and (6) show results from including additional interactions of the treatment variable, as indicated. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.9

Recovery of maternal Hgb after birth near heavy metal mines

	Altitude-adjusted hemoglobin (g/dL)						Asset index
	Baseline	Mine-year group effects	Control for height-for-age	Control for delivery setting	Placebo treatment (small sample)	Placebo treatment (large sample)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pregnancy and infancy	-0.558***	-0.581***	-0.547***	-0.429***	-0.470***	-0.652***	-0.0644**
	(0.0421)	(0.0375)	(0.0436)	(0.0454)	(0.0771)	(0.0621)	(0.0255)
HH close to mine	-0.0277	-0.00274	-0.0143	-0.0639			0.0304
	(0.0791)	(0.101)	(0.0839)	(0.0860)			(0.0523)
Pregnancy and infancy * HH close (DiD)	**-0.253****	**-0.185***	**-0.262****	**-0.330*****			**-0.0122**
	(0.0982)	**(0.107)**	**(0.104)**	**(0.116)**			**(0.0562)**
Placebo - HH in lowest wealth quintile					-0.169*	-0.232***	
					(0.0862)	(0.0551)	
Pregnancy and infancy * placebo					-0.0231	0.0783	
					(0.0825)	(0.0538)	
Number of women	5,004	5,004	4,700	3,928	6,851	14,857	4,892
Number of groups	167	521	161	158	139	269	167
R-squared	0.045	0.044	0.043	0.037	0.028	0.031	0.125

Notes. The table shows estimates of equation (5), using indicators for state-year pairs as fixed effects in all columns except column (2). Column (2) shows results using mine-year indicators. The dependent variable is Hgb at survey time among adult women in columns (1-6), and the asset index in column (7). In all columns, the sample consists of observations near mines where heavy metal contamination is to be expected. Where Hgb is the dependent variable, the sample is restricted to women who are currently pregnant, or have given birth within the three years preceding the survey, and who are known to have been resident in the current location since conception. In column (7), the sample is restricted to households in which the women included in the regression in column (1) reside. Columns (5) and (6) give results from a placebo regression, in which the treatment variable is replaced with a placebo indicator that takes value one if the respondent's household is in the bottom wealth quintile, and value zero if it is in the top wealth quintile. In column (5), the placebo sample is restricted to women who are pregnant or have given birth in the past three years, but live in households at least 20km from the nearest mine, and are observed in state-year pairs also represented in the sample in (1). In column (6), it is restricted to women who are pregnant or have given birth in the past three years, live at least 20km from the nearest mine, and are observed in country-year pairs also represented in the sample in (1). Controls include a quadratic in the respondent's age at survey time and an indicator for urban/rural status in columns (1-6). In addition, the model in column (3) includes the respondent's height for age z-score, and that in column (4) includes an indicator for whether she most recently gave birth in an improved setting. In column (7), a quadratic in the household head's age replaces that in the respondent's age. Standard errors are clustered at the mine level in column (2), and at the state level, otherwise. Significant at * 10%, ** 5%, *** 1%.

Table 1.10

Health outcomes not specifically linked to heavy metal pollution

	Panel A: Child health outcomes - cross-section				
	Infant mortality	Under-five mortality	Diarrhea	Cough	Fever
	(1)	(2)	(3)	(4)	(5)
HH close to mine	**-0.00246**	**-0.00305**	**0.0112***	**0.00480**	**0.00191**
	(0.00223)	**(0.00270)**	**(0.00579)**	**(0.00963)**	**(0.00788)**
Number of children	298,373	298,373	61,567	60,305	59,494
Number of groups	1,566	1,566	1,510	1,503	1,384
R-squared	0.002	0.003	0.029	0.007	0.01

	Panel B: Child health outcomes - panel				
	Infant mortality	Under-five mortality	Cough	Diarrhea	Fever
	(1)	(2)	(3)	(4)	(5)
Mine operating in exposure period * HH close (DiD)	**-0.00499**	**-0.00819**	**0.00392**	**-0.00260**	**-0.0234**
	(0.00745)	**(0.00864)**	**(0.0299)**	**(0.0282)**	**(0.0258)**
Exposure period	In utero	In utero	Survey year	Survey year	Survey year
Number of observations	43,057	43,057	15,325	15,449	15,576
Number of mines	259	259	236	237	230
R-squared	0.003	0.006	0.025	0.034	0.021

	Panel C: Adult health outcomes - cross-section			
	Ever miscarried	Night blindness during pregnancy	Female respondent very sick	Male respondent very sick
	(6)	(7)	(8)	(9)
HH close to mine	**0.00263**	**0.00254**	**0.00328**	**0.0120**
	(0.00460)	**(0.0104)**	**(0.00527)**	**(0.00977)**
Number of respondents	117,118	29,317	11,022	9,808
Number of groups	1,469	1,185	151	151
R-squared	0.061	0.001	0.011	0.011

	Panel D: Adult health outcomes - panel			
	Ever miscarried	Night blindness during pregnancy	Female respondent very sick	Male respondent very sick
	(6)	(7)	(8)	(9)
Mine operating in exposure period * HH close (DiD)	**-0.00236**		**-0.00845**	
	(0.0152)		**(0.0119)**	
Exposure period	Survey year	n/a	Survey year	n/a
Number of observations	29,666		4,111	
Number of mines	202		63	
R-squared	0.065		0.005	

Notes. The table reports estimates of equation (1) in the rows marked 'cross-section', and estimates of equation (2) in the rows marked 'panel'. In the latter, treatment variables are defined using the time period of exposure to pollution most appropriate to each health condition, as indicated. Only the difference in differences coefficient is reported. Cross-sectional models use indicator variables for each mine-year pair as group fixed effects; panel models, mine fixed effects and survey round dummies. The dependent variable in columns (1) and (2) is an indicator for whether a child died within the first year and the first five years after birth, respectively. In the other columns, it is an indicator for whether the respondent suffered the condition indicated - over the two weeks preceding the survey (3-5); at any point during her reproductive life (6); during the most recent pregnancy (7); or for three months or more during the year preceding the survey (8-9). Controls in columns (1-5) include an indicator for urban/rural status in all columns, a quadratic in the mother's age at birth, an indicator for gender, birth-order indicators, as well as indicator variables for the child's age (columns 3-5 only). In columns (6-9), they include an urban/rural indicator, and a quadratic in the respondent's age at survey time. In cells marked 'n/a', the model could not be estimated. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.11
Hematologic toxic effects on children

	Effects on all children under five years of age		Effects on infants					
	Cross-section	Mine-type DiD	Cross-section	Mine-type DiD	Mine-type DiD falsification tests			
	Hgb (g/dL)	Hgb (g/dL)	Hgb (g/dL)	Hgb (g/dL)	Asset index	Diarrhea	Cough	Fever
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HH close to mine	-0.0653	-0.0418	-0.0487	-0.0464	0.179***	0.00739	-0.00328	0.00196
	(0.0460)	(0.0516)	(0.0505)	(0.0569)	(0.0647)	(0.00732)	(0.0116)	(0.00923)
HHs close to a 'heavy metal' mine (DiD)		-0.109		-0.0184	0.0887	-0.000835	0.0492**	0.00500
		(0.108)		(0.119)	(0.115)	(0.0139)	(0.0216)	(0.0192)
Child in infancy			-0.358***	-0.359***	0.0318	0.0339***	0.0107*	0.0132*
			(0.0598)	(0.0651)	(0.0240)	(0.00579)	(0.00613)	(0.00699)
HH close to mine, and child in infancy			-0.0823	0.0643	-0.0175	0.0120	-0.00890	-0.00822
			(0.102)	(0.110)	(0.0617)	(0.0134)	(0.0120)	(0.0134)
Nearest mine (≤ 20km) is a 'heavy metal' mine, and child in infancy				0.00959	0.0850	-0.0198*	0.00249	0.00593
				(0.155)	(0.0750)	(0.0112)	(0.0119)	(0.0126)
HH close to a 'heavy metal' mine, and child in infancy (triple difference)				**-0.597*****	**-0.309****	**-0.0163**	**-0.0212**	**-0.0122**
				(0.199)	**(0.147)**	**(0.0240)**	**(0.0262)**	**(0.0234)**
Number of children	18,029	18,029	18,029	18,029	12,697	61,567	60,305	59,494
Number of mines	907	907	907	907	901	1,510	1,503	1,384
R-squared	0.068	0.068	0.021	0.022	0.120	0.006	0.002	0.002
DiD effect on infants				**-0.616*****	**-0.221**	**-0.0171**	**0.028**	**-0.0072**
				(0.201)	**(0.158)**	**(0.0232)**	**(0.0283)**	**(0.0261)**

Notes. Columns (1) and (3) show estimates of equation (1); columns (2) and (4-8) show estimates of equation (4). In column (3), the treatment variable in equation (1) is interacted with an indicator for whether the child was in her first year of life at survey time. In columns (4-8), the treatment variable and its interaction in equation (4) are interacted with the indicator variable for infancy. All columns indicators for each mine-year pair as fixed effects. Columns (1-4) show effects on hemoglobin levels at survey time. Columns (5-8) show effects on asset wealth and health conditions not specific to lead exposure; in columns (6-8), the dependent variables record whether a child suffered from the respective condition in the two weeks preceding the survey. Where the dependent variable is a health condition, controls include an indicator for urban/rural status in all columns, a quadratic in the mother's age at birth, an indicator for gender, birth-order indicators, as well as indicator variables for the child's age. Where the dependent variable is the asset index, controls include a quadratic in the household head's age, and an urban/rural indicator. The row labeled "DiD effect on infants" shows the sum of the coefficients "HHs close to a 'heavy metal' mine" and "HH close to a 'heavy metal' mine, and child in infancy", and tests the hypothesis that the sum is zero. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.12

Growth effects on children in the cross-section

	All children under five years of age			Infants				
	Benchmark	Never-movers only	Effect near 'heavy metal' mines	Benchmark	Effect on infants near 'heavy metal' mines			
	Height for age	Height for age	Height for age	Height for age	Height for age	Stunting	Severe stunting	Asset index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HH close to mine	0.0828**	-0.00153	0.0754*	0.0760*	0.0598	-0.0116	-0.00884	0.111**
	(0.0389)	(0.0497)	(0.0430)	(0.0418)	(0.0464)	(0.0103)	(0.00742)	(0.0469)
HHs close to a 'heavy metal' mine			0.0378		0.0726	-0.0263	-0.00668	-0.0384
(DiD)			(0.0988)		(0.102)	(0.0241)	(0.0163)	(0.0929)
Child in infancy				0.805***	0.781***	-0.164***	-0.0717***	0.0156
				(0.0368)	(0.0432)	(0.0112)	(0.00675)	(0.0156)
HH close to mine, and child in infancy				0.0304	0.0719	-0.00459	0.00597	0.00945
				(0.0528)	(0.0631)	(0.0136)	(0.00824)	(0.0403)
Nearest mine (≤ 20km) is a 'heavy metal'					0.102	-0.0571***	-0.0175	-0.0263
mine, and child in infancy					(0.0729)	(0.0217)	(0.0108)	(0.0317)
HH close to a 'heavy metal' mine, and					**-0.170**	**0.0696****	**0.0280***	**-0.0710**
child in infancy (triple difference)					**(0.110)**	**(0.0327)**	**(0.0166)**	**(0.0693)**
Number of children	40,552	16,927	40,552	40,552	40,552	40,552	40,552	28,540
Number of groups	1,244	1,041	1,244	1,244	1,244	1,244	1,244	1,243
R-squared	0.064	0.054	0.064	0.062	0.062	0.036	0.015	0.084
DiD effect on infants					**-0.097**	**0.043**	**0.021**	**-0.109**
					(0.131)	**(0.029)**	**(0.018)**	**(0.107)**

Notes. Columns (1), (2) and (4) show estimates of equation (1), and columns (3) and (5-8), estimates of equation (4). In column (4), the treatment variable in equation (1) is interacted with an indicator for whether the child was in her first year of life at survey time. In columns (5-8), the treatment variable and its interaction in equation (4) are interacted with the indicator variable for infancy. All models use indicators for each mine-year pair as fixed effects. Columns (1-5) show effects on height for age z-scores. The dependent variable in Column (6) is the prevalence of stunting, defined as a height of two σ or more below the median; in Column(7), it is the prevalence of severe stuting, defined as a height of more than three σ below the median. Column (8) shows effects on asset wealth in households in which children in the height-for-age sample live. Where the dependent variable is a health condition, controls include an indicator for urban/rural status in all columns, a quadratic in the mother's age at birth, an indicator for gender, birth-order indicators, as well as indicator variables for the child's age. Where the dependent variable is the asset index, controls include a quadratic in the household head's age, and an urban/rural indicator. The row labeled "DiD effect on infants" shows the sum of the coefficients "HHs close to a 'heavy metal' mine" and "HH close to a 'heavy metal' mine, and child in infancy", and tests the hypothesis that the sum is zero. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.13

Comparative growth effect of in utero and birth-year exposure in the panel

	In utero exposure			In utero exposure, infants only			In utero vs. birth year exposure		In utero vs. survey-year exposure		In utero and birth year exposure, mother fixed effects		
	Height	Stunting	Severe stunting	Height	Stunting	Severe stunting	Height		Height		Height		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
HH close to mine	0.191**	-0.0501**	-0.0424***	0.535**	-0.126***	-0.0448	0.0972	0.141	0.0743	0.137			
	(0.0808)	(0.0251)	(0.0110)	(0.209)	(0.0472)	(0.0346)	(0.0902)	(0.0896)	(0.107)	(0.0963)			
Mine operating during pregnancy	0.0212	0.000151	-0.00102	0.0930	-0.0369	-0.0135		0.0950		0.0953	0.175		0.0855
	(0.0614)	(0.0178)	(0.0124)	(0.124)	(0.0307)	(0.0141)		(0.0712)		(0.0826)	(0.161)		(0.173)
Mine operating during pregnancy * HH close	**-0.136**	**0.0534***	**0.0496*****	**-0.371***	**0.149*****	**0.0512**		**-0.494*****		**-0.417****	**-0.460**		**-0.422**
	(0.0981)	**(0.0274)**	**(0.0140)**	**(0.217)**	**(0.0510)**	**(0.0346)**		**(0.139)**		**(0.185)**	**(0.361)**		**(0.359)**
Mine operating in second exposure period							-0.00443	-0.0843	-0.00185	-0.0908		0.197	0.169
							(0.0595)	(0.0692)	(0.0546)	(0.0922)		(0.154)	(0.172)
Mine operating in second exposure period * HH close							-0.00860	0.423***	0.0474	0.356*		-0.163	-0.0540
							(0.110)	(0.149)	(0.130)	(0.189)		(0.424)	(0.418)
Second exposure period							Birth year		Survey year			Birth year	
Number of children	11,629	11,629	11,629	2,426	2,426	2,426	11,629	11,629	11,155	11,321	11,629	11,629	11,629
Number of fixed effects	200	200	200	186	186	186	200	200	188	191	9,408	9,408	9,408
R-squared	0.113	0.072	0.055	0.091	0.100	0.058	0.113	0.114	0.073	0.117	0.204	0.204	0.205

Notes. Columns (1-10) report estimates of equation (2), and Columns (11-13), estimates of equation (3). Columns (1-10) use indicators for each country-year pair as time effects, and columns (11-13) use country linear time trends. In columns (1-6) and (11), treatment is defined as exposure to mining *in utero*. Columns (7-10) and (12-13) compare this to the effect of exposure during a second exposure period, as indicated. The dependent variable in columns (1), (4), and (7-13) is the height-for-age z-score. In columns (2) and (5), it is the prevalence of stunting, defined as a height of two σ or more below the median; in Columns (3) and (6), it is the prevalence of severe stuting, defined as a height of more than three σ below the median. Controls include an indicator for urban/rural status in all columns, a quadratic in the mother's age at birth, an indicator for gender, birth-order indicators, as well as indicator variables for the child's age. For consistency across models, the sample is restricted to those observations where the operating status of the mine is known both in the birth year and during gestation (this removes about 3% of observations). Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Appendix 1.A: Notes on data sources and processing

DHS surveys in sample

Albania (2009)
Angola (2007, 2011)
Bangladesh (2000, 2004, 2007, 2011)
Bolivia (2008)
Burkina Faso (1993, 1999, 2003, 2010)
Burundi (2010)
CAR (1995)
Cambodia (2000, 2005, 2010)
Cameroon (1991, 2004, 2011)
Colombia (2010)
DR Congo (2007)
Côte d'Ivoire (1994, 1999)
Dominican Republic (2007)
Egypt (1992, 1995, 2000, 2003, 2005, 2008)
Ethiopia (2011)
Ghana (1993, 1998, 2003, 2008)
Guinea (1999, 2005)
Guyana (2009)
Haiti (2000, 2006)
Indonesia (2003)
Jordan (2002, 2007)
Kenya (2003, 2009)
Lesotho (2004, 2009)
Liberia (1986, 2007, 2009, 2011)
Madagascar (1997, 2009, 2011)
Malawi (2000, 2004, 2010, 2012)
Mali (1996, 2001, 2006)
Moldova (2005)
Morocco (2004)
Mozambique (2009)
Namibia (2000, 2007)
Nepal (2001, 2006, 2011)
Niger (1992, 1998)
Nigeria (1990, 2003, 2008, 2010)
Peru (2000, 2005)
Philippines (2003, 2008)
Rwanda (2005, 2008, 2010)
Senegal (1993, 1997, 2005, 2009, 2011)
Sierra Leone (2008)
Swaziland (2007)
Tanzania (1999, 2004, 2008, 2010, 2012)
Togo (1988, 1998)
Uganda (2001, 2006, 2009, 2011)
Zambia (2007)
Zimbabwe (1999, 2006, 2011)

Definition of quarries

We exclude from the analysis all mines that are best characterized as quarries. As noted, this is because we posit that quarries are sufficiently different from mines in both their economic importance and as a source of emissions to warrant treatment as a separate type of entity. Because of their economic importance, we choose to include gemstone mines in our analysis; however, we exclude mines that produce semi-precious stones.

More precisely, we define as a quarry any location where exclusively any combination of the following materials (as defined in the USGS data) is being produced: abrasive, ball clay, bentonite, brick clay, bromine, calcium, cement rock, clay, diatomite, dolomite, feldspar, fire clay, flagstone, fluorine-fluorite, fullers earth, garnet, granite, gypsum-anhydrite, halite, kaolin, kyanite, limestone, magnesite, marble, mica, mineral pigments, olivine, peat, perlite, pumice, quartz, rock asphalt, salt, sand and gravel, semi-precious stones, silica, staurolite, stone, talc-soapstone, vermiculite, travertine, volcanic materials, wollastonite, zeolites.

World metals and minerals price data

Where available, we obtain metal and mineral prices from the World Bank's Global Economic Monitor commodities data (World Bank, 2013). We add additional price series from UNCTAD (2013) (manganese ore and tungsten), and the IMF (2013) Primary Commodity Prices (iron ore and yellowcake uranium oxide). Not all metals and minerals are traded in exchanges; for those where a market price is not easily observed, we obtain aggregated transaction-level price data. We use transaction-level data for minor platinum-group metals from Johnson Matthey, the metal traders (2013), and for all other metals and minerals not covered in the sources listed above, from the U.S. Geological Survey (Kelly and Matos, 2013). We omit diamond mines from our IV analysis, since it has been argued that "there are no internationally set prices for rough diamonds . . . [and] the market prices for rough natural diamonds are almost constantly in a state of flux." (Natural Resources Canada, 2009) Prices are generally given either for units of processed metal, or units of metal content in ore. The exceptions are coal, iron ore, phosphate rock and potash, where prices are given for the raw product. We align price units with production units in our data accordingly. Where necessary, we deflate price data using the U.S. CPI published by the U.S. Bureau of Labor Statistics.

Further notes on the panel treatment definition

As noted in the main body of the paper, we do not impute mining activity in our production data, not even tacitly, by contrasting observations before and after an opening date. We do impute an absence of activity under the following restrictive conditions: we assume an absence of activity for five years prior to a mine opening date, if (i) the opening date is recorded clearly in the data, (ii) the recorded date is no more than three years earlier than the first year in which production data is available, and (iii) it is not the case that production data reflects an ambiguous start date. We consider the start of production to be ambiguous if production is reported as missing during the opening year, is known to have been zero in the year before, and is known to have been non-zero in the year after.

Further notes on the time-varying instrument

In the panel, we instrument for the current operating status of a mine using a weighted price index. Because mines typically extract several minerals, we define the price index as the world market price for each mineral produced in year $t - \tau$, weighted by the share of minerals in the previous year's production, at $t - \tau - 1$. To account for the large difference in price levels across minerals, we normalize price in the year 2005 to one. For years before the first year of production, we weight prices by the average production shares during the mine's subsequent production history; for years after the final year of production, we weight by the final year; for years in between production years, we weight by production shares in the most recent year of observed production.

Appendix 1.B: Construction and interpretation of the asset index

To obtain a sound measure of wealth in the absence of data on consumption, expenditure, or income, we compute a standard index of asset and housing characteristics, as in Filmer and Pritchett (2001). Because our asset data includes many dummy variables, we follow Sahn and Stifel (2003) in using a factor index in our main specification, rather than the more well-known principal-component index; empirically, the differences are slight.

The index is based on the largest set of assets and housing characteristics available *within each survey round*. That is, the information used in the index varies between countries and survey years. We choose this approach because (i) the set of assets recorded varies greatly across survey years, so that working with the largest common set would discard a great deal of information, and because (ii) in our very heterogenous sample, defining impacts relative to the variation in wealth within a given country and year seems more appealing than defining them relative to global variation.

We include any asset for which data is available for 90% of those households for which at least one woman answered the women's questionnaire. Empirically, little changes if we strike a different balance between data availability and richness of information.

The maximum set of variables included is the following:

Housing characteristics

- Dummy indicating whether the household has: a kitchen, a chimney.
- Dummies recording whether the dwelling uses 'rudimentary' or 'finished' building materials (as opposed to the omitted category of 'natural' building materials) for floors, walls, and the roof. The categories are country-specific and intuitive. For instance, in the 1986 Liberia survey, 'natural' roof materials were thatch and grass; the 'rudimentary' material was sheet metal; and 'finished' materials were concrete, asphalt, or asbestos.

Assets

- Share of children of no more than 15 years of age in the household who have: blankets, shoes, clothes;
- Dummy for whether the household owns any number of each of the following items: phone landline, mobile phone; stove other than open fire, electricity connection, refrigerator; radio, TV; watch, bank account; bicycle, motorbike, car.

To illustrate how the index relates to ownership of individual assets, Table 1.B shows factor loadings for those survey rounds used to illustrate results in the main body of the paper.

Table 1.B

Asset index - examples of factor loadings

	Peru 2000				Burkina Faso 2010			
	Factor loading	Mean	SD	Factor x SD	Factor loading	Mean	SD	Factor x SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Assets and housing characteristics								
Landline	0.07	0.18	0.38	0.18	0.09	0.02	0.15	0.61
Mobile phone					0.08	0.64	0.48	0.17
Electricity	0.10	0.67	0.47	0.22	0.22	0.14	0.34	0.63
Refrigerator	0.10	0.31	0.46	0.23	0.13	0.04	0.20	0.68
Radio	0.03	0.86	0.35	0.09	0.05	0.72	0.45	0.10
TV	0.13	0.68	0.47	0.27	0.21	0.17	0.38	0.55
Watch					0.04	0.41	0.49	0.09
Bank account					0.12	0.14	0.34	0.35
Bicycle	0.02	0.24	0.43	0.04	0.00	0.88	0.33	0.01
Motorbike	0.02	0.04	0.20	0.09	0.10	0.39	0.49	0.20
Car	0.03	0.08	0.28	0.11	0.07	0.02	0.13	0.56
Floor rudimentary	-0.01	0.08	0.27	-0.04				
Floor finished	0.10	0.45	0.50	0.20	0.13	0.47	0.50	0.27
Roof rudimentary					-0.01	0.03	0.16	-0.09
Roof finished	0.30	0.38	0.49	0.62	0.11	0.58	0.49	0.23
Walls rudimentary	-0.48	0.64	0.48	-1.00	-0.06	0.22	0.41	-0.15
Walls finished					0.11	0.43	0.49	0.21
Asset index								
Mean		0.00				0.00		
Standard deviation		0.98				0.94		
Minimum		-1.26				-1.15		
Maximum		1.84				3.77		
Number of households		19,452				11,329		

Appendix 1.C: Weighted cross-sectional results

In the main paper, we work throughout with unweighted regressions. This is (i) because due to sample size limitations, we cannot obtain the mine-by-mine coefficients needed for weighted estimates for all outcome variables, and (ii) because given the data structure, even the best weighted scheme is ultimately incorrect.

To gain some insight into how our unweighted estimates should be interpreted, we compare them to alternative weighting schemes in the case of the impact of mining on one outcome, namely asset ownership. The asset index is available for nearly all households in our sample. In the cross-section, we therefore have enough observations near many mines to estimate the effect of closeness separately, mine by mine. This offers an opportunity to compare our baseline unweighted estimates to estimates obtained by giving equal weight to each mine ('mine-weighted' estimates), and by computing a population-weighted average of mine-level coefficients ('population-weighted' estimates).

Table 1.C shows four treatment effect estimates obtained by using different weighting approaches, alongside the baseline estimate. Column (1) replicates the unweighted baseline estimate shown in Table 1.4, and Column (2) shows that the unweighted estimate is very similar for the sub-sample of mines for which separate estimates can be generated. Similar results emerge from pooled estimation using naïve sampling weights – namely, the original sampling weights given in the DHS data, re-scaled to account for population and sample size in the different surveys that make up our pooled sample. (Column 3)

Coefficients given in Column (4) are averages of mine-wise estimates of Equation (1), pooled with *equal weights* for each mine-year, and with standard errors as described in Deaton (1997). This is nearly a consistent estimate of the mine-weighted effect, except for the fact that each mine-level estimate is obtained from an unweighted regression over households. Finally, Column (5) shows the average of mine-wise estimates of Equation (1), pooled using population weights for each mine-year. Because sampling is stratified at the cluster level, not the mine level, we construct mine-level weights from the sum of cluster-level weights. We then use these in pooling coefficients. The resulting coefficient is intended as an estimate of the population-weighted treatment effect. It is, however, inconsistent both because the mine-level regression is unweighted (as in Column 4), and also because the population weight generated from the aggregate of cluster weights is ultimately not correct.

Both the mine-weighted and population-weighted estimates (Columns 4 and 5) are larger than the unweighted estimates; they are highly significant despite being generated by an inefficient process. Thus, while they serve as a reminder that our unweighted baseline estimates are inherently somewhat vulnerable, the weighted results confirm that there are strong and significant positive wealth effects in mining communities. Finally, we note that weighting by population, rather than applying equal weight for each mine, increases the point estimate. *Prima facie*, this suggests that treatment effects will tend to be larger in larger communities – at least among the sample of mines with dense enough population in the vicinity to allow for mine-level estimates.

Table 1.C

Weighted estimates of cross-sectional asset wealth effects

	All HHs				
	Benchmark: unweighted estimate	Unweighted estimate for mines included in mine-level estimation	Naïve weights, re-scaled to global sample	Equal mine-year weights	Mine-year sampling weights
	(1)	(2)	(3)	(4)	(5)
HH close to mine	0.105***	0.0914***	0.0963***	0.193***	0.295***
	(0.035)	(0.0104)	(0.0199)	(0.0345)	(0.0513)
Number of households	90,319	51,598	50,743		
Number of groups	1562	306	306	306	306

Notes. Columns (1-3) show estimates of equation (1), using mine-year fixed effects. Columns (4) and (5) show estimates obtained by computing equation (1) for observations near each mine separately, and then summing them as described in Appendix C. Controls include a quadratic in the household head's age, and an indicator for urban/rural status. Standard errors obtained following Deaton (1997). Significant at * 10%, ** 5%, *** 1%.

Appendix 1.D: Measurement error in geolocation

Visual inspection in high-resolution satellite images available on Google Earth of the geolocations recorded for mines in the USGS and RMD data raises some doubts about the precision of the marker positions. At times, the impression is that the marker is at some distance from visible mine features. This visual quality check is far from perfect: images available online may not have been taken at a time when the mine was operational; underground mines and small mines may not be readily visible in satellite images at any time; in many other instances, mines are very large complexes, and there is little intuition as to where the correct marker location ought to be; and in a substantial number of cases, mines are geographically clustered, and it is impossible to identify individual operations without substantial research.

However, to assess potential measurement error concerns, we benchmark geolocations obtained from the RMD business intelligence data against those recorded in an additional, entirely independent dataset (*Mining Atlas*), for the common subset of mines. To implement this test, we manually match mines from the two datasets on their name, the minerals mined, and the country in which the mine is located. Where necessary, we consult additional information, such as company records, to confirm the merge. The merge would seem to be more reliable for large mines with unusual names producing some of the more rarely mined minerals in countries where there are few mines (for instance, the Langer Heinrich uranium mine in Namibia), but less accurate without substantial additional research for mines with common names producing common metals in countries with very many mines (consider the Santa Rosa polymetallic mine in Peru). Because of this concern, we make no attempt to benchmark locations in the very large USGS dataset used in the cross-section.

As noted in the main body of the paper, in our baseline data, we drop a handful of mines where RMD geolocations are obviously erroneous, or for which the RMD and Mining Atlas geolocations are 40km or more apart. Conditional on excluding these mines, the sampling cluster distances to the nearest mine generated by the two datasets are strongly correlated, with an apparent white noise error pattern. For mines that ever appear in our analysis, the mean absolute discrepancy in distance to the nearest mine is 4.7km (with a median of 2.5km). We conclude that we should expect some attenuation bias in our results.

With two noisy, but plausibly independent measures of distance in hand, we can in the cross-section use closeness to mines as measured by one data set to instrument for closeness to mines in the other, and hence, correct measurement error.[1] Table 1.D shows results from this approach for the asset index. Column (1) mirrors the baseline cross-sectional result shown in Table 1.4, but obtained using using state-year fixed effects.[2] Columns (2) and (3) show results using closeness as measured by RMD geolocations and Mining Atlas geolocations, respectively, for clusters in the vicinity of those mines for which geolocation data is available from both sources. As is evident, the point estimates are very close to the benchmark, as well as to each other. Columns (4) and (5) show IV estimates for the sub-sample of mines present in both datasets. As expected in the presence of measurement error, both point estimates are larger than the fixed effect estimates in columns (2) and (3), though they are not significantly different.

We further note that the ratio between the OLS and IV estimates is about 0.85 for the RMD data, and 0.61 using the Mining Atlas data. Asymptotically at least, we would therefore conclude that distance as measured using RMD geolocations is a less noisy measure of true distance than the measure derived from Mining Atlas geolocations (Filmer and Pritchett, 2001). This reassures us in our choice of RMD as the basic data source. We conclude that there is evidence of measurement error in mine locations, and of resulting attenuation

[1] We use production information only from one source of geolocation data, and hence, cannot implement a similar approach in the panel.

[2] When we use mine-year effects, results are empirically very similar. However, allowing for mine-year effects necessitates a choice in the IV models of which dataset mine-year effects should be based on, and forcing a choice would seem to run counter to the spirit of instrumenting with one noisy measure for another.

bias. Bias is considerable, between 18% and 65% in the two specifications we estimate. However, our preferred RMD-based estimate exhibits the lower level of bias, and in any case, the corrected estimates are not so different from our baseline estimates as to substantially change our interpretation of wealth patterns in mining communities.

Table 1.D
Robustness to measurement error in geolocation

	Asset index				
	Benchmark (RMD - full sample)	RMD - common sample	Mining Atlas - common sample	Using Mining Atlas to instrument for RMD	Using RMD to instrument for Mining Atlas
	(1)	(2)	(3)	(4)	(5)
HH close to mine	0.105***	0.0983***	0.0969***	0.116**	0.160**
	(0.0314)	(0.0306)	(0.0324)	(0.0553)	(0.0645)
Number of households	90,319	43,985	42,190	39,073	39,073
Number of groups	554	440	435	426	426
R-squared	0.152	0.181	0.176		
First stage IV coefficient				0.712***	0.665***
				(0.0506)	(0.0576)

Notes. The table shows estimates of equation (1) in columns (1-3), and additional IV estimates in columns (4-5). Column (1) mirrors the benchmark estimate given in Table 5. Columns (2) and (3) show estimates using measures of sampling cluster distance to the nearest mine as given by the two mine location datasets, respectively. The sample is restricted to those mines where locations are recorded in both datasets. (The number of groups is not equal, because some sampling clusters are within 20km of the nearest mine in one of the datasets, but not in the other.) Columns (4) and (5) show results obtained by instrumenting for closeness as measured in one of the datasets by closeness as measured by the other. Controls include a quadratic in the household head's age, and an indicator for rural/urban status. Standard errors are clustered at the state level. IV standard errors come from cluster bootstraps using 400 repetitions. Significant at * 10%, ** 5%, *** 1%.

Table 1.E
Health impacts due to lead exposure and their functional consequences

	Health impacts on adults	Health impacts on children			
	Anemia	Anemia		Cognitive deficits	Stunting
Previously observed threshold PbB levels	≥ 50μg/dL	≥ 40μg/dL		No apparent threshold	≥ 10μg/dL
Descriptive impacts	Listlessness, reduced focus and ability to perform physical work.	Listlessness, reduced focus and ability to perform physical work.	Poor cognitive development secondary to anemia.	Poor cognitive development as a direct consequence of lead toxicity.	Lower adult height and body mass, adverse birth outcomes in stunted women. Poor cognitive development.
Estimated functional consequences in terms of productivity loss (%) in adulthood	5-17% (a)	2.5% (b)	1.2% (c)	1.6-13% (d)	54% (e)
Disability weights (% DALY per year)	0.5-5.8% (f)	0.5-5.8% (f)	2.4%	2.4-36% (g)	0.2-5.3% (h)
Direct impact used in imputing productivity loss	n/a	0.5-1.5σ decrement in performance on Bayley Scales of Infant Development (i)	1.73 IQ point decrease with 1g/dL decrease in Hgb (j)	0.3-2.7σ loss in IQ with a move from 2μg to 20μg PbB (k)	n/a
Sources	Horton (2003), Salomon et al. (2013)	Horton (2003), Salomon et al. (2013)	Stoltzfus et al. (2004), WHO (2004)	Lanphear (2005), Rau (2013), WHO (2004)	Dewey and Begum (2011), Hoddinott et al. (2011), Ricci et al. (2006), WHO (2004)

Notes. Threshold PbB levels from ATSDR (2007) and Lanphear (2005). (a) Range of direct productivity estimates from RCTs, as reported in Horton (2003). Estimates relate to the impact of iron-deficiency anemia; iron deficiency may impact productivity through channels orther than anemia. (b) Imputed productivity loss, as reported in Horton (2003). (c) Authors' calcuation of productivity loss due cognitive losses secondary to anemia. We assume mean Hgb levels in anemic and non-anemic children as observed in our sample. In addition, we follow Horton (2003) in assuming a correlation of 0.62 between childhood and adult IQ , and a wage decrease of about eight percent associated with a one-standard deviation decrement in adult IQ. (d) Authors' calculation of productivity loss due to a move from 2μg to 20μg PbB. Additional assumptions as in (c). (e) Direct productivity estimate from an RCT reported in Hoddinott (2011). (f) Range refers to weights for mild and moderate anemia. (g) Range refers to low weights for cognitive impairment secondary to anemia, and high weights for "mild metal retardation attributable to lead exposure". (h) Range refers to low weights for stunting secondary to protein-energy malnutrition, and high weights obtained by attributing a share of DALYs lost to diarrheal disease to the secondary effect on stuntintg. (i) Range of outcomes from RCTs that provided iron supplements. (j) Estimate from a meta-analysis of observational studies. (k) Lanphear et al. report a semi-log dose-response function; Rau et al. report a linear relationship. We transform them to directly compare the impact of moving from a background level of lead exposure to a level in between the thresholds previously associated with overt anemia or stunting. A standard deviation is assumed to be equivalent to 19 IQ points, as in Lanphear (2005).

Table 1.F

Differential diagnosis - effect of closeness to mines on causes of anemia other than lead exposure

	Iron deficiency			Malaria		Worms
	Youngest child given meat or eggs	Youngest child given iron-rich vegetables	Iron pills during most recent pregnancy	Tested positive for malaria	No malaria drugs during most recent pregnancy	Worm pills during recent pregnancy
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cross-section - full sample					
HH close to mine	-0.0224*	-0.0141	-0.00163	-0.0399	-0.0134	0.00848
	(0.0129)	(0.0161)	(0.00920)	(0.0274)	(0.0131)	(0.0251)
Number of observations	17,639	17,639	43,521	3,384	19,636	14,491
	Panel B: Cross-section - women's Hgb sample					
HH close to mine	-0.0134	-0.0341	-0.0156	-0.00253**	-0.00562	-0.0204
	(0.0261)	(0.0226)	(0.0181)	(0.000970)	(0.0187)	(0.0188)
Number of observations	6,223	6,223	12,678	1,615	7,125	6,478
	Panel C: Mine type DiD - full sample					
HH close to a 'heavy metal' mine	-0.0245	-0.00362	0.0190	-0.0142	0.00118	-0.00272
	(0.0307)	(0.0394)	(0.0273)	(0.0169)	(0.0271)	(0.0348)
Number of observations	17,647	17,647	37,387	1,879	18,566	13,755
	Panel D: Mine type DiD - women's Hgb sample					
HH close to a 'heavy metal' mine	-0.0131	0.0156	-0.000880	-0.00333***	-0.0358	0.0461*
	(0.0505)	(0.0501)	(0.0415)	(0.000197)	(0.0394)	(0.0270)
Number of observations	6,223	6,223	12,678	1,615	7,125	6,478
	Panel E: Panel - full sample					
HH close * mine operating in treatment period	-0.0533	-0.0342	0.0680		0.0104	-0.0207
	(0.0747)	(0.0678)	(0.0490)		(0.0757)	(0.0497)
Number of observations	3,893	3,893	7,529		4,012	2,452
	Panel F: Panel - women's Hgb sample					
HH close * mine operating in treatment period	-0.0238	0.0628	-0.0169		0.0498	-0.0699
	(0.0851)	(0.0431)	(0.0696)		(0.0997)	(0.0768)
Number of observations	1,648	1,648	3,039		1,758	1,115

Notes. The table shows estimates of equations (1), (2), and (4), as indicated. Each equation is estimated separately for the entire sample, and for the sub-sample for which we observe womens' Hgb levels. In each case, we report only the coefficients of interest, as indicated. Fixed effects, time effects, and covariates are as in the preferred Hgb models reported in the main body of the paper. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.G
Comparative effects on core outcomes near mines and smelters

	Asset index		Women's altitude-adjusted Hgb (g/dL)		Children's height for age score	
	Mines/legacies	Smelters	Mines/legacies	Smelters	Mines/legacies	Smelters
			A: Cross-section			
HH close to mine	0.109***	0.0765	-0.0948*	-0.0648	0.0822*	0.0910
	(0.0413)	(0.0666)	(0.0553)	(0.0691)	(0.0446)	(0.0762)
Number of observations	77,569	12,750	31,304	6,913	34,776	5,776
			B: Mine type DiD			
HH close to mine	0.111**	0.0294	-0.0502	0.121	0.0896*	-0.0147
	(0.0457)	(0.0960)	(0.0581)	(0.0953)	(0.0487)	(0.0837)
HH close to a 'heavy metal' mine	-0.0172	0.0870	-0.523**	-0.244**	-0.0556	0.241
	(0.102)	(0.119)	(0.224)	(0.121)	(0.128)	(0.156)
Number of observations	77,569	12,750	31,304	6,913	34,776	5,776
			C: Panel			
HH close to mine	-0.101	-0.110***	0.189	0.537***	0.195**	-0.0633
	(0.0978)	(0.00686)	(0.155)	(0.00895)	(0.0865)	(0.131)
Mine operating in treatment period	-0.0498	-0.110***	0.264*	0.222***	0.0535	-1.774***
	(0.0434)	(0.00353)	(0.144)	(0.0233)	(0.0604)	(0.240)
HH close * mine operating in treatment period	0.322***	0.183***	-0.159	-0.429***	-0.207**	0.178
	(0.107)	(0.0216)	(0.215)	(0.0552)	(0.0943)	(0.195)
Number of observations	15,688	6,891	6,680	3,165	8,726	2,903

Notes. The table shows estimates of equation (1) in the uppermost panel, estimates of equation (4) in the middle panel, and estimates of equation (2) in the bottom panel. Fixed effects and covariates are as in the preferred models in the main body of the paper. In columns (1), (3), and (5), the sample is limited to observations near mines or legacies, while the other columns, it is limited to observations near smelters. Standard errors are clustered at the mine/smelter level. Significant at * 10%, ** 5%, *** 1%.

Table 1.H

Effects on birth weight (in grams) in children under five years of age

	Panel A: Cross-sectional results				
	Benchmark	Never-movers only	Effect near 'heavy metal' mines	Effect on infants near any mine	Effect on infants near 'heavy metal' mines
	(1)	(2)	(3)	(4)	(5)
HH close to mine	7.456	-8.686	0.411	0.711	-4.651
	(11.75)	(19.32)	(13.37)	(12.73)	(14.23)
HH close to a 'heavy metal' mine			16.50		24.27
			(31.54)		(32.68)
Child in infancy				-8.653	-9.879
				(7.081)	(7.857)
HH close and child in infancy				13.12	22.12
				(15.39)	(18.52)
Nearest mine (≤ 20km) is a 'heavy metal' mine, and child in infancy					6.222
					(17.04)
HH close to a 'heavy metal' mine, and child in infancy					-34.98
					(33.97)
Number of observations	38,165	13,651	36,978	36,978	36,978

	Panel B: Panel results			
	Mine-level	Mine-level	Mother-level	Mother-level
	(6)	(7)	(8)	(9)
HH close to mine	-19.23	-28.72		
	(32.89)	(31.87)		
Mine operating during pregnancy	60.32**		103.2*	
	(24.80)		(61.67)	
HH close and mine operating during pregnancy	15.29		-246.7*	
	(36.43)		(144.9)	
Mine operating in birth year		62.54***		51.55
		(21.99)		(47.89)
HH close and mine operating in birth year		28.09		80.48
		(34.60)		(135.6)
Number of observations	10,559	10,559	10,559	10,559

Notes: The table shows estimates of equations (1), (2), and (4), as indicated. Fixed effects, time effects, and covariates are as in the corresponding results reported for child growth outcomes in the main body of the paper. The dependent variable is the birth weight in grams. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.I

Additional falsification tests - child health outcomes

	Infant mortality	Under-five mortality	Diarrhea	Cough	Fever
	(1)	(2)	(3)	(4)	(5)
			Panel A: Cross-section		
HH close to mine	-0.00246	-0.00305	0.0112*	0.00480	0.00191
	(0.00223)	(0.00270)	(0.00579)	(0.00963)	(0.00788)
Number of children	298,373	298,373	61,567	60,305	59,494
			Panel B: Cross-section - never-movers		
HH close to mine	-0.00711**	-0.00974***	0.00709	0.0292*	0.00898
	(0.00288)	(0.00369)	(0.0124)	(0.0159)	(0.0156)
Number of children	110,764	110,764	22,732	22,192	21,272
			Panel C: Cross-section - differential impact on infants		
HH close to mine and child in infancy			0.00756	-0.0143	-0.0109
			(0.0113)	(0.0107)	(0.0112)
Number of children			61,567	60,305	59,494
			Panel D: Mine-type DiD		
HH close to a 'heavy metal' mine	0.00147	0.00381	-0.00658	0.0434**	0.00128
	(0.00426)	(0.00548)	(0.0123)	(0.0208)	(0.0186)
Number of children	298,373	298,373	61,567	60,305	59,494
			Panel E: Mine-type DiD - differential impact on infants		
HH close to a 'heavy metal' mine * child in infancy			-0.0163	-0.0212	-0.0122
			(0.0240)	(0.0262)	(0.0234)
Number of children			61,567	60,305	59,494
			Panel F: Mine-level panel		
Exposure period	In utero	In utero	Survey year	Survey year	Survey year
Mine operating in exposure period * HH close	-0.00499	-0.00819	-0.00260	0.00392	-0.0234
	(0.00745)	(0.00864)	(0.0282)	(0.0299)	(0.0258)
Number of children	43,057	43,057	15,449	15,325	15,576
			Panel G: Mother-level panel		
Exposure period	In utero	In utero	Survey year	Survey year	Survey year
Mine operating in exposure period * HH close	-0.0108	-0.00514	-0.140	-0.0965	-0.114
	(0.0214)	(0.0248)	(0.125)	(0.0945)	(0.111)
Number of children	43,057	43,057	15,989	16,113	16,370

Notes. The first panel reports results from Equation (1); the following panel reports estimates from the same equation, with the sample restricted to never-movers; and the panel labled 'differential impact on infants' shows the coefficient on the interaction of the treatment in equation (1) with an indicator variable for whether a child was in infancy. The panel labeled 'mine-type DiD' reports estimates of equation (4); the following panel shows estimates of the same equation, with an additional interaction term of the DiD variable with an indicator for infancy. The mine-level and mother-level results are estimates of equations (2) and (3), respectively. All models include fixed effects and controls as in Table 13 in the main paper. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.J
Additional falsification tests - adult health outcomes

	Ever miscarried	Female respondent very sick	Night blindness during pregnancy	Male respondent very sick
	(1)	(2)	(3)	(4)
Panel A: Cross-section				
HH close to a mine	0.00263	0.00328	0.00254	0.0120
	(0.00460)	(0.00527)	(0.0104)	(0.00977)
Observations	117,118	11,022	29,317	9,808
Panel B: Cross-section - never-movers				
HH close to a mine	-0.00433	0.0155*	0.00946	0.0149
	(0.00781)	(0.00880)	(0.0173)	(0.0131)
Observations	49,817	4,459	11,701	3,716
Panel C: Mine-type DiD				
HH close to a 'heavy metal' mine	-0.00377	-0.00286	0.0310	0.0129
	(0.0109)	(0.00898)	(0.0216)	(0.0154)
Observations	117,118	11,022	29,317	9,808

Notes. The first panel reports results from question (1); the panel labeled 'never-movers' reports estimates from the same equation, with the sample restricted to households that reported never having moved from their current place of residence. The panel labeled 'mine-type DiD' reports estimates of equation (4). All models were estimated using mine-year indicators as group fixed effects, and include controls for the the respondent's age and urban/rural status of the sampling cluster. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.K.1

Effects on health insurance coverage

	Panel A: Health insurance coverage of women					
	Overall coverage			Employer-provided coverage		
	Cross-section		Panel	Cross-section		Panel
	All HHs	Never-movers		All HHs	Never-movers	
	(1)	(2)	(3)	(4)	(5)	(6)
HH close to mine	0.0282**	0.0518***	0.0260**	0.0206***	0.0276***	-0.0015
	(0.0127)	(0.0130)	(0.0125)	(0.00769)	(0.00789)	(0.0123)
Mine operating			0.191			0.0168**
			(0.12)			(0.00658)
Mine operating * HH close (DiD)			0.0571**			0.0255
			(0.028)			(0.0252)
Number of respondents	35,971	13,014	9,183	17,876	3,849	4,597
Number of groups	638	540	105	231	140	68
R-squared	0.003	0.005	0.021	0.002	0.004	0.008

	Panel B: Health insurance coverage of men					
	Overall coverage			Employer-provided coverage		
	Cross-section		Panel	Cross-section		Panel
	All HHs	Never-movers		All HHs	Never-movers	
	(7)	(8)	(9)	(10)	(11)	(12)
HH close to mine	0.00826	0.0287		0.00600	0.00995	
	(0.0225)	(0.0264)		(0.0133)	(0.0136)	
Mine operating						
Mine operating * HH close (DiD)			n/a			n/a
Number of respondents	6,978	1,601		6,148	994	
Number of groups	180	86		159	68	
R-squared	0.002	0.007		0.001	0.002	

Notes. The table reports estimates of equation (2) in columns (3) and (6), and estimates of equation (1), otherwise. Cross-sectional estimates use indicators for each mine-year pair as group fixed effects; panel estimates use mine-specific area effects, and country-year indicators as time effects. The panel cannot be estimated for the models in equations (9) and (12). All regressions control for a quadratic in the respondent's age, and an indicator for rural/urban status. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Table 1.K.2
Effects on access to health care

Panel A: Maternal and infant health care

	Delivery in improved setting			Vitamin A supplement		Time of first antenatal visit (months)		Number of antenatal visits		Post-natal check on mother		Post-natal check on child	
	Cross-section	Panel		Cross-section	Panel	Cross-section	Panel	Cross-section	Panel	Cross-section	Panel	Cross-section	Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
HH close to mine	0.0265**	-0.0147		0.00711	-0.0398	0.00151	-0.0999	0.0141	-0.0162	0.00153	0.00630	0.00210	0.00332
	(0.0130)	(0.0290)		(0.0143)	(0.0352)	(0.0417)	(0.115)	(0.0599)	(0.160)	(0.0151)	(0.0479)	(0.0278)	(0.0552)
Mine operating		0.0588***	0.0410		0.00185		-0.159*		0.336**		-0.0104		-0.0200
		(0.0217)	(0.0346)		(0.0276)		(0.0897)		(0.131)		(0.0222)		(0.0354)
Mine operating * HH close (DiD)		0.0130	-0.0917**		0.0235		0.0413		0.182		0.0268		-0.0500
		(0.0319)	(0.0447)		(0.0363)		(0.129)		(0.195)		(0.0483)		(0.0584)
Number of respondents	62,410	15,796	15,796	37,235	9,818	44,767	11,854	43,900	11,143	34,129	7,926	14,410	3,127
Number of groups	1,506	256	12,945	1,236	226	1,498	255	1,498	253	996	193	746	157
R-squared	0.021	0.106	0.024	0.000	0.062	0.005	0.050	0.008	0.087	0.003	0.038	0.002	0.092

Notes. Columns labeled 'Cross-section' show estimates of equation (1), using indicators for each mine-year pair as group fixed effects. Column (3) shows estimates of equation (3), using indicators for each mother as group fixed effects, and country linear time trends. All other columns marked 'Panel' report estimates of equation (3) using mine-level area effects, and indicators for each country-year pair as time fixed effects. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Panel B: Reasons for not seeking medical attention

	No knowledge		No money		Distance		No provider		No drugs	
	Cross-section	Panel	Cross-section	Panel	Cross-section	Panel	Cross-section	Panel	Cross-section	Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HH close to mine	0.0187*	0.0351	-0.00576	-0.0208	0.00104	0.0211	0.00134	0.112	-0.0114	0.0542
	(0.0103)	(0.0307)	(0.0105)	(0.0439)	(0.0114)	(0.0510)	(0.0164)	(0.118)	(0.0150)	(0.0961)
Mine operating		0.00165		-0.0217		-0.00803		-0.180***		-0.191***
		(0.0240)		(0.0259)		(0.0308)		(0.0454)		(0.0411)
Mine operating * HH close (DiD)		-0.0394		-0.0232		-0.0168		-0.150		-0.0998
		(0.0372)		(0.0464)		(0.0520)		(0.128)		(0.105)
Number of respondents	56,860	19,242	100,811	27,219	100,792	27,212	48,796	13,583	49,044	13,598
Number of groups	929	127	1,307	189	1,307	189	727	143	731	144
R-squared	0.002	0.011	0.007	0.018	0.020	0.024	0.001	0.007	0.001	0.008

Notes. Columns labeled 'Cross-section' show estimates of equation (1), using indicators for each mine-year pair as group fixed effects. Columns marked 'Panel' report estimates of equation (2), using mine-level area effects, and indicators for each country-year pair as time fixed effects. Standard errors are clustered at the mine level. Significant at * 10%, ** 5%, *** 1%.

Appendix 1.L: Distribution of mine-level treatment effects

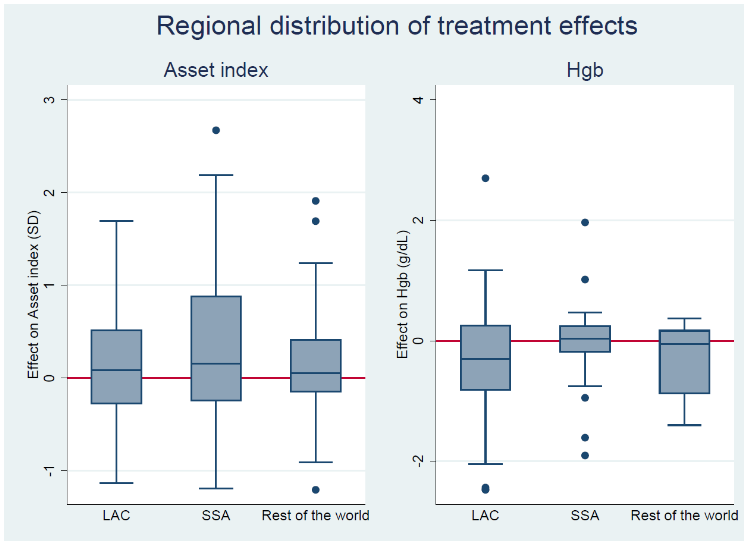


Figure 1.L.1 - Regional distribution of treatment effects

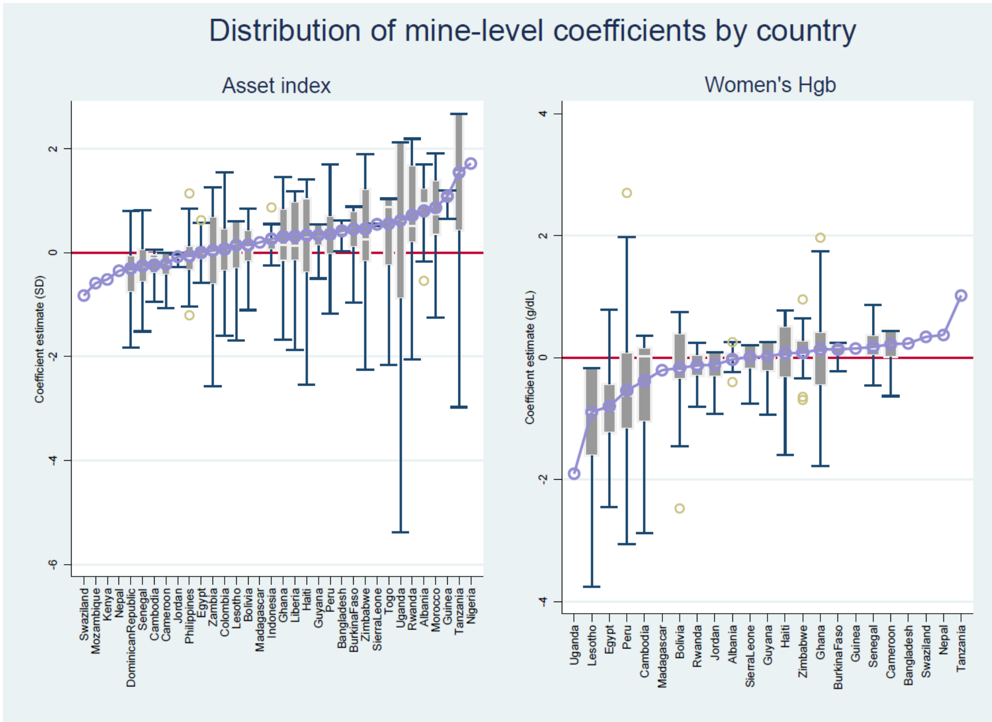


Figure 1.L.2 - Distribution of mine-level coefficients by country

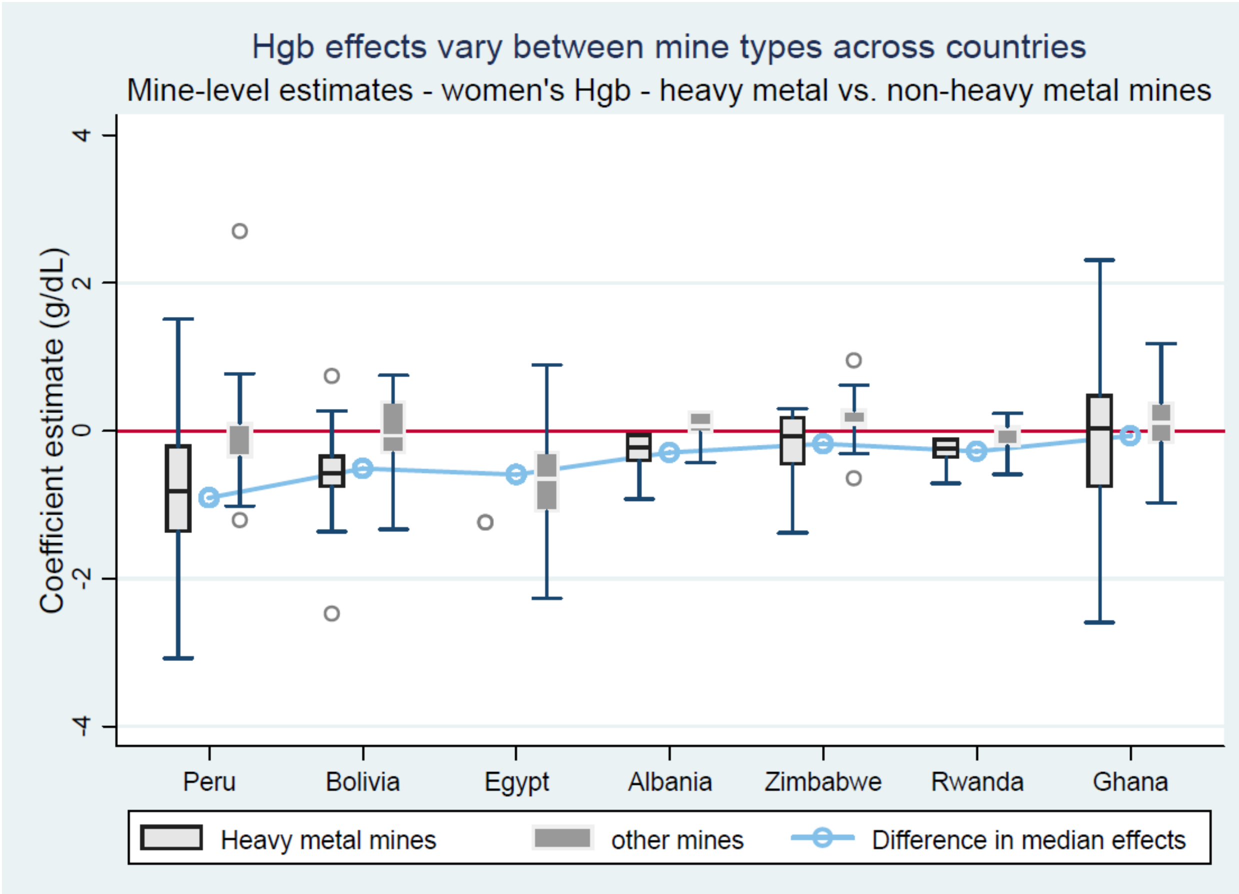


Figure 1.L.3 - Distribution of mine type difference-in-difference estimates across countries

Note: The sample shown is limited to the sub-set of countries where mine-level estimates can be obtained for at least one mine where heavy metal pollution is expected, and one mine of a different type.

Chapter 2

Cost-sharing in environmental health products: evidence from arsenic testing of drinking-water wells in Bihar, India

Cost-sharing in environmental health products: evidence from arsenic testing of drinking-water wells in Bihar, India*

Prabhat Barnwal [†] Alexander van Geen[‡] Jan von der Goltz[§]

Chander Kumar Singh[¶]

Abstract

Groundwater contaminated with arsenic of natural origin threatens the health of tens of millions of villagers across South and Southeast Asia. With a field experiment conducted in Bihar, we assess the scope for cost-shared provision of well-water arsenic tests, and study how households use the information revealed by testing. Demand is substantial, but highly sensitive to price; uptake falls from 69% to 22% of households over our price range (Rs. 10 to Rs. 50 – about equivalent to daily per capita income). Repeating the sales offer after a two-year hiatus raises overall uptake substantially, from 27% to 45%. About one-third of households with unsafe wells switch to a safer water source. Households that bought at higher prices are no more likely to switch, consistent with an absence of sunk cost or screening effects. Finally, we demonstrate that households selectively forget and remove evidence of adverse test outcomes.

*The authors especially thank the Government of Bihar, India, for its support. We gratefully acknowledge funding from the International Growth Center, under grants number IGC-13854 and IGC-14347.

[†]Department of Economics, Michigan State University

[‡]Lamont-Doherty Earth Observatory, Columbia University

[§]School of International and Public Affairs, Columbia University

[¶]Department of Regional Water Studies, TERI University

2.1 Introduction

There is pronounced policy interest in assessing how fee-based provision affects the demand for and use of basic preventive health products in low income countries. In particular, much attention has been given to the question of whether charging a fee (usually selected to share cost, rather than cover it entirely) helps or hinders sustainable access to and use of these products (Dupas, 2014a; Kremer and Miguel, 2007; Tarozzi et al., 2014). We consider these issues in the context of a preventive environmental health product, namely testing of drinking water wells for arsenic in Bihar, India.

In many low-income settings, very high social benefits are associated with basic preventive health products, such as insecticide-treated bed nets to prevent malaria infection (ITNs), or technologies to remove microbial pathogens from drinking water (Ahuja et al., 2010; Sachs and Malaney, 2002). Yet, it has proven difficult to chart a path – through private or public provision – to ensure sustainability. Since willingness to pay is limited even for very effective health interventions, market-based private provision often encounters little demand. Conversely, it has been argued that initial subsidization may present an opportunity for learning, and help future uptake. There is thus a case for public provision in low-income settings. Yet, public distribution is in turn beset with difficulties that the incentives inherent in market-based provision might help avoid – from slow and unreliable provision to poor targeting of the free good toward intended beneficiaries, and limited innovation in products and delivery. It has also been asked whether households may use products provided free of charge less reliably than they might use products for which they would have paid a fee ('screening' or 'sunk cost' effects), or whether initial subsidies may create an expectation of future free access, and lower willingness to pay ('anchoring') – though empirical evidence often does not bear out these concerns (Cohen and Dupas, 2010; Dupas, 2014b). Given the flaws of both private and public provision, cost-sharing is often suggested as a way to reduce dependency on public programs, without exposing consumers to the full cost of market provision. However, even relatively limited fees have been shown to significantly reduce take-up (Bates et al., 2012; Dupas, 2014a; Kremer and Miguel, 2007). The question of how to foster sustainability thus remains relevant.

Arsenic tests for drinking water wells share important product traits with other highly efficient preventive health interventions, and their provision poses similar questions of sustainability. Naturally elevated arsenic concentrations in well water were first reported in the mid-1980s in West-Bengal and subsequently shown to extend over a much broader area (Ahmed et al., 2006; Chakraborti et al., 2003; Fendorf et al., 2010) In areas where arsenic contamination is prevalent, tests are essential in that they provide information that is not substitutable. Because the distribution of arsenic incidence in groundwater is difficult to predict, and varies greatly even over small distances, the safety of a well cannot be predicted without a test (van Geen et al., 2002). A well that meets the WHO guidelines for arsenic in drinking water may be found in immediate neighborhood of a very unsafe well. Nor is there an easy way to design wells to be both safe and affordable: within shallow ($<$ 100 m) aquifers tapped by most private wells, there is no systematic and predictable relationship between and arsenic and well depth. At the same time, precisely because arsenic contamination varies greatly over small distances, well tests make available an effective way to avoid exposure, namely by switching to nearby safe wells. In previous interventions, about one-quarter to two-thirds of households with contaminated wells have been found to switch to safer sources (see, e.g., Ahmed et al. (2006); Chen et al. (2007); Madajewicz et al. (2007)).

Much like other basic preventive health products, arsenic tests are also very cost efficient. The cost of goods and services (COGS) for a test provided through our program was a mere USD 2.26, excluding cost purely related to data collection. (There is, of course, a potentially significant inconvenience cost to switching wells.) By stark contrast, the health consequences of chronic arsenic exposure are dramatic. Argos et al. (2010) conducted a large cohort study in an area of Bangladesh where arsenic contamination was representative of the national distribution, and estimated that 21% of all-cause deaths were due to chronic exposure by drinking water at arsenic levels above $10\mu g/l$ (the 60th percentile of the arsenic distribution in our sample). Arsenic in tubewell water has also been associated with impaired intellectual and motor function in children (Parvez et al., 2011; Wasserman et al., 2004). In consequence, there are significant effects on income and labor supply: Pitt et al. (2015) estimate that lowering the amount of retained arsenic among adult men in Bangladesh to levels encountered in uncontaminated countries would increase earnings by 9%. Matching households to arsenic exposure, Carson et al. (2011) find that overall household labor supply is 8% smaller due to arsenic exposure.

Because of their low cost and important health benefits, arsenic well tests have been provided free of charge at large scale. Large-scale testing campaigns have been carried out through public provision in rural communities across the Indo-Gangetic Plain (Ahmed et al., 2006; Fendorf et al., 2010). However, these important programs have not come close to comprehensively covering the geographic area where arsenic is of concern – including in our study area. Due to the explosive growth of tube well use, they may also need complementing where they have once been carried out. Thus, after a single blanket testing covering five million wells by the government of Bangladesh in 2000-2005, no further country-wide public programs have been undertaken as of the time of writing. In consequence, recent estimates suggest that more than half of currently used tube wells in Bangladesh have never been tested for arsenic (van Geen et al., 2014). Public provision has hence not fully met the need for testing, and a permanent network of test providers may be required to ensure coverage. This prompts the question whether cost-shared private provision might provide a sustainable complement to public provision, and whether there is the prospect of a market for arsenic tests in which local entrepreneurs would have an incentive to seek out untested wells (indeed, wells are drilled by small entrepreneurs).

In this paper, we shed light on this issue using a randomized control trial conducted in 26 villages in Bihar, India, from 2012-2015. In order to elicit demand, we offered tests at prices between Rs. 10 to Rs. 50, randomized at the village level. The highest price level was approximately equal to one day of per capita income in Bihar, or one-third of the full cost of goods and services.[1] We find that there is a considerable demand for arsenic testing: at the mean across price groups, and over the duration of our intervention, 45% of households purchase the test. However, demand drops steeply with price, in line with demand elasticities found in other studies of highly effective preventive health care products (Cohen and Dupas, 2010; Kremer and Miguel, 2007). To our knowledge, no study has previously estimated the demand curve for diagnostic testing of water source quality for arsenic. One related study by George et al. (2013) considers demand for arsenic testing at a single fixed price in Bangladesh, and shows that education and media campaigns increased adoption.

To further assess the question of sustainability, we repeat the sales offer two years after the initial campaign, at the same (nominal) sales price. We record significant additional demand at the time of the repeat offer, with overall coverage rising from 27% to 45%. Data limitations do not allow us to ascertain what mechanisms lie behind additional demand – wealth increases, learning, or the direct effect of repeating the offer (what one might call a 'marketing' or 'nudge' effect). However, from the vantage point of policy interest in sustainability, the reduced-form effect of making a repeat offer is highly relevant. The observed additional demand is perhaps particularly remarkable because the opportunities for learning are somewhat circumscribed by the fact that arsenic tests are an experience good only in a very limited sense. Thus, once some consumers buy tests, others may observe that neighboring wells test positive for arsenic, and may learn about opportunities to switch – but because the health impact of arsenic are slow in onset, health benefits are not immediate observable.

Our study further contributes to the literature by investigating how households respond to information on the arsenic status of their well. In a follow-up survey conducted three months after the first wave of test offers, about 31% of households whose wells had unsafe

[1] http://www.indiaenvironmentportal.org.in/files/file/Economic-Survey-2014-bihar.pdf.

levels of arsenic reported having switched to a safer tube well for their drinking and cooking water needs. This rate is in line with previously reported switching rates, though at the lower end of the spectrum (Ahmed et al., 2006; Bennear et al., 2013; Chen et al., 2007; George et al., 2012; Madajewicz et al., 2007; Opar et al., 2007). We find no effect of price paid on the probability of switching to safer water sources, in line with earlier studies rejecting a 'screening' or 'sunk cost' effect in the use of other preventive health products (Dupas, 2014a).

In a novel result, we find strong evidence of selective recall of test results. Thus, about half of the household whose wells tested unsafe were unable to recall their well status correctly (with no corresponding difference among owners of safe wells). Some households proactively discarded placards attached to wells to indicate that drinking water was not safe. Stigma, concerns over reduced property value, or obstacles to switching might explain this choice.

Two limitations arising from the study's implementation are worth noting. A review of the field work finds that in the first phase of test sales, enumerators did not systematically collect data from all households approached with a sales offer. To mitigate the resulting obstacles for demand estimation, we collected recall data on sales offers and purchases during the second offer phase. Secondly, an attempt to create a well owner-level panel was unsuccessful, since well tags attached during the first phase proved to be far less durable than expected, and could not be comprehensively tracked.

The remainder of the paper is structured as follows. Section 2.2 discusses the details of the experiment and data. Results are presented in Section 2.3, and Section 2.4 concludes.

2.2 Details on Experiment, Data and Methodology

2.2.1 Study setting and sample

Our study is set in a region in the Indo-Gangetic plains in Bihar, India, arsenic levels are elevated in a significant proportion of drinking water wells. Chakraborti et al. (2003) first showed that a significant proportion of wells in the region was elevated in arsenic by extending their testing campaign upstream along the Ganges from the state of West Bengal. Arsenic testing is a new service in the study area: tests are not available in the private market (nor are they elsewhere in South Asia), and while Nickson et al. (2007) report that about 5,000 wells have been previously tested in the general area, it has not previously been covered by any government-sponsored blanket testing of wells.[2] Within the general study area, we selected Bhojpur district to conduct our intervention. Within this large district (1,045 villages are recorded in the Census), we select a study area of four blocks (sub-districts) adjacent to the village where arsenic was first reported in Bihar (Chakraborti et al., 2003). We discuss external validity of our results below. Within these, we choose 26 villages of moderate size (50-400 households) for this study, based on a high probability of arsenic incidence, as indicated by distance from the river.[3] Our endline survey identifies 4,084 well-owner households in total.[4]

To elicit demand, we used a simple revealed preference approach – namely, making take-it-or-leave-it offers of arsenic tests at a certain price to households in the sample villages. As is obvious, a take-it-or-leave-it offer elicits only a bound on each household's willingness to pay. For instance, if a household accepts to purchase a test at Rs. 30, we can only infer that its willingness to pay was at least Rs. 30. Similarly, rejection only suggests that willingness to pay was less than the asking price.

We randomly assigned each village to one of five price levels at which households were

[2]Nickson et al. (2007) report arsenic testing of about 5,000 wells in six out of 14 sub-districts of our study district. The sub-districts were not identified in the study, and it is hence not possible to precisely compare the number of wells tested to the number of local wells. However, the share of wells tested was certainly a small fraction of the 335,000 wells reported in the 2011 Census for the entire study district. 26% of wells tested unsafe.

[3]The original intention was to work in a sample of 25 villages, i.e., five villages in each of our five price groups. However, enumerators erroneously visited two villages of the same name during initial field work. We included the additional village as the 26th for the rest of the program.

[4]We cross-checked the number of households recorded in our study against 2011 Census data for 21 out of 26 villages that could be matched to the census. For these villages, the census shows 4,497 households that own a hand pump, whereas we record 3,322 attempted sales in the same 21 villages - that is, 74% of the census population . The discrepancy is in significant part due to the failure to include entire parts of a few villages, because enumerators believed these to be distinct villages.

offered arsenic tests for purchase, rising from Rs. 10 to Rs. 50, in increments of ten. It was felt that offering different prices to households *within* a given village would be seen as violating fairness norms, and would deter purchases.[5] We therefore chose not to randomize our prices within villages. The highest price was chosen based on initial local focus group discussions; it is slightly higher than the average daily per capita income of Rs. 45 in Bihar in 2011-12. Revenue from test sales was used to partially cover the enumerators' salaries and travel cost. The cost of the test kits alone was about USD 0.35 (about Rs. 21 at January 2014 exchange rates); the COGS for testing, including wages, quality control, and test result placards amounted to USD 2.26 (Rs. 136). (Metal well tags intended purely for data collection added an additional USD 0.48 (Rs. 29).) The highest price charged therefore more than covered the cost of the test kits, and about one-third of the entire COGS. We did not add a treatment arm that would have offered tests free of charge, because of a strong expectation that take-up would be near-universal at zero cost. This expectation was based on prior experience in arsenic testing campaigns, and was confirmed further when free tests were offered with near-complete take-up in four pilot villages visited for the design of our experiment. It is also in line with broader evidence from the lab (Shampanier et al., 2007) and from field experiments (Cohen and Dupas, 2010; Kremer and Miguel, 2007).

2.2.2 Implementation – testing campaign and surveys

Testers were locally recruited from among college graduates, and trained prior to the roll-out of the campaign. (By way of contrast to van Geen et al. (2014), where village health workers conducted tests.) Testing then proceeded in two waves; the first conducted in 2012-13, and the second, in 2014-15. (See Table 2.1 – henceforth, for simplicity, we refer to the first round of testing as having taken place in 2012, and the second round, in 2014.)

[5]This consideration obviated the use of alternative techniques for eliciting willingness to pay, such as the Becker-DeGroot-Marschak (BDM) mechanism and other auction-based methods. In any case, auctions would have been unlikely to be efficient mechanisms, given the potential buyers' uncertain and likely correlated beliefs over the value of arsenic tests.

2.2.2.1 First wave of testing – initial sales offer

The first wave of testing began with focus group meetings in each village. To increase awareness of the arsenic issue, a large poster was put on display, showing a satellite image of a pilot village along with color markers indicating the arsenic status of tested wells (Figure 2.2). The poster served the additional purpose of making tangible the great spatial variation in arsenic contamination, and the resulting opportunities for well switching. Following the focus group meetings, testers began to offer tests door-to-door; where a sale was made, tests were conducted using a reliable field kit that requires approximately 15 minutes per test (van Geen et al., 2014). The protocol foresaw that for all households approached with a test offer, GPS locations and basic data on the household would be collected. However, in contrast with what was intended, testers did not record data from *all* households that did not purchase a test. We discuss the resulting challenges for demand estimation, and our solution approach, at length in Appendix 2.A. During the initial wave of test offers, a total of 1,212 tests were sold across the 26 sample villages (Table 2.A.1, Column 3). The results of each test were posted on the pump-head of the well that was tested, with an easy-to-read metal placard, color coded red for unsafe wells ($> 50\mu$g/l arsenic), green for 'borderline safe' wells where arsenic is of some concern (> 10-50μg/l), and blue for safe wells ($\leq 10\mu$g/l) (Figure 2.3). The cut-off values were chosen to correspond with the Indian national safety standard for arsenic of 50μg/l that was current as of the time of the test campaign, and the WHO guideline of 10μg/l (the government of India – unlike the government of Bangladesh – has since matched its standard to the WHO guideline). Smaller placards with a unique well ID were also attached to each pump-head in anticipation of a future response survey. Regrettably, well ID tags proved to be less durable than hoped, and less than 5% of tags placed in 2012 were still attached in 2014. It was hence not possible to reliably link wells across survey rounds. Immediately after the first wave of arsenic testing was completed, village-level maps were exhibited in each village, showing the geo-locations of safe, borderline safe and unsafe wells, with the goal of illustrating, where relevant, that the proximity of safe wells would make well-switching feasible. (Geo-locations were jittered to preserve anonymity.) During home visits, households were alerted to the fact that switching from unsafe or borderline safe wells to neighboring safe wells would be an effective way to avoid arsenic exposure. The first phase of the project concluded with a follow-up survey conducted approximately three months after testing was completed. Enumerators visited all households with a high-arsenic well-head and collected information on whether they now drew water from neighboring safe wells.

2.2.2.2 Second wave of testing – sales offer repeated

In a second phase, commencing in 2014 – some two years after the initial visits – we offered the tests again in the same set of villages, and at the same price assigned initially. Across the 26 villages, a total of 4,084 households were approached with the intention of making a sales offer (Table 2.4, Column 4). In the second round, data was collected systematically from every household where a respondent could be interviewed, including from households that did not wish to buy the tests. Each house was visited at least two times to ensure high coverage. After two visits, about 14% of households could not be surveyed because no adult member was present or willing to answer questions; sales offers could be completed in 3,528 households. The enumerators reported that, to avoid embarrassment, some households who were unwilling to purchase tests at the asking price avoided being interviewed. For a conservative demand estimate, we therefore work throughout with the number of households approached for sales, rather than the number of households where a sales offer could be completed. A total of 719 tests were sold in this second phase (Column 5). The household survey administered in the second round gathered socio-economic and demographic information, along with GPS locations of the wells. It also collected information on recall of tests being offered and purchased in 2012, along with recall of test results. This recall data allows us to work around some of the constraints posed by the implementation issues encountered during the first wave of offers.

2.2.3 Summary statistics

Summary statistics from the 2014 survey show modestly well-off village communities (Table 2.2). Households are of moderate size (3.9 members on average). Most (89%) own at least one mobile phone, and most (70%) live in houses made from durable building materials ('pucca'). Ownership of bikes (68%) and cows (67%) is common, though fewer households own consumer durables or have access to sanitation, and very few own cars.

Table 2.2 also shows a randomization check on observables. Here, and throughout the paper, we analyze data using ordinary least squares regression, with straightforward specifications. In all regressions, we report cluster bootstrapped standard errors to account for randomization at the village level. As Table 2.2 shows, price category dummies are jointly significant at the 90% level for two out of the eleven variables tested. The two instances where there are significant differences (ownership of cars and access to sanitation) appear isolated, and would suggest opposite signs in a relationship between price and ownership. There is therefore no indication that the price groups in question are systematically any more or less wealthy than the other groups.

To give a sense of the external validity of our results, Table 2.3 compares household wealth proxies in the 2011 Census for our sample villages, the four blocks that nest them, Bhojpur district, and the state of Bihar. As is evident, households in our sample villages are similarly well-off as the mean household in the blocks (Panel A) and Bhojpur district (Panel B). They are, however, better off than the average household in Bihar, with a far higher share of houses made from durable materials, greater literacy, and ownership of household assets up to 10pp higher for many categories (Panel C). While we show below (in Section 2.3.1.2) that asset ownership does not strongly predict willingness to pay, we might expect demand in our sample villages to be representative of Bhojpur district, but at weakly higher than in Bihar at large.

2.3 Results

2.3.1 Demand for well arsenic testing

Demand for fee-based arsenic tests in the study area is substantial. Overall, after excluding 74 repeat purchases by households who had their wells tested in both 2012 and 2014, a total of 1,857 tests were sold at randomly assigned prices across the 26 sample villages over the entire duration of the program (2012-2015). This implies that arsenic testing covered about 45% of households approached for sales (Table 2.4, Column 10).[6] An example of test results in one village is provided in Figure 2.1; a map displaying the proportion of safe, unsafe, and untested wells in each village is shown in Figure 2.4. It pools results from the first and second test phase. In total, using the national and WHO thresholds of 50 and 10μg/l, respectively, 50% of wells tested 'safe' ('blue'), 31% tested 'borderline safe', and 19% tested 'unsafe' ('red'). As expected, test results varied over small distances, and there is a wide spread in the shares of unsafe wells across villages, ranging from 2% to 77%.

Demand in the first round of sales alone was 27% across price groups in our preferred recall estimate (Column 8). Demand at the time of the second offer was 18%, after adjusting for repeat purchases (Column 9). As noted, demand estimation for the first round of sales is complicated by incomplete data collection. In Appendix 2.A, we discuss our solution approach, and assess robustness. In the following, we work with recall data systematically collected during the second test wave to determine 2012 demand, both because it is more internally consistent, and because it yields more conservative estimates (overall demand was 30% using an alternative approach of imputing demand from 2012 sales and the 2014 sample size).

[6]To estimate total coverage after two offers, we add first and second-round coverage, correcting for repeat purchases. We define second-round purchases to have been repeat purchases in 74 instances where households recall having bought the test in 2012, and purchased another test in 2014. Households had been advised that, since arsenic levels in ground water are stable over time, wells need not be tested repeatedly.

2.3.1.1 Price sensitive of demand

In line with prior research, we find that demand is highly sensitive to price (Figure 2.5, Table 2.5). The mean elasticity across sales at different price levels in our data is -0.36 in the first round, and -0.47 in the second round. At the lowest price of Rs. 10 (USD 0.15 at market rates at the time of the repeat offer), 40% of households purchase the test after one offer, and 69% after two offers (Table 2.4, Columns 7 and 10). While our experiment did not include an arm with zero price offer, uptake of free tests can be assumed to be nearly 100% (as discussed in Section 2.2.1). Thus, while there is significant demand at Rs. 10, charging this small amount, rather than offering the test for free, reduces coverage after two sales offers by about one-third. Demand further drops precipitously at higher prices, and at Rs. 50, reduces to less than one-sixth of households after one offer, and less than one-quarter after two offers.

This pronounced sensitivity is in line with demand behavior observed in other recent studies of preventive health products such as ITNs or rubber shoes in developing countries (Cohen and Dupas, 2010; Dupas, 2014b; Kremer and Miguel, 2007; Meredith et al., 2013). The fact that arsenic tests arguably were less well-known to consumers than products studied elsewhere was not reflected in distinctly higher price elasticity.

Perhaps the most natural comparison in terms of the nature of products offered is to Berry et al. (2012), who study willingness to pay for water filters to remove pathogens in northern Ghana. Berry et al. report that, while 95% of respondents had non-zero willingness to pay (an analogue of near-universal take-up at zero cost), charging a price equivalent to 116% of daily income (or 30% of the filter's cost) reduced demand to 21%.[7] This is comparable to outcomes in our experiment at a price of Rs. 50 and after one sales offer: demand of 15% at a price equivalent to 111% of average daily income, and 30% of the full cost of goods and services.

[7]Demand figures from Dupas (2014a). Figures are not directly reported in Berry et al. (2012). Share of income based on USD 4.20 (GHS 3) price and 2010 (current) per capita GDP of USD 1,323.

Our demand estimates compare well with results shown by George et al. (2013), who estimate demand for arsenic tests in Bangladesh at a single price point of USD 0.28 in 2011 – the equivalent of about Rs. 10 in 2014 in our setting. George et al. find 53% uptake in the control group, where no dedicated awareness campaign is conducted, and 93% uptake in each of two treatment arms with an awareness campaign. Our demand estimate at Rs. 10 is in between these two values after two offers, but far below after a single offer. This is perhaps intuitive: arsenic test were not widely known in our intervention area, while George et al. (2013) worked in Bangladesh, where government-sponsored blanket testing and many other interventions have significantly raised awareness of arsenic.

2.3.1.2 No buyer selection at different price levels

We next test how sales price correlates with buyer characteristics in terms of household wealth proxies. Table 2.6 shows regression results for second-wave buyers (results for 2012 buyers are similar, and omitted for conciseness). As is evident, few asset categories are correlated with sales price. For those that do correlate, selection was limited to the two highest price levels. Given the large drop in demand associated with a price increase from Rs. 10 to Rs. 20 (13pp, or 45% in relative terms), it is perhaps surprising that there is virtually no distinction in observed asset ownership between households that buy at these price levels. The absence of a wealth pattern suggests that, either, purchasing decisions were driven by different valuation of the product among similar households, or marginal utility of consumption differed in ways that do not correlate with characteristics we observe. Investment in sanitation – i.e. having a latrine facility in the house – is correlated with purchase decisions at high price levels (about one household in three among those who buy at Rs. 10 owns a latrine, but two in three do among those who buy at Rs. 50). This result might well speak to a concern over hygiene and health driving both investments.

2.3.1.3 No residential sorting

We test whether households can predict arsenic contamination, and potentially, sort accordingly in choosing their residence. As noted, the distribution of arsenic in groundwater wells is hard to predict; it would be surprising if we were to observe sorting. Appendix Table 2.C.1 confirms this notion, in keeping with findings in Madajewicz et al. (2007). There is no relationship between well characteristics (age, depth, and price) and the probability of high contamination – that is, households do not appear to specify well design to effectively avoid arsenic (Column 1). Nor is there a distinct relationship between asset ownership and arsenic status of wells that would suggest residential sorting (Column 2).[8]

2.3.1.4 Additional demand when offer is repeated after two years

A key feature of our experiment is that in each village, the initial test offer was followed by a repeat offer after some two years had elapsed – at the same (nominal) sales price. Our purpose in re-offering the arsenic test was to assess whether additional demand (i.e. from households who did not purchase in the first phase) could be elicited after a two-year delay. We repeated the offer *at the same price charged initially*, as opposed to repeating it at a *uniform* price as in Dupas (2014b). This allows us to study the (reduced-form) effect of making a repeat offer at different price levels, a question of immediate policy interest. However, we sacrifice the ability to directly test for learning as a specific mechanism driving demand at the time of the second offer.[9]

We find that repeating the offer after a two-year delay did indeed generate substantial

[8]Given the small number of high-arsenic wells, tests are run separately for each asset category to avoid over-fitting. Due to multiple hypothesis testing, the standard errors reported in Appendix Table 2.C.1 are too small. We omit any adjustment because the absence of sorting emerges even when precision is overstated.

[9]Appendix 2.B shows an alternative test for learning. The reason why we cannot assess learning as in Dupas (2014b) is as follows. Our product is distinct from the ITNs offered in Dupas (2014b) in that there is no reason for households to repeat arsenic tests, whereas there is reason to purchase ITNs again after some time. Still, if we had made the second sales offer at a uniform price, we might have tested for learning by using first-round price to instrument for first-round demand, and then study the effect of first-round demand on second-round demand through peer learning. This is not possible, however, when price levels are the same in the first and second round: as an instrument, price would clearly violate the exclusion restriction.

additional demand. Thus, purchases at the time of the second offer raise total coverage by some 18 percentage points (pp), from 27% to 45% (Table 2.4, Columns 7 and 10). Demand is more price-sensitive than at the first offer (Figure 2.5). However, we observe an effect of repeating the sales offer on coverage at any price level, with increases ranging from 70% of the original sales at Rs. 10 to 19% at Rs. 40.

From a policy perspective, the effect of making a repeat offer is remarkable: price matters greatly for demand, but at any price level considered here, repeating the offer meaningfully increases coverage (and from a business perspective, sales). Irrespective of the channels – learning, income growth, or marketing intensity, this simple finding underscores the need for a more careful assessment of experimental evidence generated with offers available only for a short period.

Because we lack a household panel, and because there may be some error in recall of first-round tests, we cannot completely rule out the concern that some of the demand at the second offer may be driven by households that may not have been approached during the first offer phase in 2012. However, the observable evidence offers significant reassurance. About 70% of the new purchases in 2014 are made by households who recall being offered the test in 2012, but did not purchase (Table 2.4, Columns 5-6). Perhaps most compellingly, the pattern of 2014 demand is very similar among those who recall having been made an earlier offer and the overall sample (Column 10).

It is intriguing to ask why there is a high level of demand when a repeat offer is made within the relatively short time frame of two years. However, our data does not allow us to conclusively assess this question; Appendix 2.B shows some limited evidence. (i) Strong state-level growth in nominal income between survey rounds suggests that changes in wealth between the first and second offer may have played a role; our survey data on asset ownership is consistent with this mechanism, but not conclusive. The absence of a correlation between wealth and price among buyers is at odds with this explanation (see Section 2.3.1.1). (ii) Learning may have lead households to adjust their valuation of arsenic testing. The product's characteristics were not familiar to potential customers at the time of the first offer, and the initial wave of tests may have allowed households to change their beliefs about the possibility of contamination, and opportunities to switch, although the health benefits of switching cannot be observed within two years. We obtain the 'expected' sign in a test with a credibly causal interpretation, but the results are not significant. (iii) In the absence of conclusive evidence on wealth or learning effects, one could speculate about a direct effect of repeating the offer – what one might call a 'marketing' or 'nudge' effect. We consider it a priority for further work to assess the importance of such an effect.

2.3.2 Household response to test results

2.3.2.1 Well switching

We next consider how households use the information revealed by arsenic testing. Particular importance attaches to whether households switch from highly contaminated wells to safe water sources. Within the context of the wider literature on preventive health products, this can be viewed as equivalent to behavioral issues surrounding the use of products once they have been purchased. Thus, it is the act of switching to a safe water source that brings about health benefits after the purchase of a test – and switching imposes further inconvenience cost. Similarly, after the purchase of an ITN or a drinking water filter, it is the act of sleeping under the net or filtering water that generates health benefits, and each may be associated with inconvenience.

Among households that purchased the test in 2012, 31% reported that they had switched to safer water sources at the time of the follow-up survey. This is a low switching rate, but not an atypical response. A number of similar studies in Bangladesh have reported switching rates of 26-39% (Ahmed et al., 2006; Bennear et al., 2013; Chen et al., 2007), although others find higher rates, in between one-half and two-thirds of affected households (George et al., 2012; Madajewicz et al., 2007; Opar et al., 2007). In line with prior evidence (Chen et al., 2007; Opar et al., 2007), we find that distance to safer wells is an important predictor of switching (Figure 2.6).

The somewhat subdued response to information could be related to the limited number of wells identified to be safe, because of lower take-up of the for-fee service, as opposed to blanket testing. It could also plausibly be due to restrictions on sharing water based on caste affiliation and religion. – Among households in our survey, 90% report that they prefer to exchange water within their own caste or group of relatives. Similarly, in Uttar Pradesh, a state adjacent to Bihar, caste in particular has been found to be a major factor in impending water trade within a village (Anderson, 2011).

We further find that the propensity to switch does not depend on the purchase price (Table 2.7).[10] We interpret this result to demonstrate an absence of screening or sunk cost effects. Both effects would tend to increase usage with price, and imply that highly subsidized provision might lead to 'overinclusion' of those who do not sufficiently value the product.[11] Our result further bolsters recent findings that have suggested that, for preventive health care products, there is little empirical evidence of overinclusion in subsidized provision (Cohen and Dupas (2010); Dupas (2014a) – see Berry et al. (2012) and Ashraf et al. (2007) for experimental evidence of screening, but not sunk cost effects).

2.3.2.2 Selective recall of test results and loss of test placards

We find strong evidence of selective recall. During follow-up, households not only avoid reporting adverse arsenic test outcomes, but take direct action to remove markers of unwelcome results.

Table 2.8 offers a test for selective recall. It compares the proportion of tests in each cat-

[10] To guard against concerns that the tests for individual price categories shown in Table 2.7 might be under-powered, we confirm that there are no significant differences when we regress on a dummy variable for 'high' price level, under any possible cutoff (results available upon request).

[11] In our setting, the respective arguments are as follows: 'those who decided to buy at high price care more about health from the outset, and will therefore be more likely to switch wells'; and 'those who buy at high prices have invested more in the test, and will hence more highly value the information it yields'.

egory of arsenic contamination levels (Red/high, Green/moderate, and Blue/safe) observed in first-round test outcomes recorded in 2012 to the proportion of tests in the same category of outcomes *recalled* in 2014. We adduce three different measures of recalled arsenic status – namely, (1) those households where the test placard was still affixed to the well; (2) those where the placard had been removed from the well, but was still kept in the house; and (3) those where the placard was neither on the well nor kept in house, but the respondent reported being able to remember the arsenic contamination status. As is evident, the proportion of wells respondents believe to be unsafe is consistently some nine to eleven percentage points lower than the true proportion of red tests recorded in 2012 (Columns 1, 4, 7, and 10). It is particularly striking that such a discrepancy exists even among households where the test placard was still attached to the well: since it is inconceivable that red tags are more likely to be accidentally lost than others, this is clear evidence of intent either to hide the well's status, or to avoid being reminded of it (Column 1). The magnitude of the effect is very substantial: 20% of wells tested 'red' in 2012 – and hence, a decrease of the share of 'red' wells by about 9-11pp implies that about half of the households with wells that were high in arsenic intentionally sought to hide the test outcome. We also note that respondents who did not produce a placard tended to preferentially indicate that wells were tested 'green' – suggesting that households prefer to claim a medium arsenic level in their highly contaminated wells (Column 8). Conversely, as Appendix Table 2.D.1 shows, wells in households that opted to repeat the arsenic test in 2014 were more likely to have tested 'green' than those only tested once. This suggests that households who initially received 'mixed news' were more likely to hope for a different outcome than those who received clearly 'good' or 'bad' news.

These findings are consistent with general theoretical and experimental evidence of 'self-serving bias' and 'over-confidence' (see, e.g., Eil and Rao (2011)). More practically, we note that efforts to hide unsafe well status could be related to low well switching rates in various ways. It could be that well owners hide bad news because it is (for unrelated reasons) impossible to take action to remedy the situation, as evidenced by the relatively low switching rates reported above. It is also possible that both the reluctance to share and the propensity to hide bad news speak to a social stigma or material loss (e.g., in house value – for the United States, Boyle et al. (2010) find a temporary 1% reduction in residential sales values associated with a 10μg/l increment in arsenic levels) being attached to owning an unsafe well. We note that there is some indication that wealthier households may be more likely to hide adverse test results, potentially because of greater concerns over stigma or material loss. To show this, we compare test results and recall as above – but distinguish between households that owned and did not own consumer durables (the one asset ownership indicator collected consistently in both survey rounds). (Table 2.9) As is evident, while all households under-report, households that do own durables are about twice as likely to do so; the difference is significant for the larger samples.

2.4 Summary and Policy Discussion

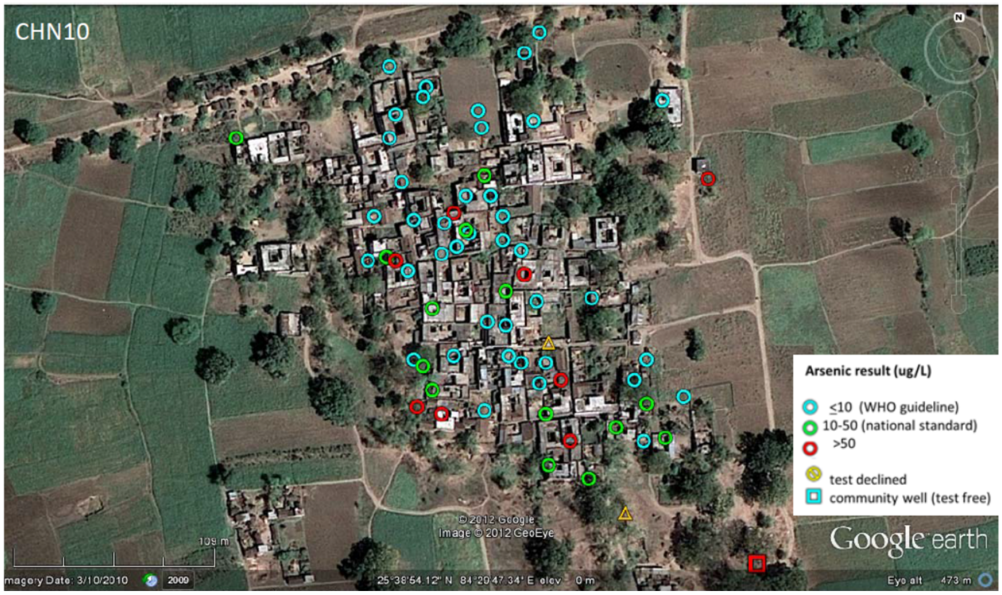
We have shown experimental evidence from Bihar, India, on the demand for and use of a preventive environmental health product – a water quality diagnostic test for arsenic contamination – when offered at a fee. Demand is substantial, but highly sensitive to price. Compared to the near-universal adoption found under free provision, two-thirds of households purchased tests at the lowest price, and about one-third at the highest price over the duration of the project. A key finding of our study is that a repeat offer made within two years of the original offer is met with significant demand, raising total coverage by 18pp, from 27% to 45%.

Our results affirm that in preventive health products, subsidies remain critical in ensuring high coverage. However, cost-shared provision might still have a useful role to play in providing an ongoing testing service in the absence of or in between public testing campaigns. In particular, one could imagine a business model in which independent testers generate their own wages, while NGOs conduct awareness campaigns, provide test kits, train testers, and implement quality control (for instance, GIS tracking and re-testing of a subsample of wells). Yet, market demand was not quite sufficient to cover wages. In 2012, expected daily revenue was about Rs. 200 (revenue per offer made was highest in the Rs. 30-50 price range, at about Rs. 8; on average, testers visited about 25 households per day). By way of contrast, under local labor market conditions, testers might have expected a daily wage in the range of Rs. 300-400.

Through a follow-up survey conducted after the first wave of sales, we assess how households respond to information furnished through well testing. About one-third of households with unsafe wells switch to less perilous water sources. This is in the lower range of switching rates found in other studies of arsenic testing. Preferences for sharing within caste groups may have limited opportunities to draw water from safer sources – an important consideration for future arsenic testing campaigns in Bihar. The probability of switching did not depend on the price paid for the test; further evidence against the empirical importance of a possible 'screening' or 'sunk cost' effect in preventive health products

By comparing the share of wells with safe and unsafe arsenic levels between test results collected in 2012 and results recalled in 2014, we show that households avoid reporting adverse test results, and indeed, remove well tags indicating arsenic contamination. This may speak to discomfort with knowledge of well status in the context of low switching rates, stigma, or concerns over property value. The reaction is certainly policy relevant – in particular when allowing for the possibility that the *ex ante* decision to purchase a test might be affected by any motivation to avoid bad news. In practical terms, the finding suggests that in future testing campaigns, it may not be worth incurring the high cost associated with durable metal placards to make test results visible.

Figure 2.1: Example of well arsenic distribution in a village in Bhojpur district, Bihar (India)



Note: a sample village map from the study is shown with the outcomes of arsenic testing. Red circles denote drinking water wells that are highly contaminated with arsenic; green circles show wells with intermediate arsenic levels; blue circles show wells that are low in arsenic and safe to drink from.

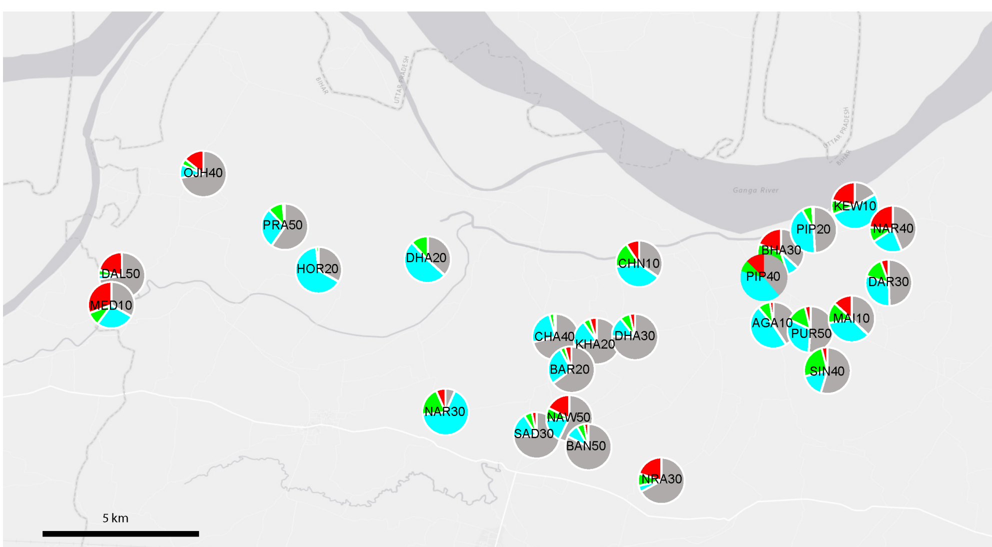
Figure 2.2: Satellite maps from nearby villages were shown in focus group meetings

Note: village meetings and exhibition of posters showing safe and unsafe wells from near by villages. The geo-location of wells were jittered because of privacy concerns.

Figure 2.3: Metal Placard showing arsenic status after testing

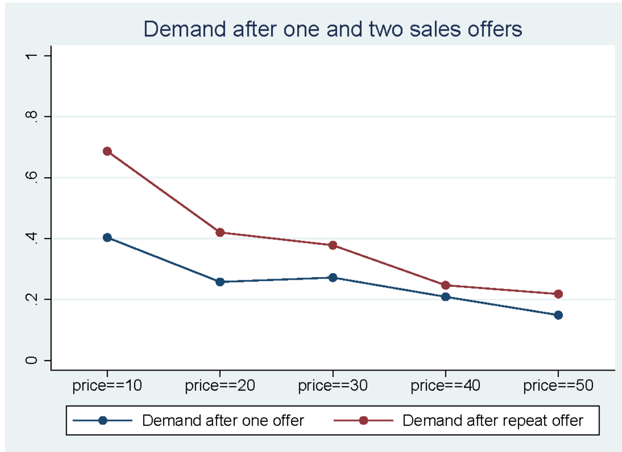
Note: red (Arsenic high), green (Arsenic moderate) and blue (Arsenic low) placards were fixed on the tubewells after arsenic testing.

Figure 2.4: Map showing village locations with the arsenic test outcomes



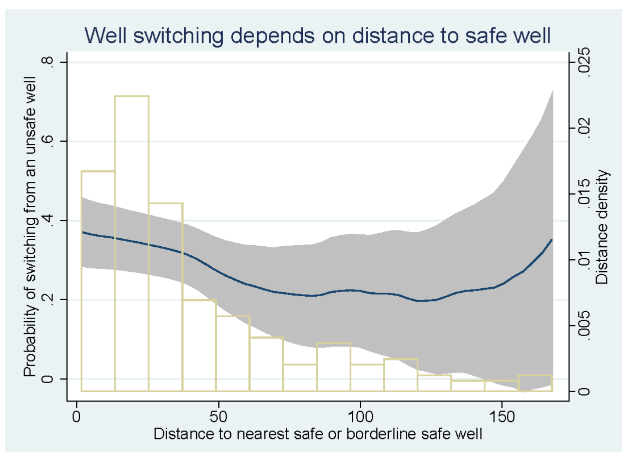
Note: the map shows the location of villages, take-up and outcome of the arsenic testing in subject area. Red (Arsenic high), Green (Arsenic moderate) and Blue (Arsenic safe) colors show the outcome of arsenic testing. Grey color shows the proportion of untested wells.

Figure 2.5: Demand curves after one and two sales offers



Note: the plot shows demand patterns after one offer (2012) and after two offers. 2012 demand estimates are obtained from recall of sales offers and purchases as measured in the 2014 survey. See Appendix A for discussion.

Figure 2.6: Switching conditional on distance to blue/green



Note: the graph shows the probability that household whose wells tested 'red' (high arsenic) in 2012 switched to a safer ('blue' or 'green') well, conditional on distance (in metres) to the nearest safer well. Local polynomial fit with confidence interval; histogram of distances overlaid.

Table 2.1: Fieldwork timeline

August 2012	Arsenic testing in pilot villages
November 2012 - February 2013	First round of arsenic testing
February 2013 - May 2013	Follow-up survey of well switching
November 2014 - January 2015	Second round of arsenic testing

Table 2.2: Summary statistics and randomization balance

	Households Members			Housing characteristics		Asset ownership					
	Total Adults (1)	Infants (2)	Children (3)	'Pucca' (4)	Has latrine (5)	Car (6)	Cell (7)	TV (8)	Bike (9)	Motorbike (10)	Cow (11)
Mean across price groups	3.893	0.201	0.370	0.700	0.326	0.0286	0.885	0.204	0.676	0.213	0.665
Mean at Price = Rs. 10	3.741	0.169	0.354	0.795	0.278	0.0384	0.855	0.198	0.722	0.214	0.638
Regression coefficients											
Price = Rs. 20	0.678	0.0830	0.238	-0.227	-0.0667	-0.0110	0.0124	0.0308	-0.0277	-0.0515	-0.00546
	(0.672)	(0.104)	(0.250)	(0.154)	(0.0855)	(0.0196)	(0.0735)	(0.105)	(0.0518)	(0.0578)	(0.0944)
Price = Rs. 30	-0.729	0.0618	-0.134	-0.0372	0.0257	-0.0127	0.0532	0.0214	-0.0469	0.00206	0.125
	(0.584)	(0.148)	(0.228)	(0.102)	(0.102)	(0.0166)	(0.0634)	(0.114)	(0.107)	(0.0367)	(0.0852)
Price = Rs. 40	0.268	0.141	0.0633	-0.142	0.166	-0.0276*	0.0623	-0.00814	-0.137	0.0297	0.00104
	(0.749)	(0.126)	(0.242)	(0.100)	(0.112)	(0.0146)	(0.0565)	(0.128)	(0.128)	(0.0321)	(0.0795)
Price = Rs. 50	0.439	0.176	0.157	-0.0304	0.270**	0.000127	0.0576	-0.0392	-0.0583	0.0644	0.0387
	(0.998)	(0.167)	(0.312)	(0.108)	(0.125)	(0.0223)	(0.0699)	(0.0766)	(0.0793)	(0.0574)	(0.0875)
N	3,526	3,528	3,522	3,758	3,528	3,527	3,528	3,528	3,528	3,528	3,527
R-squared	0.040	0.004	0.019	0.040	0.059	0.003	0.007	0.003	0.009	0.009	0.011
Joint significance											
Wald chi2(df)	4.848	2.295	2.464	3.558	15.08	9.317	1.752	0.811	1.685	3.793	4.509
Prob > chi2	0.303	0.682	0.651	0.469	0.00455	0.0536	0.781	0.937	0.793	0.435	0.342

Note: the table shows overall mean values of key demographic and asset variables observed in 2015, alongside regression results for differences in means across price groups. A test for joint significance of the price dummies is reported in the bottom rows. Cluster bootstrap standard errors in parentheses (400 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.3: External validity: characteristics of sample villages compared to census block, district, and state means

	Housing characteristics		Asset ownership					Household characteristics		
	'Pucca' (1)	Latrine (2)	Cell phone (3)	Bike (4)	Motorbike (5)	Car (6)	TV (7)	Scheduled caste (8)	Literate (9)	Employed (10)
Sample villages	0.659	0.203	0.617	0.592	0.119	0.0174	0.224	0.168	0.611	0.328
Panel A										
Census blocks where villages are situated	0.598	0.258	0.594	0.525	0.113	0.0182	0.224	0.154	0.589	0.298
Difference	-0.061	0.0547*	-0.0226	-0.0671*	-0.00658	0.00082	-0.000193	-0.0138	-0.0218	-0.0304
Panel B										
Bhojpur district	0.627	0.224	0.598	0.509	0.101	0.0184	0.182	0.162	0.583	0.31
Difference	-0.0315	0.0205	-0.0187	-0.0831**	-0.0184	0.000976	-0.0426	-0.00583	-0.0276	-0.0188
Panel C										
Bihar	0.461	0.19	0.517	0.496	0.0773	0.0161	0.128	0.179	0.505	0.343
Difference	-0.197***	-0.0132	-0.0993*	-0.0957**	-0.0421***	-0.00136	-0.0962***	0.0116	-0.106***	0.0141

Note: the table shows mean values of key demographic and asset variables observed in the 2011 Census, for 21 out of 26 sample villages that could be matched with the census, the four census blocks that nest these villages (Panel A), and the district (Panel B) and state (Panel C) where they are all located. Mean values are shown for each group, alongside the difference between the mean for the respective group and the mean for our sample villages. Significance of differences obtained from robust standard errors (omitted for readability); *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.4: Test offers, sales, and demand

	2012 offers and sales		2014 offers and sales			Demand estimates			
Price (Rs.)	Recalled offers	Recalled sales	Sales offers	Sales	Sales among HHs recalling 2012 offer	2012 demand (recall)	2014 demand	2014 demand given 2012 offer	Coverage after two offers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
10	615	249	960	288	187	0.40	0.30	0.30	0.69
20	804	206	1,105	183	135	0.26	0.17	0.17	0.42
30	460	125	815	117	74	0.27	0.14	0.16	0.38
40	441	92	653	86	72	0.21	0.13	0.16	0.25
50	350	52	551	45	34	0.15	0.08	0.10	0.22
All	2,670	724	4,084	719	502	0.27	0.18	0.19	0.45

Note: the table summarizes the number of offers and sales in both phases of the experiment, alongside the resulting demand levels. Sales reported in Column (5) include repeat purchases, while coverage after two offers in Column (10) has been adjusted by excluding 74 repeat purchases. See Appendix A for additional results and discussion.

Table 2.5: Estimated demand

	First-round demand (recall) (1)	Second-round demand (2)
Price = Rs. 20	-0.146	-0.134*
	(0.184)	(0.0723)
Price = Rs. 30	-0.132	-0.156*
	(0.163)	(0.0867)
Price = Rs. 40	-0.195	-0.168*
	(0.164)	(0.0903)
Price = Rs. 50	-0.255	-0.218***
	(0.167)	(0.0725)
Mean at Rs. = 10 (constant)	0.403***	0.300***
	(0.151)	(0.0702)
Observations	2,666	4,084
R-squared	0.034	0.037

Note: the table shows estimated demand for each individual round of test offers. Demand for 2012 is estimated based on recall data collected in 2014. See Appendix A for an alternative estimate. Cluster bootstrap standard errors (based on 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.6: Do purchase decisions at high price levels correlate with wealth?

	House Type		Other Assets					
	'Pucca'	Has latrine	Car	Cell	TV	Bike	Motorbike	Cow
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Linear specification								
Price	-0.00162	0.00747**	0.000154	0.00156	0.00180	-0.000673	0.00296***	9.18e-05
	(0.00294)	(0.00302)	(0.000374)	(0.00164)	(0.00324)	(0.00197)	(0.00100)	(0.00237)
Panel B: Breakdown by price levels								
Price = Rs. 20	-0.189	-0.0350	-0.00346	-0.0304	0.0459	-0.0366	0.0297	-0.0546
	(0.136)	(0.114)	(0.0164)	(0.120)	(0.135)	(0.0705)	(0.0742)	(0.0873)
Price = Rs. 30	-0.0367	0.0171	0.00884	0.0121	0.0394	0.0425	0.0279	0.0882
	(0.119)	(0.136)	(0.0184)	(0.0757)	(0.143)	(0.0778)	(0.0422)	(0.0805)
Price = Rs. 40	-0.173	0.254**	-0.0121	0.107***	0.0837	-0.0805	0.115***	-0.0501
	(0.118)	(0.116)	(0.0135)	(0.0407)	(0.183)	(0.146)	(0.0428)	(0.102)
Price = Rs. 50	0.0112	0.334**	0.0168	0.00559	0.0489	-0.0168	0.116***	-0.0221
	(0.0922)	(0.135)	(0.0235)	(0.0733)	(0.150)	(0.0824)	(0.0417)	(0.107)
Mean at Price= Rs. 10	0.803	0.330	0.0267	0.886	0.223	0.789	0.221	0.685
N	1,301	1,366	1,365	1,366	1,366	1,366	1,366	1,365

Note: the table shows correlations between purchase price and wealth proxies among households that bought a test during the second round of offers in 2014. Panel A shows results from a linear regression in price; Panel B shows results from a regression on price indicators. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.7: Effect of price paid on behavioral response to information

	Switched to safe well	Switched to safe or moderately contaminated well
	(1)	(2)
Mean across price groups	0.280	0.308
Price = Rs. 20	0.242	0.227
	(0.275)	(0.277)
Price = Rs. 30	-0.0326	0.00227
	(0.216)	(0.227)
Price = Rs. 40	0.0254	0.0292
	(0.228)	(0.226)
Price = Rs. 50	0.0424	0.0773
	(0.123)	(0.110)
Constant (mean at Rs. 10)	0.258***	0.273***
	(0.0981)	(0.1000)
Observations	211	211
R-squared	0.018	0.014
Joint significance		
Wald Chi2	0.96	1.13
Prob > Chi2	0.916	0.889

Note: the table shows the probability that households whose wells had unsafe arsenic levels ('red') switched to safer wells. Arsenic test results from 2012 data; self-reported switching data from 2013 follow-up survey. Column (1) considers switching only to wells with safe ('blue') levels of arsenic; Column (2) considers switching to safe or moderately contaminated ('green') wells. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.8: Selective recall of arsenic test outcomes

	Fixed on Well			Kept in House			Recall of Placard Color			All three combined		
Placard color	Red (1)	Green (2)	Blue (3)	Red (4)	Green (5)	Blue (6)	Red (7)	Green (8)	Blue (9)	Red (10)	Green (11)	Blue (12)
Second Phase	-0.0942***	0.0584	0.0358	-0.0925**	0.155***	-0.0621	-0.116***	0.0555*	0.0601	-0.0955***	0.118***	-0.0221
	(0.0262)	(0.0419)	(0.0374)	(0.0422)	(0.0513)	(0.0744)	(0.0253)	(0.0312)	(0.0435)	(0.0233)	(0.0317)	(0.0403)
N	1,529	1,529	1,529	1,379	1,379	1,379	1,762	1,762	1,762	1,840	1,840	1,840
R-squared	0.010	0.004	0.001	0.006	0.016	0.002	0.020	0.004	0.003	0.014	0.018	0.000

Note: the table compares the proportion of 'red' (unsafe), 'green' (moderately contaminated) and 'blue' (safe) wells in the recorded results of tests conducted in 2012, and in household recall obtained in the 2014 survey. The coefficient on 'second phase' reflects the difference in shares of each test result category. Columns (1-3) show results for households where placards indicating test results were still affixed to the well; Columns (4-6) show results for households where placards were no longer affixed, but were kept in the house and shown to the enumerator; (Column 7-9) show results for households where the placard was no longer available, but the respondent recalled the test result. Columns (10-12) pool all recall information on well status. The sample size reflects the sum of all tests recorded in 2012, along with the number of households for which information in a given category was available in 2014. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p < 0.1$.

Table 2.9: Selective recall and household assets

	Placard color red			
	On well	Kept in house	Recalled	All
	(1)	(2)	(3)	(4)
Second phase	-0.0831***	-0.0688	-0.0919***	-0.0760***
	(0.0285)	(0.0507)	(0.0286)	(0.0256)
HH owns consumer durables	0.0423	0.0423	0.0423	0.0423
	(0.0402)	(0.0405)	(0.0406)	(0.0397)
Second phase * HH owns consumer durables	-0.0571	-0.0661	-0.0903**	-0.0728*
	(0.0495)	(0.0662)	(0.0409)	(0.0407)
Observations	1,497	1,350	1,730	1,808
R-squared	0.012	0.007	0.023	0.016

Note: the table shows differences in the share of 'red' wells in 2012 tests and 2014 recall as in Table C, but conditional on ownership of (any) consumer durables. The coefficient on 'HH owns consumer durables' is the same across all four samples by construction: it is only the composition of the 2014 recall sample that changes, not the composition of the 2012 test sample. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.A Comparison of 2012 demand estimates based on recorded and recall sales data

As noted in the main body of the paper, during the first offer phase in 2012, enumerators did not systematically collect data from all households - chiefly, some households that did not want to purchase the test were omitted. (This is evident in the comparison of Columns 2-4 in Table 2.A.1.) In addition, anecdotal evidence raises a concern that enumerators may have offered tests less systematically in parts of the villages where people showed strong reservations against the idea of arsenic tests being offered for a fee (rather than free of charge) during focus group meetings.

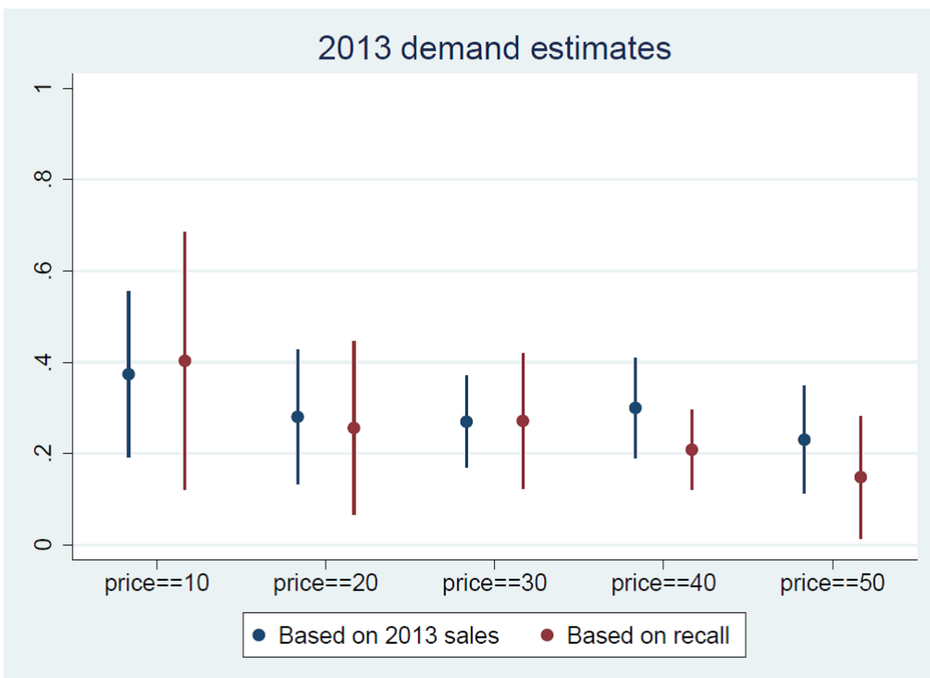
We hence face a considerable challenge in reliably assessing baseline demand, since the number of households to whom the test was offered in 2012 cannot be completely ascertained. We address this challenge with the following strategy. (1) We first compute demand based on recall data collected in the 2014 follow-up survey (i) on whether households were offered the test at baseline, and (ii) on whether they purchased the test at baseline. (Table 2.A.1, Columns 5-6.) This estimate is correct to the degree that there is no correlation between the decision to purchase in 2012 and recalling the offer when surveyed in 2014.

To assess whether the recall-based estimate is reasonable, we also (2) estimate demand from the 2012 sales (Column 3), based on the assumption that as many households were approached during the 2012 campaign as during the 2014 campaign (Column 4). This estimate is correct to the degree that (i) sales approaches were comprehensive in 2012 (while numerators neglected to keep records of some visits), and (ii) the number of households has remained constant between survey rounds.

Reassuringly, as is evident from Table 2.A.1 and Figure 2.A.1, the estimates obtained by recall and by imputing the number of sales offers are well-aligned in the aggregate (27% and 30%, respectively) and in the Rs. 10-30 groups. They diverge more at higher prices, though never significantly so. As a corollary, there is a good match between the ratio of *recalled* 2012 sales to *recorded* 2012 sales (0.65) on the one hand, and the ratio between *recalled* 2012 offers and *recorded* 2014 sample size on the other (0.60). This suggests that recall error is similarly likely for offers and sales, and provides at least some reassurance that the 2012 data is affected by failure to record unsuccessful sales attempts, rather than selective sales attempts.

Although first-round data collection did not follow protocol completely, we are hence able to offer two sensible demand estimates, and show that they match up well with each other. In the main body of the paper, we discuss results based on recall data – arguably, the more internally consistent approach, as well as the more conservative demand estimate.

Figure 2.A.1: Comparison of demand estimate from first phase data and recall



Note: the plot shows demand estimates obtained by scaling recorded sales in the first round of offers (2012) to 2014 sample size, and from offers and sales recalled in 2014.

Table 2.A.1: Test offers, sales, and demand

	Recorded 2012 offers and sales			Recalled 2012 offers and sales		Demand estimates	
Price (Rs.) (1)	Recorded offers (2)	Recorded sales (3)	*Sample 2014* (4)	Recalled offers (5)	Recalled Sales (6)	2012 demand (recorded sales) (7)	2012 demand (recall data) (8)
10	431	361	*960*	615	249	0.38	0.40
20	423	310	*1105*	804	206	0.28	0.26
30	352	218	*815*	460	125	0.27	0.27
40	327	196	*653*	441	92	0.30	0.21
50	289	127	*551*	350	52	0.23	0.15
All	1822	1212	*4084*	2670	724	0.30	0.27

Note: the table summarizes data used in computing the 2012 demand estimates shown in Figure 2.A.1.

2.B Why is there substantial demand at the time of the repeat offer?

This appendix summarizes evidence on what might explain demand at the time of the repeat offer. On balance, the evidence is inconclusive. Patterns in wealth proxies are consistent with a contribution of growing income and wealth. We note, however, that this is at odds with the absence of a correlation of wealth proxies with sales price among buyers shown above. A test for learning that allows for a sound causal interpretation is consistent in sign, but not significant.

2.B.1 Wealth effects

There is mixed evidence on increased wealth as a driver of repeat offer demand. As reported above, we find that observable wealth does not correlate systematically with willingness to pay. Indeed, one of the two wealth proxies that does correlate – ownership of a latrine – can be read as a marker of difference in concern over health that might affect valuation of the arsenic test as much as it may speak to lower marginal utility of consumption.

Still, there are some good reasons to ask whether rising wealth may have to some degree contributed to generating additional demand.

The most important piece of *prima facie* evidence is the rapid economic growth Bihar experienced between sales rounds. Per capita real income rose precipitously, at a rate of about 10% per year between 2012 and 2014.[12] In line with such a favorable development, owership of consumer durables among households who purchased tests in the first round of offers (the one asset category we can reliably compare among both survey rounds, and the one group of consumers sampled in a consistent way) rose by 5pp from a baseline value of 23% between 2012 and 2014 (result not shown). Because the tests were offered at the same *nominal* price in both phases, inflation further reinforced this effect. In total, nominal per

[12] State GDP growth for India from `http://planningcommission.nic.in/data/datatable/data_2312/`

capita income grew by some 38% between the two offers.

Secondly, patterns in asset ownership among buyers groups and across time are consistent with a wealth effects – though they do not offer a very powerful test. Our data allows in principle for two tests to reject wealth effects (at the mean). Most obviously, we can compare wealth among the two groups of buyers *at the time of purchase*, that is, in 2012 and 2014, respectively. This comparison could furnish some evidence against wealth effects if it were to emerge that second-round buyers were less well-off at the time of purchase than first-round buyers were at the time their wells were tested (with the assumption that the two groups initially had the same valuation of the tests). We can only draw this comparison on the ownership of (any) consumer durables; questions used to collect ownership information for all other categories of assets differed too much between the 2012 and 2014 surveys. For consumer durables, there is no significant difference between buyer groups, and the coefficient is centered near zero (Panel A in Table 2.B.1). This finding is consistent with wealth effects (new buyers catching up in wealth to original buyers), but also does not exclude a contribution of learning.

Beyond the ownership of consumer durables, we are constrained to comparing wealth as observed in the year 2014: among households that bought in 2012 and households that bought in 2014. This comparison could also reject wealth effects, namely if second-round buyers were weakly better off in 2014 than first-round buyers (and we were willing to assume that growth in wealth among the two groups was such that the ranking was not reversed since 2012 – which would then imply, less appealingly, that the wealthier group initially had a lower valuation of the tests). Our data suggests quite clearly that the opposite was the case: first-round buyers were better off than second-round buyers when surveyed in 2014 (Table 2.B.1). Difference in ownership of durables such as TV and consumer durables are significant, second round buyers have significantly less education than first round buyers, and there are notable differences in caste composition.[13]

[13]We note that, strictly speaking, we are comparing between one group observed pre-treatment (2014

2.B.2 Learning

Arsenic tests in themselves are distinctly a non-experience good: a one-off application which does not directly affect the consumer. It is therefore most plausible to suggest that learning might be chiefly driven by increased awareness of the probability of arsenic contamination, and of opportunities to switch to safe wells.

We test in the following way for evidence of learning after the first wave of tests. Because the distribution of arsenic in ground water varies substantially and unpredictably over small distances, variation in the results of first-round tests is exogenous. We posit that different distributions of first-round results at the village level may induce differential effects on second-round demand. In particular, we speculate that, when a high share of wells tested 'unsafe' during the first wave, concern in the village community over arsenic contamination might have been raised, translating into learning – namely, greater awareness of the health risks associated with arsenic, and the benefits of testing and well-switching. Empirically, the relationship between second-phase purchases and the share of wells tested 'unsafe' in the first phase has the expected sign, across a range of specifications (Table 2.B.2). However, results are not significant with cluster bootstrap standard errors.

buyers) and one group observed post-treatment (2012 buyers). However, since the health effects of Arsenic are long-term, one would not expect a strong treatment effect a mere two years after the test, even conditional on households effectively avoiding exposure. We acknowledge that in principle, Arsenic testing could have had effects upon wealth through conduits other than health – for instance, a change in the value of houses with wells tested safe/unsafe, or a change in social status with implications for future wealth.

Table 2.B.1: Household characteristics of first and second phase buyers

	Panel A: as observed at time of purchase		
	2014 buyers	2012 buyers	2014 vs. 2012
	(1)	(2)	(1) - (2)
HH has consumer durables	0.225	0.226	-0.00135
	(0.0404)	(0.0276)	(0.0392)

	Panel B: as observed in 2014		
	2014 buyers	2012 recall	2014 vs. 2012 recall
	(1)	(2)	(1) - (2)
Household characteristics			
Number of HH members	4.919	4.311	0.608
	(0.367)	(0.325)	(0.382)
Infant living in HH	0.302	0.223	0.0798**
	(0.0459)	(0.0246)	(0.0370)
Child living in HH	0.488	0.438	0.0497
	(0.0585)	(0.0618)	(0.0657)
Housing characteristics			
House pucca	0.701	0.756	-0.0553
	(0.0556)	(0.0504)	(0.0391)
Has latrine	0.330	0.408	-0.0778
	(0.0551)	(0.0496)	(0.0553)
Asset ownership			
HH has consumer durables	0.225	0.301	-0.0766*
	(0.0404)	(0.0563)	(0.0405)
Has cell phone	0.912	0.861	0.0507
	(0.0230)	(0.0578)	(0.0460)
Has TV	0.208	0.298	-0.0905**
	(0.0372)	(0.0573)	(0.0424)
Has bicycle	0.783	0.811	-0.0285
	(0.0187)	(0.0402)	(0.0382)
Has motorbike	0.248	0.261	-0.0131
	(0.0254)	(0.0243)	(0.0260)
Has cow	0.680	0.680	6.24e-05
	(0.0417)	(0.0319)	(0.0353)
Caste			
Scheduled caste or tribe	0.0163	0.0386	-0.0223
	(0.00852)	(0.0240)	(0.0226)
Other backward caste	0.227	0.127	0.0995**
	(0.0518)	(0.0298)	(0.0411)
Kshatriya	0.0767	0.124	-0.0473
	(0.0309)	(0.0455)	(0.0371)
Brahmin	0.251	0.388	-0.137***
	(0.0658)	(0.0646)	(0.0510)
Baniya	0.297	0.203	0.0940*
	(0.0670)	(0.0446)	(0.0537)

Note: the table shows characteristics of households that bought tests in 2014 (Column 1) and 2012 (Column 2), and the change between the two phases (Column 3). Panel A shows ownership data as observed at the time of purchase; Panel B shows data as observed in 2014 – that is, 2014 values for those who buy in 2014 in Column (1), and 2014 values for those who recall having purchased in 2012 in Column (2). Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.B.2: Do first-round test results relate to second-round demand?

	Demand in Second Phase				
	(1)	(2)	(3)	(4)	(5)
Share of wells in village tested arsenic high (red) in first round	0.0384	0.0699	0.0437	0.0933	0.117
	(0.112)	(0.125)	(0.107)	(0.114)	(0.130)
	[0.0301]	[0.0384]	[0.0301]	[0.0326]	[0.0404]
Controls					
Price	Yes	Yes	Yes	Yes	Yes
First-round demand	No	No	Linear	Quadratic	Quadratic
Wealth proxies	No	Yes	No	No	Yes
N	4,084	3,002	4,084	4,084	3,002
R-squared	0.037	0.060	0.051	0.059	0.082

Note: the table summarizes the correlation between arsenic test outcomes in the first phase and the demand in second phase. In each column, the dependent variable is demand for well tests in the second phase of offers, and the coefficient of interest is the share of wells that tested 'red' (high arsenic) among wells tested in the first offer phase. All models include price controls; Columns 3-5 control for first-round demand, and Column 5 controls for wealth proxies. We consider Column 4 to show the preferred specification. Cluster bootstrap standard errors (400 replications) in parentheses, classical standard errors in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.C.1: Sorting on well status

	High arsenic well ('red')	
	Well characteristics	Wealth proxies
	(1)	(2)
Well age	-0.00234	
	(0.00297)	
Well depth	0.00114	
	(0.00129)	
Well cost	1.48e-06	
	(8.68e-06)	
		Coefficients from univariate regressions
Has car		0.172
		(0.132)
Has cellphone		-0.0148
		(0.0875)
Has more than one cellphone		-0.0558
		(0.0840)
Has TV		-0.00610
		(0.0615)
Has bicycle		0.0626*
		(0.0345)
Has motorbike		-0.0285
		(0.0426)
Has cow		0.102**
		(0.0407)
Has more than one cow		0.0529
		(0.0480)
Has consumer durables		0.0377
		(0.0657)
'Pucca' house		-0.0255
		(0.0572)
Has latrine		0.0981
		(0.0668)
Adult household members		-0.00480
		(0.00913)
Infants in household		0.0125
		(0.0202)
Children in household		-0.00866
		(0.0242)
Observations	677	719
R-squared	0.007	n/a

Note: the table shows correlations among wells tested in 2014, between the probability of a well having high arsenic status (at least $50\mu g/l$) with characteristics of the well (Column 1) and the household (Column 2). To avoid evident overfitting problems, regression coefficients show in Column 2 were obtained by performing univariate regressions for each characteristic. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.D.1: Decision to re-test depends on contamination status

	Well contamination status		
	Red	Green	Blue
	(1)	(2)	(3)
Test purchased in both 2012 and 2014	-0.0411	0.172***	-0.130*
	(0.0582)	(0.0598)	(0.0792)
Share among wells tested once only	*0.257*	*0.274*	*0.468*
Observations	719	719	719
R-squared	0.001	0.013	0.006

Note: the table compares the proportion of 'red' (unsafe), 'green' (moderately contaminated) and 'blue' (safe) wells in the recorded results of tests conducted in 2014, among households that recalled preciously purchasing a test, and households that recalled a prior offer, but no purchase. Arsenic levels are stable over time, so test results obtained in 2012 can be assumed to have been identical to those measured in 2014. Cluster bootstrap standard errors (obtained from 400 replications) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 3

Geolocation error and the use of DMSP-OLS night lights as high resolution wealth proxies

Geolocation error and the use of DMSP-OLS night lights as high resolution wealth proxies

Jan von der Goltz [†]

Abstract

Is night lights data from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) observed precisely enough to measure human activity at high spatial disaggregation? Night lights are routinely used as proxies of ground-based activity at the level of countries, sub-national regions, or metropolitan areas. Due to the data's resolution (30 arc seconds), they might also be useful in studying processes at much higher geographic disaggregation – for instance, at the level of towns or villages. Yet, DMSP-OLS data are recorded with geolocation error that could interfere with such uses. I use a new data set of 185 calibration sites that are small, bright, and remote, to assess the offset between the actual location of light sources and their recorded location in the most commonly used yearly night lights data product. The error is small enough to be ignored, even in applications where the spatial scales of interest are on the order of a few kilometers. Root mean square error is a mere 0.52km in zonal and 0.67km in meridional direction. I illustrate the potential and limits of very high-resolution applications by benchmarking light data on household asset wealth in all official localities in Mexico. Night lights are a strong proxy measure of cross-sectional wealth differences even within small administrative units, in particular in the poorest, least populous, and most dimly lit regions. However, the analysis of changes over time is more subtle.

[*] I thank Chris Small for guidance and advice, and seminar participants at Columbia University for useful feedback. Dheeraj Sharma and Liwei Mao provided excellent research assistance.

[†] School of International and Public Affairs, Columbia University

3.1 Introduction

It is known that satellite observations of night-time luminosity (henceforth, 'night lights') from the Defense Meteorological Satellite Program's Operational Linescan System ('DMSP-OLS') are subject to several sources of measurement error: overglow, saturation, imperfect alignment of measurements over time, and error in geolocation registration. One would expect such measurement error to matter particularly when the data is used to study processes at small spatial scales At the same time, due to the relatively high resolution of the gridded data products (30 arc seconds – about 1km at the equator), there is increasing interest in using the data precisely as a proxy for human activity at high spatial disaggregation – an immediate extension of their routine use as proxies of socio-economic phenomena at the level of countries, sub-national units, or metropolitan areas.

Understanding the characteristics and impact of measurement error is therefore highly relevant to the future development of the night lights research agenda. In this paper, I assess the magnitude of one such source of error, namely error in geolocation registration. Secondly, I conduct an extensive benchmarking exercise to test the limits of the viability of night lights as a welfare proxy at very high spatial disaggregation – namely at the level of individual towns and villages in Mexico, 2000-2005.

Previous work by Elvidge et al. (2004) (henceforth, 'Elvidge et al.', without publication year) demonstrated that there is error in the geospatial registration of night lights data ('geolocation error') – that is, objects known to have a certain true location are sometimes observed in the gridded nightlights data in a grid cell that does not contain the true location. This misalignment is often visible to the naked eye when night light images are overlaid, for instance, with coast lines. I extend the analysis in Elvidige et al. by assessing geolocation error (i) in a large set of calibration sites spread across the world that allows me to check for variation in error with location, and (ii) in the annual 'stable lights' composites data commonly used in social science research, rather than in individual orbit data. I confirm that there is error, but report that it is quite small – indeed, even smaller than measured for individual orbits at the set of test sites in Elvidge et al. In consequence, geopositional error interferes only mildly with the measurement of brightness at my calibration sites.

Based on this assessment, I then conduct a benchmarking exercise to investigate whether lights are a meaningful proxy of welfare at a very high level of spatial disaggregation, namely the level of all villages and towns in Mexico. This application is intended to test an extreme example of how one might think to use night lights at small spatial scales.

Night light data has long been used to study socio-economic phenomena such as urban extent or population density. More recently, in the wake of seminal work by Chen and Nordhaus (2011) and Henderson et al. (2012), it has been applied to estimating output at the national level and at the second and third-tier administrative level, as well as for geographic areas not defined by administrative boundaries (Alesina et al. (forthcoming); Doll et al. (2006); Ghosh et al. (2009); Michalopoulos and Papaioannou (2013)).

Some work has also studied the structure of lights within large cities in developing countries, to assess city development in China (Baum-Snow et al., 2012), and the impact of US military strategy in Baghdad (Agnew et al., 2008). Most similar to the application assessed here is a forthcoming paper by Storeygard (forthcoming) tracking city lights in 15 African countries, and a working paper by Abrahams (2015), who shows that production in the towns of Palestine correlates strongly with lights, as well as work by Min et al. (2013) and Min and Gaba (2014) tracking the signature of electrification for a sample of villages in Senegal and Mali, and in Vietnam, respectively.

However, relative even to the objects considered in these studies, many of the localities studied here are small, with a median population of less than one hundred. What is more, they are not necessarily remote from brighter objects, so that it is not possible to take the standard approach of employing an algorithm to bound the lit area associated with a town and measuring an aggregate of lights. The challenge is hence to identify with sufficient precision the light emitted by small and dim objects that may be in the vicinity of brighter objects. I show that even in such a highly challenging application, night lights can provide some information on welfare differences in space – even within small administrative units – and over time. I also demonstrate, however, that the relationship is substantially noisy, and in the case of changes over time, not easily modeled.

In the remainder of the paper, Section 3.2 summarizes methods and describes the data used, in particular the calibration sites collected for the assessment of geoposition accuracy. Results in Section 3.3 provide an assessment (a) of the extent of geospatial error in the annual gridded DMSP-OLS data product, and (b) resulting measurement error in brightness. Section 3.4 tests the viability of night lights data as a proxy of human activity at the level of towns and villages, by assessing their performance in tracking wealth differences. Section 3.5 concludes.

3.2 Data and methods

Lights data

I use yearly composites of DMSP-OLS data publicly available from NOAA-NGDC, and described, for instance, in Elvidge et al. (2009). (Image and Data processing by NOAA's National Geophysical Data Center. DMSP data collected by the US Air Force Weather Agency.) The composites aggregate lights that are 'stable', rather than ephemeral, in cloud-free overflights. Brightness measurements are expressed as 6-bit digital numbers (DN), ranging from zero to 63. The sensor saturates, e.g., over bright city centers. Noise has been filtered out of the data by using low-end luminosity thresholds set according to values observed in pixels known to be unlit. Noise filtering has obvious downsides in the detection of dimly lit locations, but seems indispensable in extracting a viable signal.

Calibration sites

To assess geolocation error in the night lights data, I use a purpose-built dataset of 185 calibration sites selected to be small, bright, and remote sources of light. The location of these sites is shown in Figure 3.1; they are listed in Appendix 3.A. Night light and high-resolution images of two calibration sites (randomly drawn from the sample) are shown in Figure 3.2. I select the calibration sites in a three-step process.

First, I collect a set of 519 objects that one might expect to fit three key criteria for viability as a calibration site, namely: (1) high peak brightness, (2) small spatial extent, and (3) remoteness from other sources of light. This set of candidate sites includes oil and gas wells on-shore and off-shore; refineries; mines; small islands with military bases, tourist facilities, or other installations; lighthouses; and Arctic and Antarctic research stations and settlements. I also include the calibration sites used in Elvidge et al. (2004), to be able to compare error in the individual orbit data assessed in their study to error in the annual composites used here. I obtain the geolocation of each candidate site from high-resolution imagery in Google Earth (precisely, I visually estimate the center point of the light-emitting installations at each calibration site, and record the geolocation of the center point).

Secondly, I obtain night light images of each of these candidate sites for the years 1992, 2003, and 2009. The image is centered on the grid cell containing the true position of the site, and extends ten grid cells in each cardinal direction. An algorithm selects the brightest pixel within a four-grid cell box around the center grid cell, as follows. (1) When there is a unique pixel with peak brightness, it selects this pixel. (2) When there are several equally bright pixels, it selects the pixel with the highest average brightness in the directly adjacent grid cells. (3) When several pixels have the same average surrounding brightness, it chooses randomly.

I visually screen these images, and select a sub-set of 185 calibration sites that meet the following criteria in at least one year: (1) the object is clearly visible; (2) it is compact; (3) there are no other objects close by that might be confounded with the calibration site; (4) its center is well-articulated – in particular, it either (i) has higher brightness than all of the adjoining grid cells, or (ii) while several pixels share the same peak brightness, the light structure is visually centered on a certain grid cell. This excludes sites that either present as a fuzzy area of low brightness, or as a very large source of light with a saturated center.

Thirdly, for the resulting 185 calibration sites, I obtain night light images for each mission-year from the 1992 F-10 data to the 2009 F-16 data. I use both visual inspection and automated evaluation to construct a baseline sample, and two differently defined auxiliary samples for robustness checks, as follows. (Images of all calibration sites from each mission-year are available in the supporting online materials, as is a table recording whether they were included in the sample.)

1. I select the baseline sample for calibration by (i) excluding observations where the grid cell with peak brightness had to be selected randomly, and (ii) visually assessing whether the image fits the criteria set out above for a given each mission-year. This is the case for 2,165 observations (i.e., unique combinations of calibration site and mission-year).

2. The first auxiliary sample consists of a sub-set of 1,137 observations within the baseline sample that, when inspected, provide a particularly crisp image of the calibration sites, and where there was a single grid cell with peak brightness.

3. The second auxiliary sample adds to the baseline sample an additional 1,718 observations in which the object is clearly visible and can be uniquely identified, but the centroid is less well-defined, and the grid cell with peak brightness may have been randomly chosen.

For each observation in the baseline and auxiliary samples, I take the following measurements. I assume that the grid cell with peak brightness within four grid cells of the centroid (obtained from the algorithm described above) records brightness at the calibration site. This assumption would appear to be natural, as long as (i) there are no other sources of light nearby, and (ii) overglow falls off monotonically away from the true source. Based on this assumption, I then record (1) the distance in terms of grid cell count in meridional and zonal direction between the grid cell where the calibration site should be recorded if there were no geolocation error, and the grid cell in which it is observed in the night lights data;

(2) the great circle distance (on the WSG84 spheroid) between the calibration site's true geolocation, and the centroid of the grid cell where the site is recorded in the night lights data; and (3) brightness in the grid cell in which the site is observed, as well as in the grid cell where it should be observed if there were no geolocation error. To compare to Elvidge et al.'s results, I also (4) calculate root mean square error (RMSE) in meridional and zonal direction: I square the great circle distance in meridional and zonal direction, take its mean across all observations in sample, and then take the square root of the mean.

Welfare data

For the purpose of assessing whether geolocation error interferes with the use of night lights as a wealth proxy for small localities, I use a 'marginality index' for each official locality in Mexico, recorded in 2000 and 2005 (Consejo Nacional de Población, México (CONAPO), 2000). The index processes Census information on eight population and housing characteristics, namely: the share of adults (1) who are analphabets, and (2) without primary education; the share of residences (3) with dirt floors, and without access to (4) sanitation, (5) electricity, (6) piped water, and (7) refrigeration; and (8) the natural logarithm of the average number of residents per room among residences in the locality. It is computed using principal component analysis, as is standard in the literature. Indices based on asset ownership, education, and housing characteristics have been shown to correlate well with core measures of welfare such as consumption or income, and are widely used to assess welfare in a data-poor environment. (See, e.g., Filmer and Pritchett (2001).) The dataset covers more than 100,000 localities, from hamlets of three inhabitants to Mexico City. Many of these settlements are very small: the median population is 87, and the mean population, 903. In addition, I use data on household expenditure from one round of a panel survey conducted for the Progresa-Oportunidades cash transfer program in the year 2000 (Secretaría de Desarollo Social (SEDESOL), 2000). It covers 506 villages, and some 25,000 households.

3.3 Results

3.3.1 Error in geolocation

Error is small, and lower than in the individual orbit data considered in Elvidge et al.

Error in recorded geolocations is very modest in the annual composite data. Geolocation is correctly registered in 30-39% of observations across the different missions. (Figure 3.3) Perhaps more impressively, in between 90-97% of observations, the error is no more than one grid cell. This corresponds to an RMSE across the missions of 0.51km in zonal direction, and 0.66km in meridional direction. (Table 3.1) Precision is very slightly higher when I use only the crispest calibration sites; it is somewhat lower if I include sites where the structure of lights is less clear (0.60km and 0.74km in zonal and meridional direction, respectively).[1]

The RMSE observed in the yearly composites is hence similar to the values of 0.74-1.13km found in Elvidge et al. (p. 288), if slightly lower. This holds when I measure error only at the test sites studied in Elvidge et al.: observed RMSE for these sites in the yearly composite data is 0.66 (0.61) in zonal (meridional) direction (Table 3.2). (Limiting the sample further to the years of data included in Elvidge et al. does not change the finding. It is worth noting that, in the yearly composites data, the sample for this set of calibration sites is small (90 observations across all years, and 47 for the common years), and differences cannot be interpreted with confidence.) Figure 3.3 shows that the direction of bias in the yearly data is also less distinct than in the earlier assessment, while the two sets of measurements agree in showing lower error in zonal direction than in meridional direction.

While I cannot conclusively assess what explains the higher apparent precision observed

[1] The relatively mild difference between the datasets is reassuring for the following (somewhat subtle) reason. When I limit the baseline set of calibration sites to those with the crispest structure of night lights, I reduce measurement error from including sites where my algorithm would perform poorly (for instance, those with diffuse dim lights). However, true error in the accuracy of recorded geolocation might be correlated with these visual characteristics of the calibration sites. This might lead me to underestimate the error in geolocation. The welcome implication of the relatively small differences between the baseline data and the restricted/expanded datasets is that any such error, if present, would appear likely to be small.

in my sample, one would expect geoposition error to mechanically decrease in the yearly composites, because each observation in the yearly data can be seen as a sample mean of individual orbit data. Other possible explanations seem less likely: sample composition is unlikely to be a driver, since, as mentioned, precision is quite similar whether measured in the full sample or at the Elvidge et al. sites.

Secondly, while I do employ a different search algorithm than was used in the earlier study, results are (as mentioned) not very sensitive to considering the auxiliary samples based on variations in the algorithm. Thirdly, although to avoid misattribution, I limit the search for the brightest pixel to locations no more than four grid cells away from the true location (as opposed to a maximum of five grid cells in Elvidge et al.), Elvidge et al. report very few instances of offsets greater than four grid cells; and even within four grid cells, my data (Figure 3.3) shows higher concentration near the true position than the earlier study.

Error changes with location only at high latitude

I leverage the geographic dispersion of my calibration sites to test whether precision and direction of bias varies systematically with location. Among most groups of sites, there are no conspicuous differences. (Table 3.2) The one clear pattern to emerge is an effect of high latitude on the direction of the mean zonal offset.

As is to be expected, because the grid cell width of 30 arc seconds corresponds to smaller zonal distances at high-latitude calibration sites than at lower latitudes, the average absolute offset in terms of gridcells is mechanically higher in zonal direction, but not in meridional direction. RMSEx is higher for Antarctic sites, but since there is no difference among the Arctic sites, this might be due to small-sample variation. However, in terms of the direction of the offset, for both groups of high-latitude calibration sites, the offset tends to be toward the West of the true position. This is robust to removing the mean offset for each mission-year, and to removing the predicted offset for a fifth-order polynomial in longitude (hence, flexibly controlling for any possible effect of longitude). Figure 3.4 shows the mean zonal offset in terms of grid cells for each group of calibration sites, ordered in sequence of rising mean latitude. A similar pattern is evident when I graph the relationship between offset and latitude for individual observations (removing the effect of mission year and longitude in the same manner). There is no comparably obvious relationship between the zonal offset and longitude, or between region and the meridional offset. (Results available upon request.)

The direction and magnitude of offset varies with mission-years, but without an obvious pattern

Both the magnitude and direction of the offset vary between missions, and over time within missions. Average offset distance over the lifetime of each mission ranges from a minimum in zonal (meridional) direction of 0.36km (0.46km) for F12 to a maximum of 0.43km (0.58km) for F15 (F10). (Figure 3.5) In terms of the direction of the offset, bias in zonal direction is relatively small (or statistically zero), with the exception of F15. Bias is somewhat more pronounced in meridional direction for all missions, without a clear pattern. Year-on-year variation is much more substantial both in mean offset distance (ranging between 0.29-0.55km in zonal and 0.35-0.88km in meridional direction), and in the mean direction of the offset (between -0.44 and 0.51 grid cells in zonal direction, and -0.54 and 0.85 in meridional direction). With sample sizes of 60-80 observations per mission-year (as compared to 200-600 per mission), this greater observed variation in performance is unsurprising.

3.3.2 Error in the measurement of luminosity

The results shown so far suggest that error in geolocation is small in the annual composite lights data. However, the decisive question for the possible use of night lights as proxies for human activity at high spatial resolution is whether brightness at the recorded centroid performs well as a measure of actual brightness.

Reassuringly, this appears to be the case in measuring levels of brightness, and with somewhat more noise, also in the measurement of changes in brightness. In the baseline sample, digital numbers obtained at the true location and the maximum digital number observed within four grid cells surrounding the true location correspond closely, with a bivariate R-squared of 0.97, and a median (mean) difference of 1 DN (1.8 DN). Similarly, the yearly aggregate data at the true location performs well, if somewhat less strongly, in proxying changes in peak brightness. The correlation for year-on-years changes in brightness as measured by the same mission is 0.85 (with a bivariate R-squared of 0.73).

The magnitude of the expected error depends on the brightness of the light source observed. As Figure 3.6 shows, levels of brightness are measured without error in just under 60% of observations at low peak brightness, and in just under 40% of observations at high peak brightness. (Notice that brightness may be recorded without error even when peak brightness is not observed in the grid cell that contains the calibration site. This is the case when the center grid cell and the grid cell with peak brightness have the same DN, but mean brightness is higher in the pixels surrounding the 'peak brightness' grid cell than in the pixels surrounding the center grid cell.) Error is virtually always less than 3 DN at low peak brightness, and in excess of 70% of observations at high peak brightness. It is very rarely larger than 8 DN at any peak brightness level. For changes in brightness, the right panel in Figure 3.6 shows similar non-exceedance patterns, shifted in correspondence with higher overall error. To remove any potential effect of the known discrepancies in sensor performance between missions, I compare only among observations taken by the same mission in different years of observation.

Two points are worth making. One, while absolute error is lowest for dimly lit sites, error relative to peak brightness is highest among them (consider, e.g., the level of noise implied by a roughly 50% probability of any error at a true brightness of 3-4 DN). Two, expected absolute error initially rises little in brightness, up to about 12 DN – and conversely, error relative to peak brightness declines steeply. These observations may help explain behavior encountered in using very low brightness thresholds in applications. For instance, Small et al. (2011) describe how spatial patterns in urban extent change profoundly when thresholds below around 8 DN are used. High noise levels at low true brightness may be a key driver of this behavior.

Thus, the data perform notably well in measuring brightness in the presence of geolocation error. However, some perspective is needed. My calibration sites are remote from other sources of light, and I have shown above that geolocation error is not very large. Hence, perhaps counterintuitively, the known phenomenon of overglow in the present setting works to reduce measurement error in brightness: geolocation error is such that the recorded brightness in most cases is the brightness within about one grid cell of the true location; even where there is an error of one grid cell, overglow from the single adjacent source of light ensures that brightness in the two grid cells will be highly correlated. This raises the question how well the uncorrected data will perform in proxying differences in welfare among small locations that may not be remote from other sources of light – and where overglow will therefore tend to attenuate differences in brightness. I address this question in the following application.

3.4 Application: using night lights to measure welfare in small, dimly lit locations

The cross-sectional relationship between brightness and welfare in small locations is strong

Figure 3.7 shows non-parametric and linear fits of the relationship between observed digital number and the marginality index for all 107,218 official localities in Mexico, in the year 2000. (Because the welfare level is expressed as marginality, higher index values indicate lower wealth.) I use F14 data for the year 2000, and use brightness observed in the grid cell containing the true location of each town (that is, I make no attempt to correct for geolocation error). As is evident, the relationship is strong. It is also non-linear, and marginality drops most steeply with brightness among the most dimly-lit localities. Conversely, the relationship between marginality and the natural logarithm of brightness is nearly linear, as shown in the subplot on the right hand side. Table 3.3 shows that R-squared for the linear fit over the entire range is about 0.15; it is 0.23 for a more fitted model including an indicator variable for unlit observations, and a linear term in the log of non-zero brightness. – This is a meaningful correlation, though below R-squared in regressions of GDP on average brightness in lower-resolution applications.

More importantly, there is a meaningful relationship between brightness and marginality even when I consider only the way localities differ within small neighborhoods. Specifically, I study deviations between values measured for each locality from the mean value in each municipality (municipio), the second-level administrative unit in Mexico. This is a demanding level of geographic disaggregation: there are 2,442 municipalities in my data, and with a median population of 32,224, these are not large administrative units. Figure 3.8 shows that there is a relationship between deviations in marginality from the municipality mean and deviations in brightness from the municipality mean. The (within) R-squared for the linear specification suggests that variation in night lights within municipalities accounts for some 6% of the variation in marginality. (Table 3.3, Column 2) The relationship is again non-linear in digital numbers; it is approximately linear in the log of brightness, with a within R-squared of 0.12 (allowing for an indicator for unlit observations, as above). Fit varies over the range of observed brightness – specifically, as I discuss below, night lights perform best as a measure of deviations in wealth in municipalities that are on average dimly lit.

Figure 3.9 compares the relationship between marginality and brightness as measured in F14 and F15 data. It also compares results using brightness at the true location or peak brightness within one grid cell of the true location (as shown above, the target site is very likely to be observed within this buffer). As is evident, there are slight differences in levels in the relationships; however, the shape is very similar (and while I omit confidence intervals for readability, so is the amount of noise). Regression results show somewhat steeper relationships between marginality and brightness at the true location than between marginality and brightness within one grid cell. The linear coefficient differs little between missions, though the fit is slightly better for F15 (results available upon request). Because there is little apparent difference in the measurements yielded by the two missions and selection algorithms, in the remainder of the paper, I report results from a baseline specification using F14 data and brightness at the uncorrected true position.

Night lights perform best as a measure of local wealth in marginal and dimly-lit areas

Since municipalities in Mexico vary enormously in population, wealth, and brightness, it is worthwhile asking where the relationship between brightness and wealth is most pronounced. Figure 3.10 shows the distribution of coefficients from regressions of marginality index values on night lights, *computed separately for the localities in each municipality*, over the mean brightness among the localities in the municipality. (Figure 3.10 shows coefficients only for municipalities with at least 20 localities, that is, those with reasonable sample size. With the exception of the most brightly lit municipalities, results are very similar when I graph all coefficients.) As is evident, the relationship is steepest (and t-statistics are highest) for the dimmest municipalities: it is most powerful for an average brightness of up to perhaps 15-20 DN, and viable up to at least 40 DN. This echoes, of course, the non-linear shape in the relationship between night lights and marginality in the cross-section, shown in Figure 3.6. Similar patterns emerge when I graph the steepness of the relationship between marginality and brightness over mean marginality (Figure 3.11), or over population size (not shown). While there are meaningful relationships at all levels, they are steepest among the poorest and smallest communities.

Changes in brightness track changes in welfare, but the relationship is subtle

For the purpose of tracking the evolution of localities over time, I compare changes in night lights to changes the marginality index between the years 2000 and 2005 for each locality. I discuss results obtained without inter-calibrating the night light data from the two mission-years; inter-calibration would clearly be appropriate (see Elvidge et al. (2009)), but does not materially change the results discussed here, while significantly complicating the exposition.

Simple regression results suggest that increases in brightness are associated with reductions in marginality – but the effect is extremely weak (Table 3.3, Column 3). However, closer analysis shows that this is due to the fact that the relationship between changes in brightness and changes in marginality is non-monotonic. This non-monotonicity arises due to the structure of the data studied here, in the following way. In two-thirds of localities, brightness decreases; in more than one-fourth, there is no change. Only 7% of localities increase in brightness. Of the non-negative changes, 91% are observed in localities that were unlit in the base year. That is, to the first order, the relationship between non-negative changes in brightness and marginality reflects how brightness relates to wealth in localities that had no visible lights in 2000. The relationship between decreases in brightness and marginality reflects mostly processes in localities that were moderately bright in 2000 (80% of the sample comes from localities with DN < 15, and 90% from localities with DN < 27). At the same time, localities with zero brightness in 2000 saw on average a 0.04σ increase in marginality, while localities with baseline brightness up to DN 15 saw a 0.07σ decrease.

As Figure 3.12 shows, this leads to a break in the relationship between changes in brightness and changes in marginality at zero – and hence, to a fallacy of composition when estimating a linear relationship on the full sample. (The upper panel shows the relationship in the full same, and the lower panel, for those values of changes in brightness where there were at least 100 observations – or about 0.1% of the sample.) Once I split the sample into observations that were unlit in the base year (Table 3.3, Column 4) and those that were lit (Column 5), a far more meaningful relationship emerges: for instance, localities that were unlit in the base year and recorded a low brightness of 4 DN in 2005 are expected to have experienced a decrease in the marginality index of one-fourth of a standard deviation.

Cross-sectional relationships between brightness and welfare hold across a broad range of measures

To ascertain whether the relationship between local brightness and welfare is specific to the particular welfare metric used thus far – the marginality index – I also assess the relationship between brightness and mean household expenditure, as well as brightness and the individual measures of wealth used in the marginality index.

For the purpose of studying expenditure, my sample consists of the 505 villages initially tracked in the evaluation of the Progresa-Oportunidades social welfare program. As is evident from Figure 3.13, the cross-sectional relationship in the expenditure data for the year 2000 is strong and approximately linear in the range of brightness values where nearly all villages are observed, up to perhaps a value of 10 DN. An increase in brightness of 1 DN is associated with an increase in expenditure of about 2.6%. (There is no relationship when I remove municipality means, as done above for the marginality index. This is an unsurprising limitation, given that the 506 villages are scattered across 191 municipalities.)

Secondly, Figure 3.14 shows the relationship between night lights and the individual components of the marginality index. While the relationship is strong and intuitive in each case, its shape differs across components. It is enticing to speculate about possible uses of these differences, and in particular, the near-discrete jump in electrification levels between unlit and lit localities. I leave further investigation for future work.

3.5 Conclusion

This paper has assessed the magnitude and characteristics of error in geospatial registration in the annual stable composite DMSP-OLS night lights data, and has illustrated how it impacts the use of night lights data for the purpose of tracking welfare at very high spatial resolution.

While my measurements confirm that there is error in recorded geolocations, error is small in the yearly composite data. In the overwhelming majority of cases, the calibration site is either recorded in the correct grid cell, or within no more than one grid cell of the correct location. With the exception of high latitudes, the error also does not vary systematically with location or mission. In consequence, at the calibration sites studied here, brightness at the true location closely tracks brightness at the observed location, without any correction applied to the data.

These results raise the prospect that, despite geolocation error, night lights might be successfully used without geographic adjustments to measure human activity – and in particular, wealth – at high spatial disaggregation. However, my calibration sites were selected to be remote from other sources of light; hence, the overglow phenomenon tends to abate geospatial measurement error. By way of contrast, in using night lights to study small localities that are not remote from other sources of light, one would expect geoposition error to compound measurement error induced by overglow. I therefore assess the suitability of the yearly composite night lights as a wealth proxy at very high resolution in a large benchmarking exercise, using data on all official localities in Mexico.

I find that there is a meaningful relationship between brightness and a standard marginality index even at this extreme level of disaggregation. A clear relationship also emerges between brightness and expenditure in a smaller sample of villages. Of particular practical importance is the observation that there is a stable relationship at the level of deviations from municipality means – and most strongly so in the poorest, least populated, and most dimly lit municipalities. Night lights may hence have a useful role to play in studying well-being in small locations, even when survey or administrative information is available at the level of quite fine-grained administrative units – such as for instance in poverty mapping.

An intuitive relationship can also be found between changes in brightness and changes in wealth over time. However, in the data analyzed here, the relationship is subtle in its dependence on brightness in the base year. It must be modeled carefully. This is of some concern for the use of nightlights as a proxy for changes in wealth in small localities. Local patterns in brightness and in the relationship between growth and baseline wealth will be important for how changes in brightness should be interpreted. Based on the data analyzed in this study, one would suggest that attempts to use night lights to proxy changes in wealth at very high spatial disaggregation should involve some benchmarking of the night lights signal on local welfare data. Results from a 'hands-free' approach, in which changes in brightness might be used without any validation as a pure proxy measure of changes in wealth, might be quite prone to misinterpretation.

In summary, an extensive calibration and benchmarking exercise suggests that DMSP-OLS data can provide some insight into wealth at extremely high spatial resolution – in a set of localities with a median population below one hundred, and among a set of villages with a mean expenditure per adult equivalent of a mere US$1.27 (2005 PPP). At the same time, it is clear that the application considered here pushes the night lights data to the limits of its usefulness. While the observed relationships are steep and intuitive, there is considerable variance, and the fit is loose. I leave for further investigation the question whether recently proposed corrections for overglow (Abrahams et al., 2015) might further add to the usefulness of night lights data at high resolution.

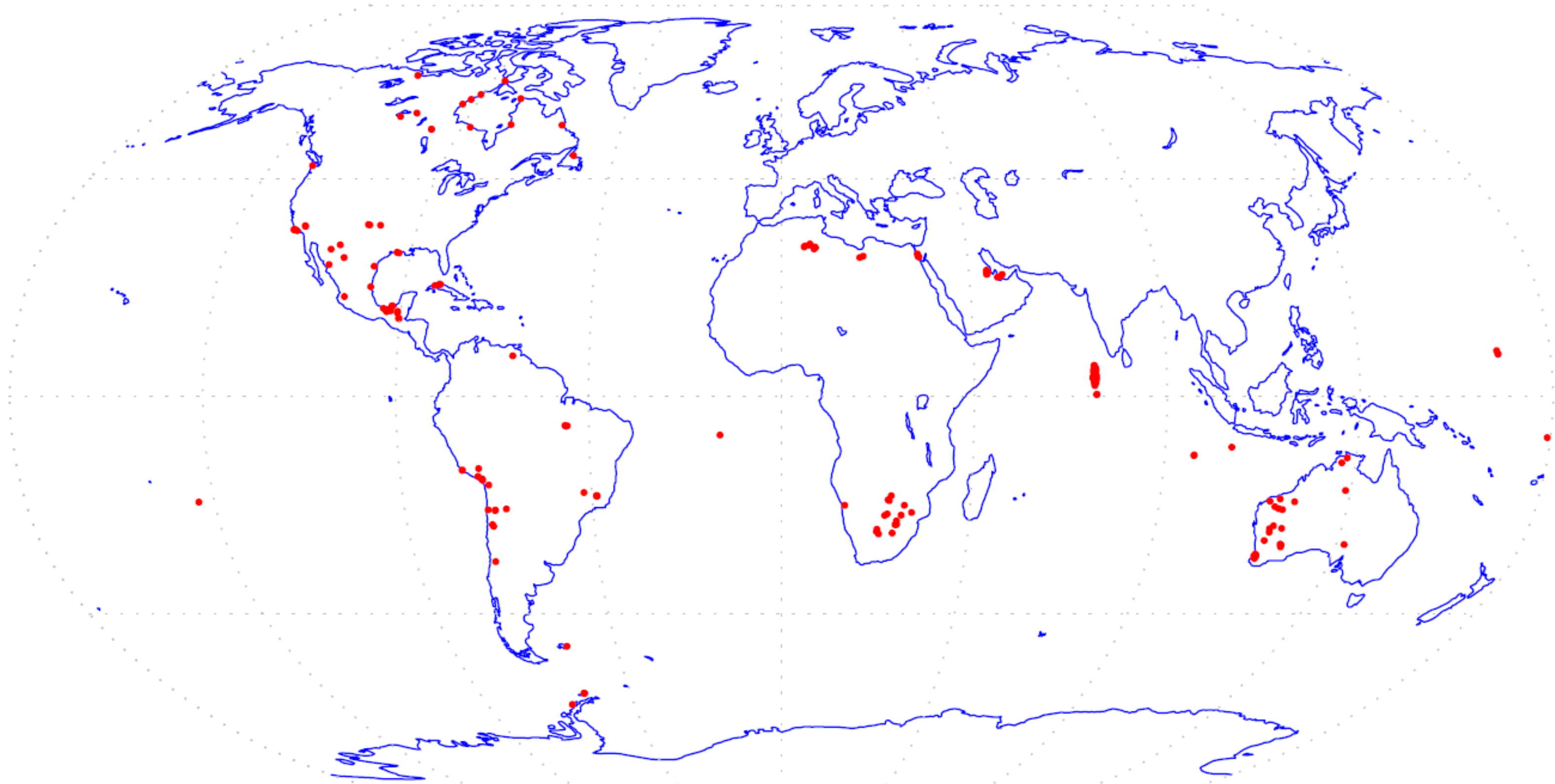
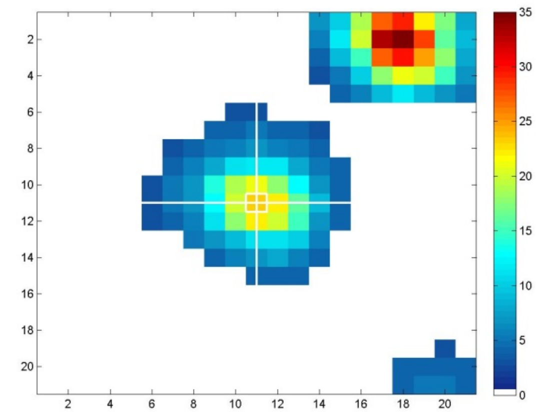


Figure 3.1. DMSP-OLS geolocation calibration sites

(a) Evaporation plant, Atacama desert, Chile – F14 2000 data.



(b) Ascension Island RAF base – F15 2005 data.

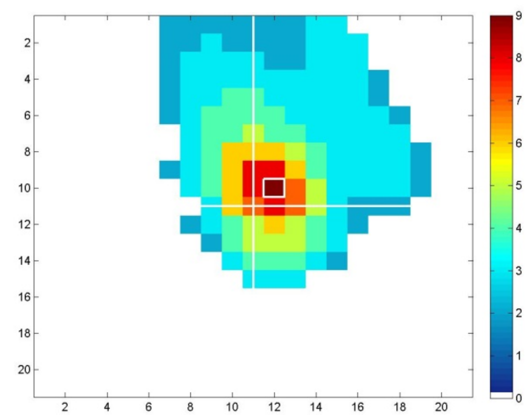


Figure 3.2. Examples of calibration sites.

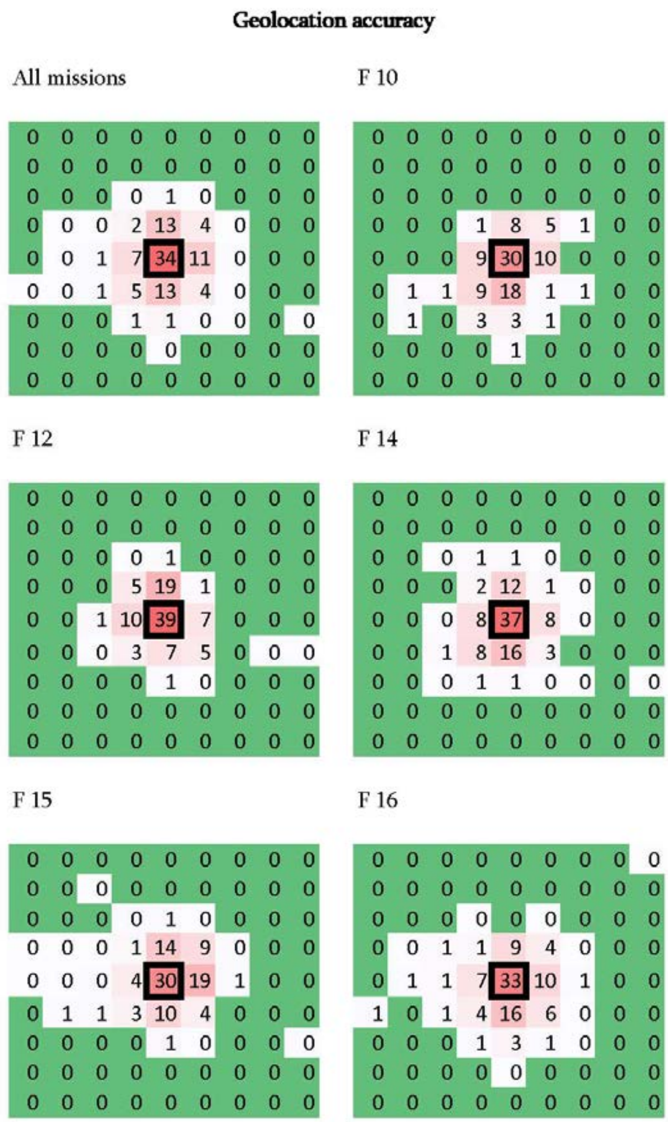


Figure 3.3. Distribution of pixels with peak brightness around the pixel in which the calibration site should be contained. The numbers in the grid cells reflect the percentage of observations in which peak brightness occurred in the respective grid cell. Darker shades of red indicate grid cells in which peak brightness is observed increasingly often. (Pixels shaded green are those where peak brightness is never observed.) The pixel that should contain the calibration site is shown at the center of each sub-plot, bounded in bold. The first sub-plot shows results when pooling all observations; the remaining sub-plots show results for each individual mission, as indicated.

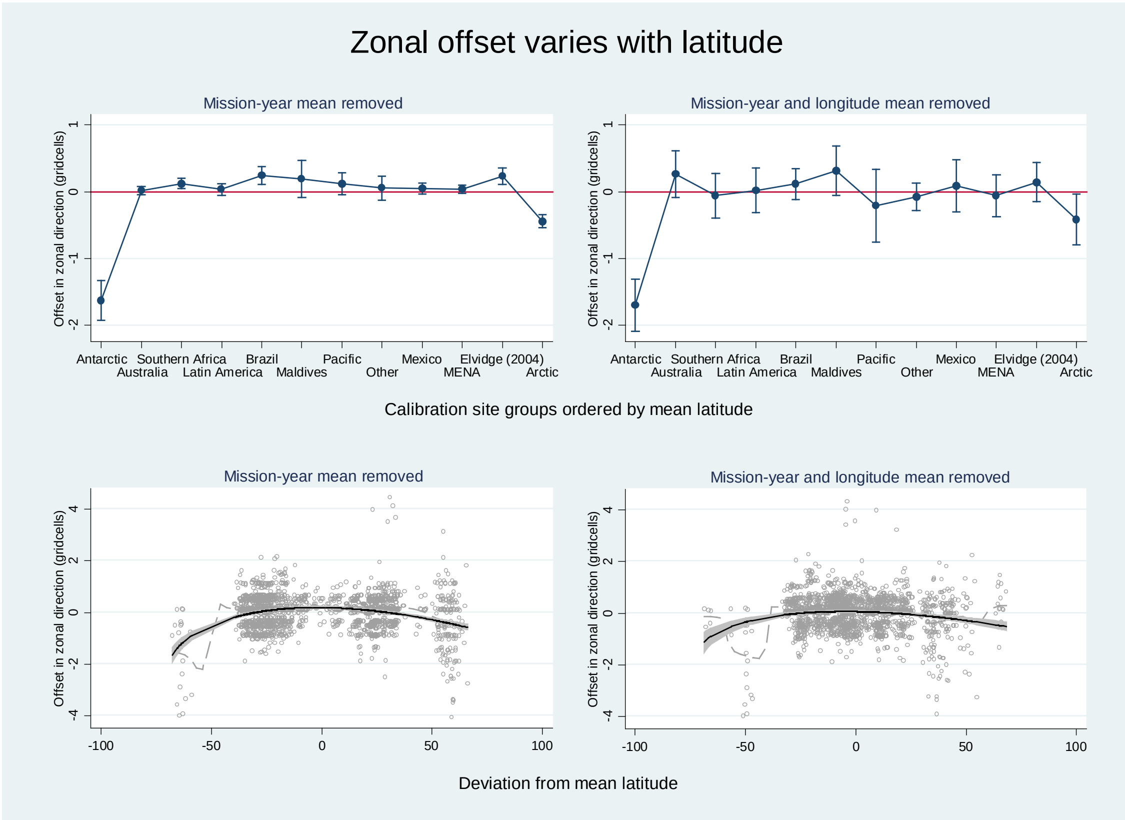


Figure 3.4. Zonal offset direction varies with latitude. The top panel shows mean offset (in terms of grid cells) for groups of calibration sites, ordered by increasing mean latitude. The lower panel shows the relationship between offset and latitude at the level of individual observations. The scatter plot shows the relationship for each observation. The solid black line shows a quadratic fit of the data, with its confidence interval shaded in gray. The dashed gray line shows a local polynomial fit. Offset estimates shown in the left-hand subplots remove potential mission-year effects by plotting deviations from mission-year means; estimates in the right-hand subplots additionally remove effects of longitude by plotting deviations from a fifth-order polynomial fit in site longitude.

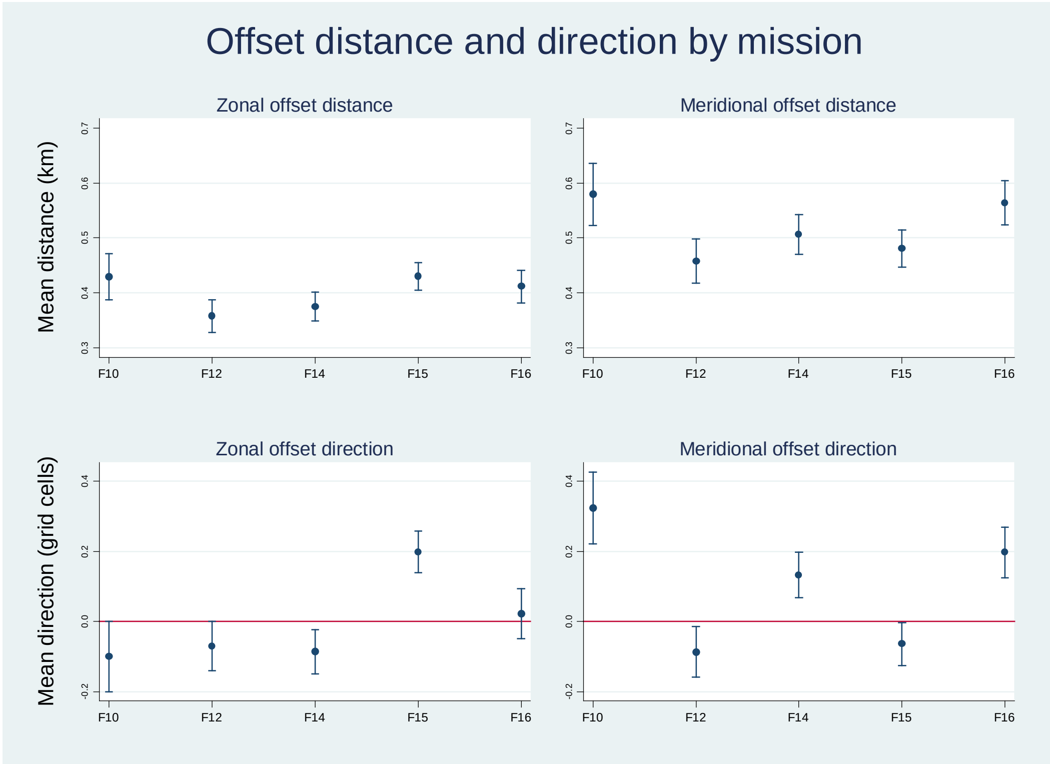


Figure 3.5. Offset distance and direction varies by DMSP mission. The figure shows mean offset (and confidence intervals) in zonal (left-hand side) and meridional direction (right-hand side) for the missions F10-F16. The top panel measures offset distance in kilometers; the bottom panel shows offset direction in terms of grid cell count.

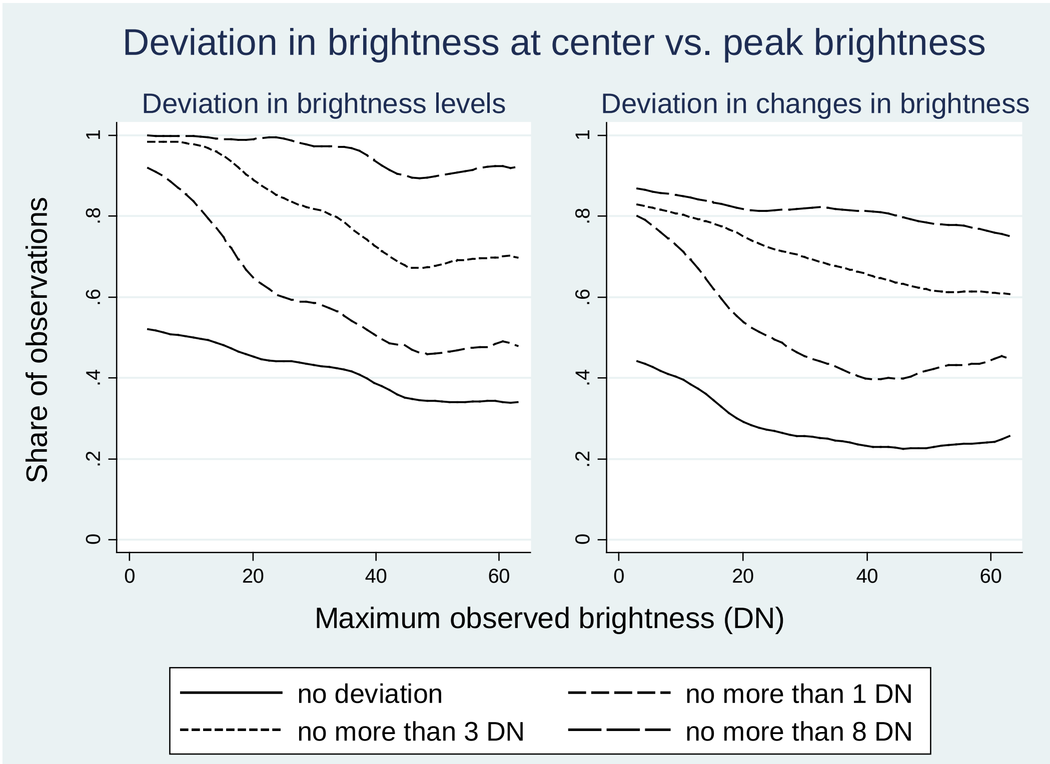


Figure 3.6. Probability that error in brightness measurement does not exceed certain thresholds, at different levels of peak brightness. The left-hand subplot shows error in the measurement of levels of brightness, and the right-hand subplot, in the measurement of changes in brightness over time. Error in the measurement of levels of brightness is defined as the deviation between peak brightness within four grid cells of the 'center' grid cell that should contain the calibration site, and brightness measured in the center grid cell. Error in the measurement of changes is defined as the difference between the change in peak brightness within four grid cells and the change in the center grid cell. The solid and dashed lines show the share of observations in which error does not exceed 0-8 DN. Note that brightness may be recorded without error even where geolocation is not accurately recorded. This is the case where there is no difference in brightness between the center grid cell and the peak brightness grid cell, but mean brightness in the surrounding pixels is higher for the peak brightness grid cell – which is therefore selected by the algorithm used in this paper.

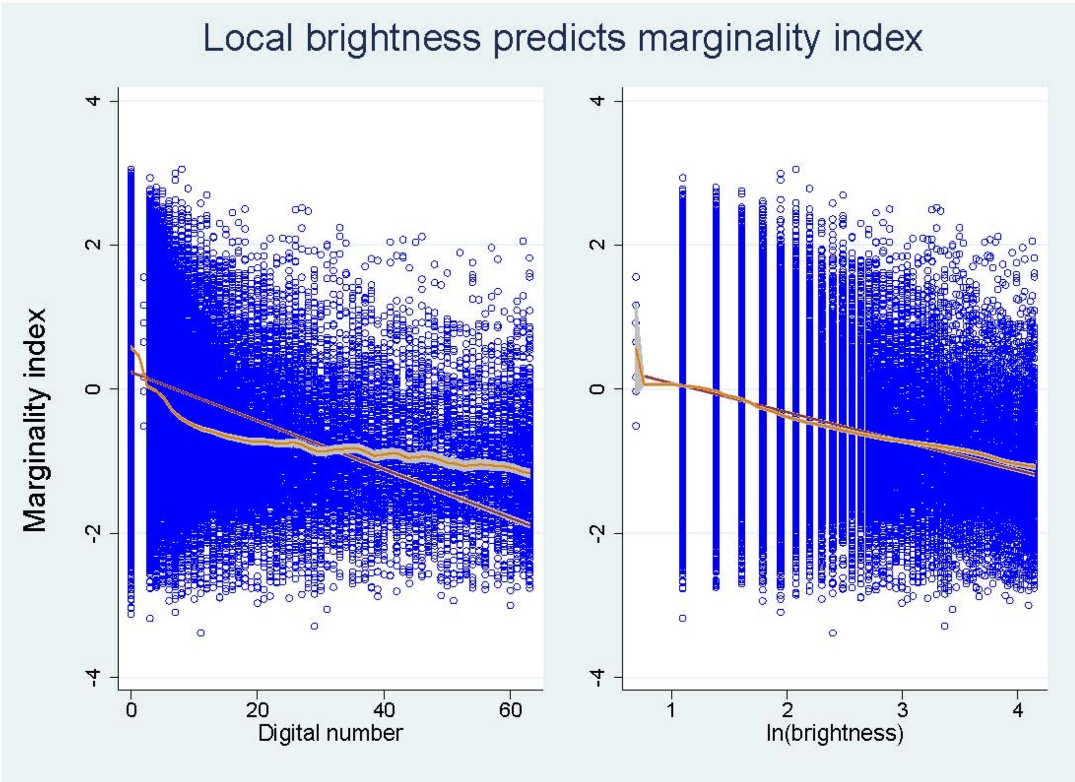


Figure 3.7. There is a strong relationship between brightness and marginality in the cross-section. Each subplot shows a scatterplot of locality-levels values, along with a linear and local polynomial fit, and their confidence intervals. Marginality is shown in units of standard deviations; brightness is expressed in digital numbers in the subplot on the left hand side, and as the natural logarithm of digital numbers in the subplot on the right hand side.

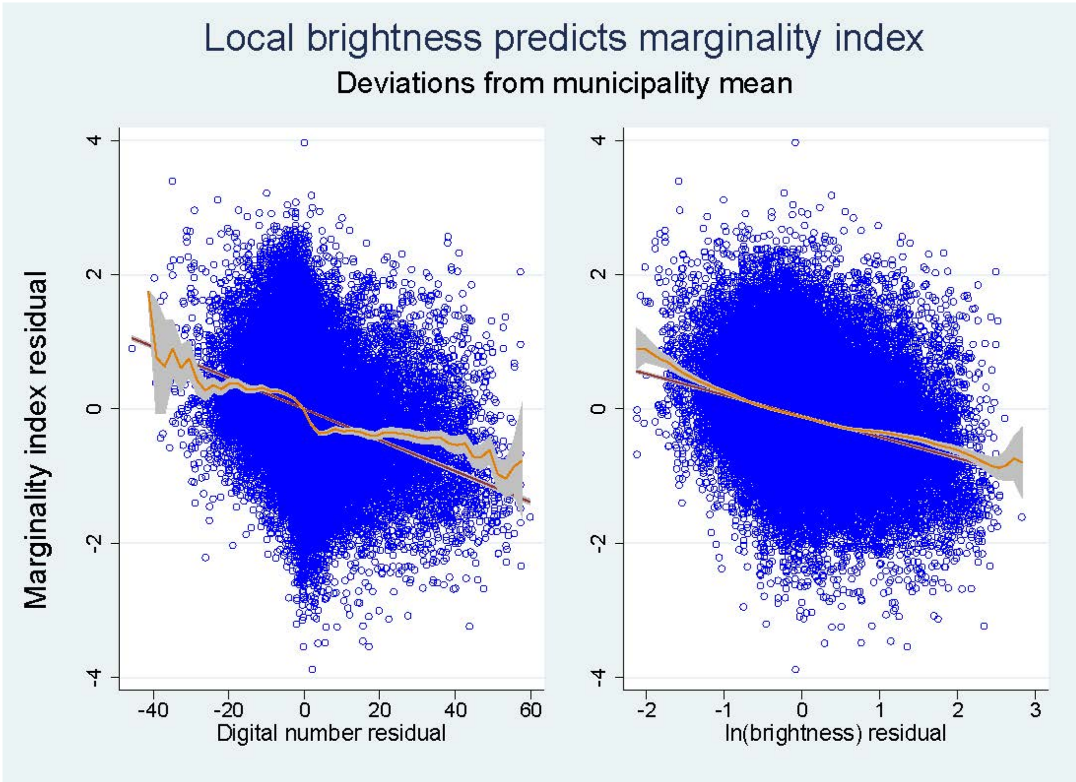


Figure 3.8. The relationship between brightness and marginality persists when considering deviations from municipality means. The figure plots deviations in marginality from the mean value in a municipality (in units of standard deviations) over deviations in brightness from the municipality mean – expressed in digital numbers in the subplot on the left hand side, and expressed as the natural logarithm of digital numbers in the subplot on the right hand side. Each subplot shows a scatterplot of locality-levels values, along with a linear and local polynomial fit, and their confidence intervals.

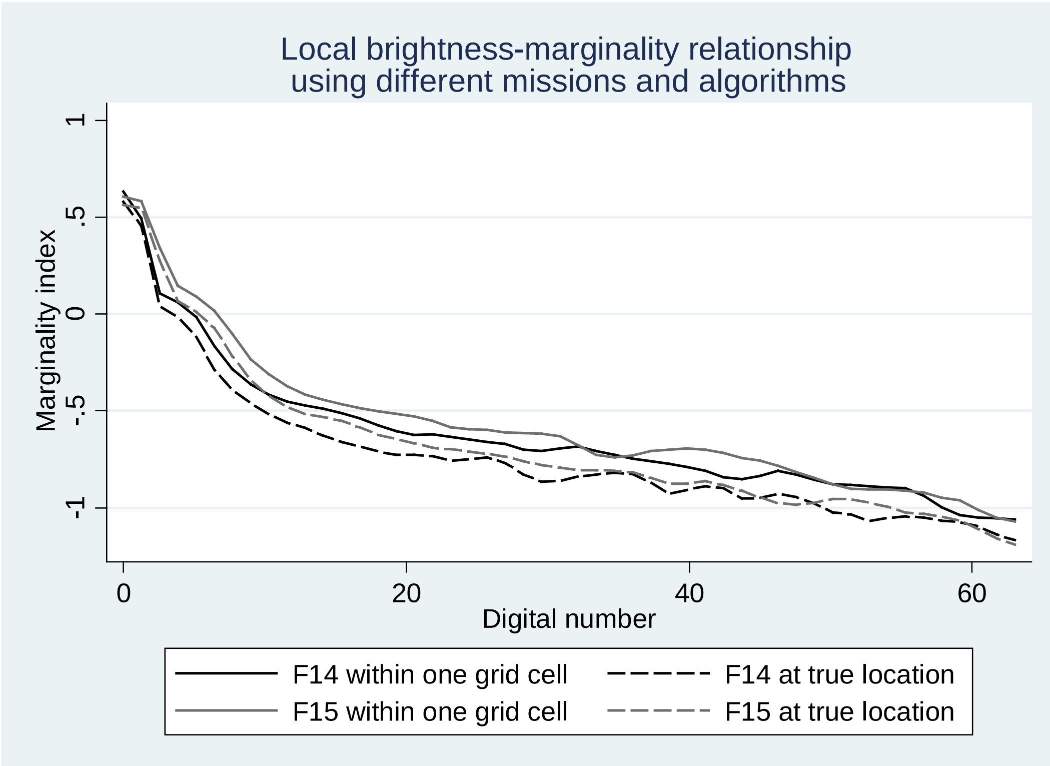


Figure 3.9. The relationship between brightness and marginality is robust to variations in data and data processing. The figure plots local polynomial fits of the relationship between brightness (in DN) and marginality (in standard deviations) for data obtained for all localities in the year 2000 from the F14 and F15 missions. It also shows the relationship when the measurement is taken in the grid cell containing the true location of the calibration site, and when I use peak brightness within one grid cell of the center, instead.

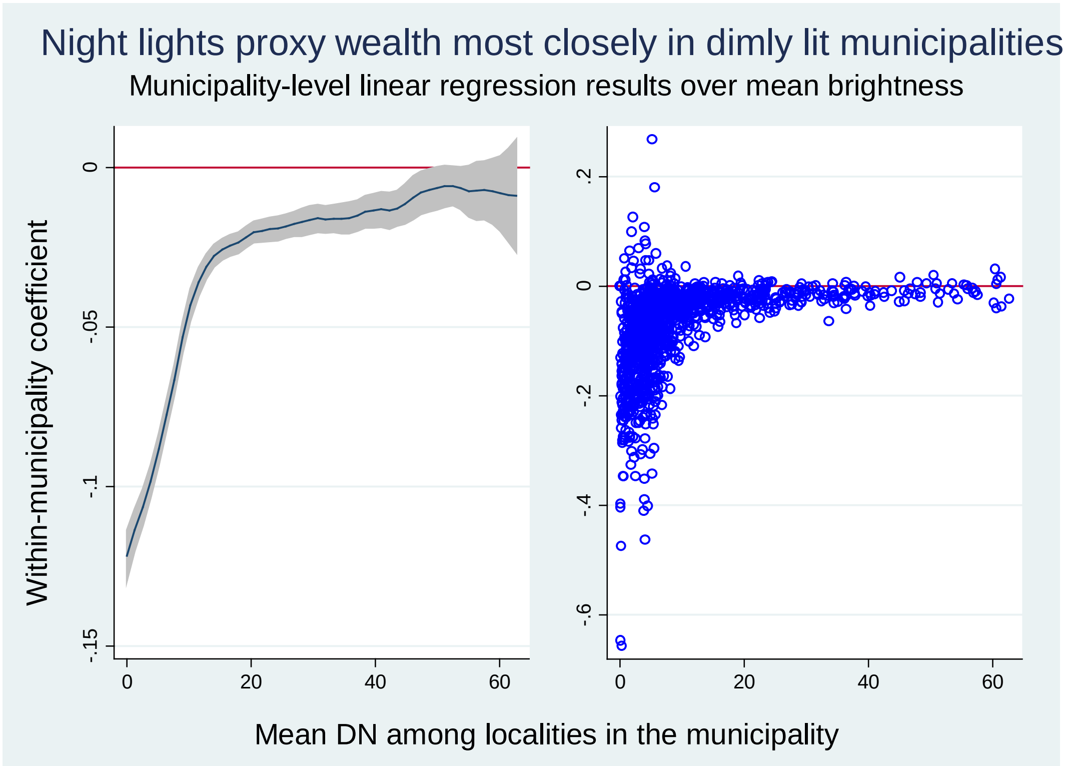


Figure 3.10. Locality-level marginality and brightness are most closely associated in municipalities that are dimly lit on average. The figure shows coefficients from linear regressions of marginality on brightness computed separately for each municipality in the sample. The subplot on the right shows a scatter plot of regression coefficients over mean brightness among the localities within the municipality for which the coefficient was computed. The subplot on the left shows a local polynomial fit of the data, with its confidence interval.

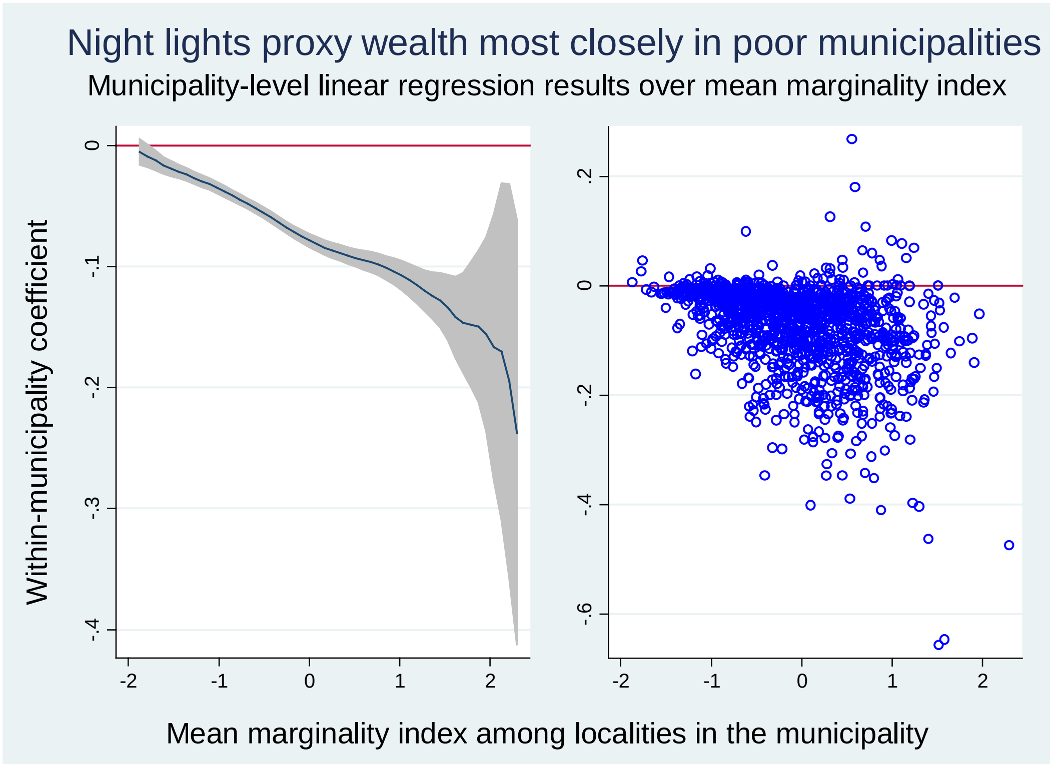


Figure 3.11. Locality-level marginality and brightness are most closely associated in municipalities where marginality is high on average. The figure shows coefficients from linear regressions of marginality on brightness computed separately for each municipality in the sample. The subplot on the right shows a scatter plot of regression coefficients over the mean of the marginality index among the localities within the municipality for which the coefficient was computed. The subplot on the left shows a local polynomial fit of the data, with its confidence interval.

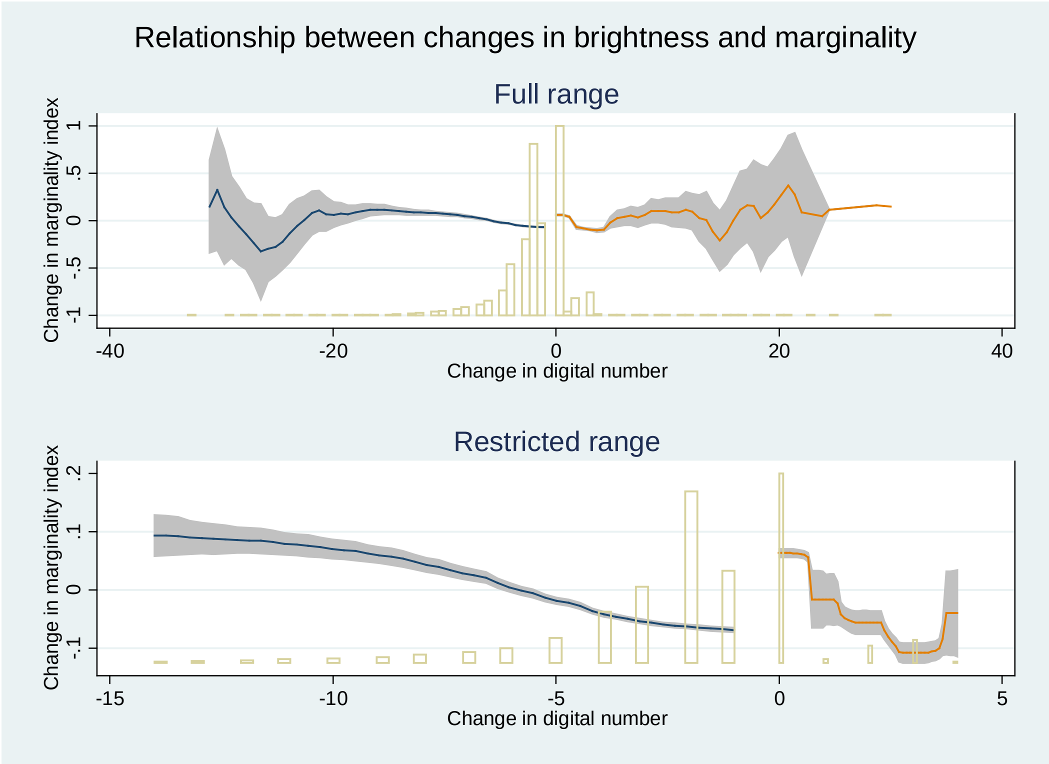


Figure 3.12. There is a strong relationship between locality-level changes in marginality and brightness, but in the data used here, it exhibits a subtle non-monotonic pattern. The figure shows local polynomial fits (and confidence intervals) of the relationship between changes in brightness (expressed in DN) and changes in the marginality index (expressed in standard deviations) at the locality level between the years 2000 and 2005. The histogram shows the distribution of changes in brightness. The subplot on top shows the relationship over the full range of observed changes; the subplot on the bottom shows in greater detail the relationship over the range of changes where there are at least 100 observations for each value.

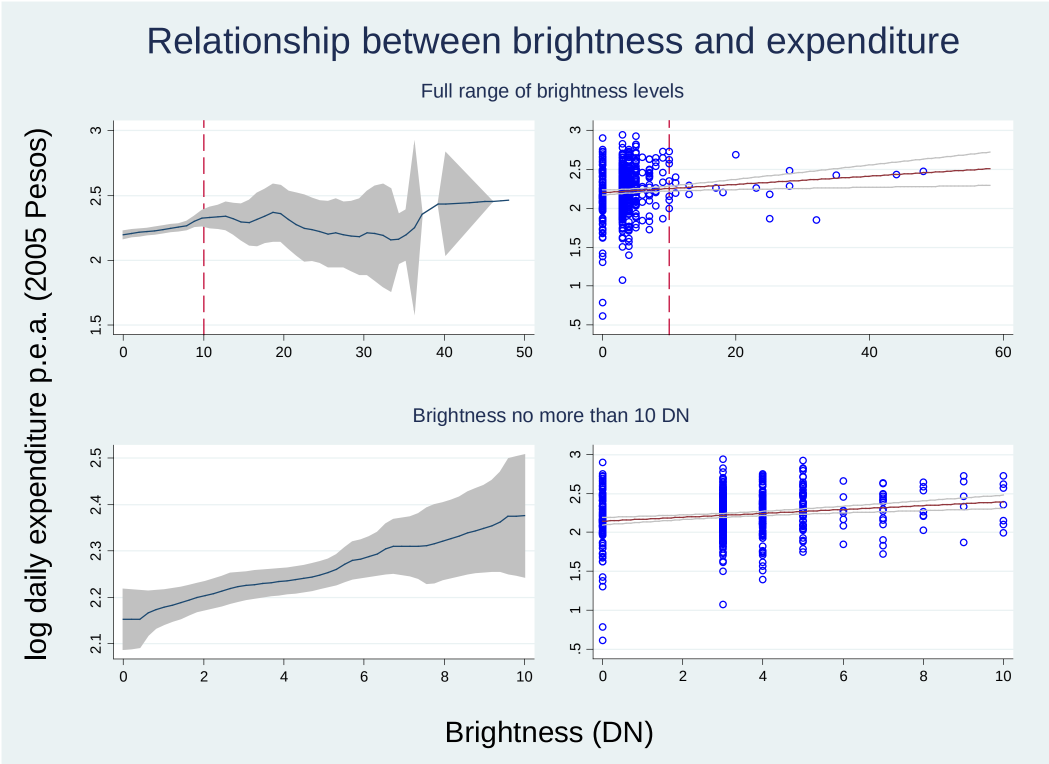


Figure 3.13. Cross-sectional relationship between brightness and expenditure at the village level in the year 2000 for villages surveyed for the Progresa/Oportunidades program. The subplots on the right show a village-level scatter plot of mean log expenditure per equivalent adult (in 2005 Pesos) over brightness (in DN), along with a linear fit of the data (and its confidence interval). The subplots on the left show a local polynomial fit of the same relationship, along with its confidence interval. The top panel shows the entire range of observed brightness values; the bottom panel shows in greater detail the relationship for brightness values not exceeding 10 DN.

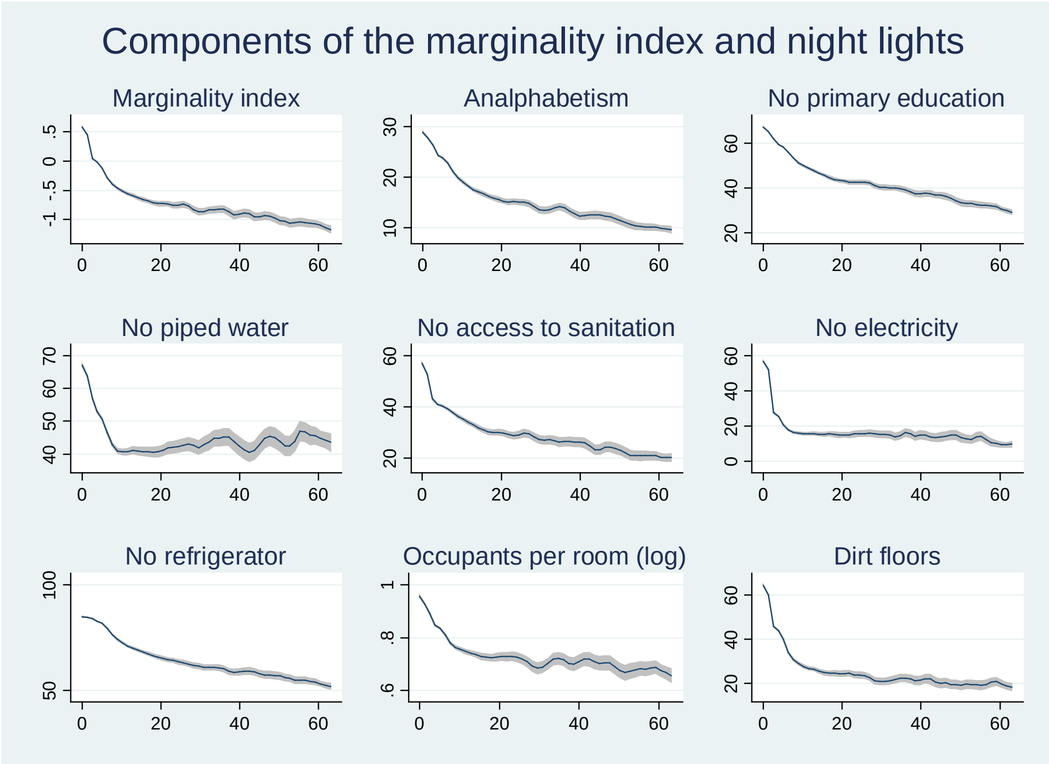


Figure 3.14. The individual indicators used in computing the marginality index exhibit different cross-sectional relationships with brightness. The figure shows local polynomial fits (with their confidence intervals) of locality-level values of the individual welfare indicators used in computing the marginality index to brightness, expressed in DN. The marginality index is expressed in standard deviations; analphabetism, lack of primary education, lack of access to piped water, sanitation, electricity, and refrigeration, as well as the presence of dirt floors in residences are given as percentages of the locality's population; the number of occupants per room is the locality average.

Table 3.1 - Offset characteristics

	Zonal (x) direction			Meridional (y) direction		
Sample	Baseline	Restricted	Expanded	Baseline	Restricted	Expanded
RMSE	0.51	0.48	0.60	0.66	0.63	0.74
Absolute deviation	0.43	0.35	0.52	0.51	0.48	0.56
Mean offset	0.01	0.04	0.01	0.07	0.05	0.10
Median offset	0	0	0	0	0	0
Maximum negative offset	-4	-4	-4	-4	-2	-4
Maximum positive offset	4	4	4	3	3	4
Observations	2165	1137	3883	2165	1137	3883

Notes: the table provides different measures of geolocation error in zonal and meridional direction. Results are shown for three samples, as described in Section 2 of the paper. Statistics provided include RMSE in km (obtained from calculations described in Section 2 of the paper); and the mean absolute deviation, mean, and median offset in terms of grid cells, as well as the largest offset in terms of grid cells in negative and positive direction. The number of observations refers to unique mission-year-site combinations.

Table 3.2 - Regional effects

	Observations	Zonal (x) direction			Meridional (y) direction		
		RMSE	Absolute deviation	Mean	RMSE	Absolute deviation	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Antarctic	19	0.96	1.58	-1.58	0.62	0.42	-0.32
Arctic	169	0.58	0.92	-0.44	0.67	0.57	0.34
Australia	489	0.42	0.36	0.02	0.73	0.59	-0.14
Brazil	93	0.55	0.45	0.24	0.70	0.48	0.05
Elvidge (2004)	90	0.66	0.42	0.33	0.61	0.43	-0.03
1992-2001 only	*47*	*0.64*	*0.34*	*0.30*	*0.54*	*0.40*	*-0.15*
Latin America	227	0.56	0.47	0.03	0.68	0.43	-0.03
Maldives	22	0.46	0.32	0.23	0.60	0.32	0.14
MENA	411	0.51	0.37	0.03	0.62	0.48	-0.05
Mexico	245	0.48	0.34	0.04	0.60	0.49	0.22
Other	50	0.44	0.30	0.06	0.68	0.72	0.60
Pacific	63	0.45	0.27	0.14	0.62	0.35	0.13
Southern Africa	266	0.48	0.39	0.12	0.63	0.45	0.29
All	2165	0.51	0.43	0.01	0.66	0.51	0.07
1992-2001 only	*1152*	*0.49*	*0.41*	*-0.01*	*0.65*	*0.49*	*0.09*

Notes: the table shows for each of the geographic regions represented in the calibration site sample different measures of geolocation error in zonal and meridional direction. Column (1) gives the number of unique observations (i.e., mission-year-site pairs) for each region. Columns (3) and (6) show the mean absolute deviation, and columns (4) and (7) the mean deviation in terms of grid cell counts. Columns (2) and (5) show root mean square error, computed as described in Section 2 of the paper.

Table 3.3 - Relationship between brightness and marginality index

	Cross-section		Deviations from municipality mean		Locality-level panel		
					All localities	Unlit in base year	Lit in base year
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Digital number	-0.0337***		-0.0231***		-0.00222***	-0.0624***	-0.0113***
	(0.000254)		(0.000285)		(0.000546)	(0.00309)	(0.000616)
Unlit		0.0906***		0.125***			
		(0.00928)		(0.00904)			
log(digital number)		-0.403***		-0.310***			
		(0.00371)		(0.00429)			
Observations	107,218	107,218	107,218	107,218	181,576	55,170	126,406
R-squared	0.150	0.237	0.059	0.129	0.002	0.020	0.015
Number of groups			2,442	2,442	90,788	27,585	63,203

Notes: the table shows regression results for a linear relationship between the marginality index (expressed in units of standard deviations) and brightness, experessed in digital numbers. Column (1) shows basic cross-sectional results; column (2) shows cross-sectional results when including indicator variables for each municipality in sample. Columns (3-5) show the relationship between changes in birghtness and changes in marginality over time at the locality level. Column (3) shows results unconditional on brightness in the base year; columns (4-5) disaggregate between localities that were unlit and those that were lit in the base year, as described in Section 4 of the paper. Significant at * 10%, ** 5%, *** 1%.

3.A Appendix – Calibration sites

Antarctic and sub-Antarctic			
32	RAF Mount Pleasant Falklands	58°27'36.00"W	51°49'37.71"S
42	Bellinghausen and Estrellas station (Russia, Chile)	58°57'43.63"W	62°11'58.82"S
43	Palmer station (US)	64°3'10.31"W	64°46'27.14"S
63	King Sejong Station (Korea)	58°47'17.34"W	62°13'23.14"S
Arctic and sub-Arctic			
148	Voisey Bay camp	62°4'47.25"W	56°24'40.12"N
149	Beaver Brook mine, Newfoundland	55°46'15.67"W	49°50'36.83"N
150	Ivujivik, Nunavut	77°54'41.70"W	62°24'58.30"N
151	Umiujaq, Nunavut	76°32'59.53"W	56°33'8.42"N
152	Fort Severn, Ontario	87°37'54.56"W	55°59'27.42"N
153	Arviat, Nunavut	94°3'39.49"W	61°6'29.61"N
154	Whale Cove, Nunavut	92°34'44.69"W	62°10'23.14"N
155	Chesterfield Inlet, Nunavut	90°42'15.43"W	63°20'26.74"N
156	Repulse Bay, Nunavut	86°14'12.03"W	66°31'21.62"N
157	Kugluktuk, Nunavut	115°5'52.42"W	67°49'31.77"N
158	Pipe Lake mine, Manitoba	98°9'39.84"W	55°29'43.39"N
159	Black Lake mine settlement, Saskatchewan	105°36'12.71"W	59°7'45.41"N
160	Paint lake mine, Manitoba	98°9'40.00"W	55°29'44.44"N
161	Cluff lake mine, Saskatchewan	109°35'45.43"W	58°21'25.38"N
Australia			
103	Roxby Downs mining town	136°53'59.49"E	30°33'42.75"S
104	Jabiru mine	132°55'27.59"E	12°40'57.58"S
125	Murray Basin mine	116°27'27.12"E	32°56'24.17"S
126	Mine, Darling Range	116°21'30.95"E	32°44'24.68"S
127	Mine, Hamersley Range	118°40'25.50"E	21°10'55.74"S
128	Mine, Hamersley Range	122°11'45.17"E	21°44'16.53"S
129	Mine, Kalgoorlie area	121°36'53.80"E	31°2'20.02"S
130	Mine, Kalgoorlie area	121°34'42.75"E	30°36'12.93"S
131	Plant, Kalgoorlie area	121°27'23.65"E	30°35'21.58"S
132	Mine, Kalgoorlie area	121°21'3.73"E	30°29'19.37"S
133	Mine, NT	133°49'18.71"E	19°26'40.44"S
134	Mine, NT	131°47'5.33"E	13°43'2.80"S
135	Mine, W Australia	116°4'11.27"E	33°14'18.78"S
136	Mine, W Australia	116°18'21.89"E	33°26'48.53"S
137	Mine, W Australia	115°54'53.14"E	32°55'0.44"S
138	Huntly Bauxite mine	116°9'38.09"E	32°33'36.45"S
139	Mine, W Australia	117°9'36.19"E	29°45'17.33"S
140	Mount Magnet mine	117°48'55.28"E	28°1'47.15"S
141	Weld Range mine	117°38'59.69"E	27°19'1.52"S
142	Mine, W Australia	118°25'52.58"E	26°42'49.50"S
143	Mt Keith mine	120°33'6.38"E	27°12'50.63"S
144	Tom Price mine	117°46'21.35"E	22°44'50.22"S
145	West Angelas mine	118°45'58.39"E	23°10'10.13"S
146	Mt Whaleback mine	119°41'20.21"E	23°22'5.11"S

147	Pannawonica mine settlement	116°19'30.62"E	21°38'15.40"S
Brazil			
58	Mine, Minas Gerais	43°52'39.05"W	20°25'9.50"S
59	Mine, Minas Gerais	46°50'51.20"W	19°50'40.77"S
112	Smelter, Minas Gerais	43°45'36.86"W	20°32'50.96"S
113	Plant, Minas Gerais	43°57'58.00"W	20°34'59.90"S
114	Mining town, Para	50°4'4.93"W	6°4'13.46"S
115	Mine, Para	50°18'4.37"W	6°6'50.90"S
116	Mine, Para	50°34'44.31"W	6°1'53.61"S
Elvidge (2004) replication			
18	Gaviota 1	120°16'42.25"W	34°21'5.69"N
19	Channel Island 3	119°24'3.60"W	34°7'31.20N"
20	Gaviota 2 West	120°10'1.41"W	34°22'40.41"N
21	Gaviota 2 East	120°7'13.43"W	34°23'29.53"N
22	Structure East of CA City	117°51'31.79"W	35°9'11.38"N
23	Gaviota Plant	120°12'10.13"W	34°28'26.23"N
Latin America			
16	Large mine, Peru	75°7'33.39"W	15°12'14.89"S
34	Industrial complex, Anzoategui, Venezuela	62°50'33.45"W	8°21'23.71"N
38	Mine, Tacna, Peru	70°37'33.35W	17°16'25.14S
39	Mine, Moquegua, Peru	70°43'22.77W	17°2'27.90S
40	Mine, Parinacota, Chile	69°16'37.93W	18°19'33.92S
46	Mine, Arequipa, Peru	71°35'29.38"W	16°31'4.38"S
51	Mine, Pasco, Peru	70°46'18.02"W	17°3'53.13"S
52	Mine, Tacna, Peru	71°18'46.59"W	14°53'20.71"S
55	Mine, Chile	70°27'12.03"W	34°5'54.81"S
56	Mine, Calama, Chile	70°3'46.31"W	23°26'21.18"S
57	Mine, Chanaral, Chile	69°28'48.38"W	26°26'18.69"S
62	Mine, Jujuy, Argentina	65°40'53.40"W	23°12'47.86"S
95	Port facilities, Ica, Peru	75°14'11.98"W	15°15'33.03"S
109	Chile evaporation plant	68°20'12.57"W	23°29'17.16"S
110	Chile evaporation plant	68°23'30.81"W	23°33'51.82"S
111	Mine, Atacama, Chile	69°15'55.12"W	26°49'44.22"S
Maldives			
28	Gulhi Falhu harbor	73°26'46.21E	4°10'57.87N
29	Bandos resort island	73°29'29.36E	4°16'8.06N
64	Kanduhulhudhoo	73°32'21.40"E	0°21'4.21"N
65	Gemanafushi	73°34'7.08"E	0°26'33.18"N
66	Buruni	73°6'27.63"E	2°33'32.03"N
67	Thimarafushi	73°8'36.06"E	2°12'20.23"N
68	Bandidhoo	72°59'26.79"E	2°56'12.06"N
69	Velavaru Island	73°0'58.54"E	2°58'47.66"N
70	Kolhufushi	73°25'27.83"E	2°46'37.76"N
71	Resort-Alimatha Island	73°29'52.27"E	3°35'37.46"N
72	Resort-Dhiggiri	73°29'14.74"E	3°38'42.41"N

73	Mirihi Island Resort	72°46’50.30"E	3°37’10.56"N
74	Hotel Lily Beach Resort	72°57’13.88"E	3°39’12.93"N
75	Diamonds Thudufushi Island Resort	72°43’51.44"E	3°47’9.87"N
76	Madhibadhoo	72°58’7.36"E	3°45’26.15"N
77	Constance Moofushi Resort	72°43’40.62"E	3°53’5.30"N
78	Madoogali Resort	72°45’12.38"E	4°5’43.89"N
79	Havaveli Resort	72°55’10.17"E	4°2’17.72"N
80	Maayafushi Resort	72°53’15.69"E	4°4’25.52"N
81	Fihaalhohi Island Resort	73°22’2.28"E	3°52’37.20"N
82	Adaaran Club Rannalhi	73°21’27.29"E	3°54’11.01"N
83	Kurendhoo	73°27’51.93"E	5°20’1.30"N
84	Thulhaadhoo	72°50’25.17"E	5°1’22.40"N
85	Holhudhoo	73°15’45.77"E	5°45’18.66"N
86	Manadhoo	73°24’48.03"E	5°46’0.72"N
87	Hilton Mildives - Iru Fushi Resort and Spa	73°19’25.19"E	5°44’38.60"N
88	Maduvvari	72°53’45.41"E	5°29’10.21"N
89	Foakaidhoo	73°8’54.19"E	6°19’33.30"N
90	Feydhoo	73°2’50.94"E	6°21’35.71"N
100	Huvashen Fushi Resort	73°22’13.60"E	4°22’5.36"N
Middle East and North Africa			
123	Oil platform, Gulf of Suez	33°4’18.40"E	28°51’3.17"N
124	Oil platform, Gulf of Suez	33°7’47.43"E	28°58’10.19"N
162	On-shore installation, Gulf of Suez	33°13’25.38"E	28°43’48.74"N
163	On-shore installation, Gulf of Suez	32°56’2.34"E	29°21’55.78"N
164	Oil installations, Das Island, Persian Gulf	52°52’29.03"E	25°9’8.99"N
165	Arzanah Island, Persian Gulf	52°33’36.86"E	24°47’1.67"N
166	Persian Gulf Island	51°43’35.97"E	24°34’44.23"N
167	Oil camp Algeria	5°45’8.24"E	30°58’40.56"N
168	Oil installations Algeria	5°30’21.23"E	30°46’59.62"N
169	Oil installations Algeria	8°2’14.22"E	30°50’2.46"N
170	Oil installations Algeria	8°15’56.77"E	30°36’38.45"N
171	Oil installations Algeria	7°54’21.70"E	30°23’9.48"N
172	Oil installations Algeria	6°56’16.66"E	31°23’29.41"N
173	Oil installations Algeria	6°47’33.26"E	31°11’29.47"N
174	Oil installations Libya	19°46’11.68"E	28°54’51.19"N
175	Oil installations Libya	18°55’5.03"E	28°34’54.56"N
176	Oil installations Saudi Arabia	49°31’2.96"E	25°40’48.80"N
177	Oil installations Saudi Arabia	49°15’39.05"E	25°38’37.41"N
178	Oil installations Saudi Arabia	49°12’38.14"E	25°31’50.09"N
179	Oil installations Saudi Arabia	49°15’8.00"E	25°25’24.38"N
180	Oil installations Saudi Arabia	49°13’55.39"E	25°21’41.15"N
181	Oil installations Saudi Arabia	49°14’56.63"E	25°15’47.11"N
182	Oil installations Saudi Arabia	49°20’35.69"E	25°18’43.98"N
183	Oil installations Saudi Arabia	49°13’1.56"E	25°8’2.66"N
184	Oil installations Saudi Arabia	49°14’38.90"E	25°52’46.62"N
185	Oil installations Saudi Arabia	49°13’55.12"E	25°58’18.90"N
Mexico, Central America, and Southern U.S.			

1	Oil/gas installation, PetÃĺn, Mexico	90°47'4.03"W	17°31'52.48"N
2	Oil/gas installation, Texas, U.S.	98°4'17.61"W	26°48'40.08"N
3	Oil/gas installation, Veracruz, Mexico	94°12'8.49"W	18°7'43.28"N
4	Oil/gas installation, Guatemala	90°11'56.89"W	16°0'44.25"N
5	Oil/gas installation, Guatemala	90°20'6.95"W	16°7'1.66"N
6	Oil/gas installation, Chiapas, Mexico	93°21'51.60"W	17°28'48.59"N
7	Oil/gas installation, Chiapas, Mexico	93°22'49.91"W	17°33'53.74"N
8	Oil/gas installation, Chiapas, Mexico	93°17'57.34"W	17°35'26.67"N
9	Oil/gas installation, Tabasco, Mexico	93°21'42.70"W	17°39'24.98"N
10	Oil/gas installation, Mexico	93°30'4.14"W	17°51'52.06"N
11	Oil/gas installation, Tabasco, Mexico	92°37'23.48"W	17°53'33.96"N
12	Oil/gas installation, Texas, U.S.	101°59'34.58"W	35°21'9.75"N
13	Mine, Cuba	82°40'53.09"W	22°47'59.80"N
14	Oil/gas installation, Cuba	81°37'16.77"W	23°9'0.41"N
15	Mine, Mexico	107.585W	31.2327N
17	Mine, Mexico	105°48'49.90"W	28°36'17.09"N
33	Airport, Matanzas, Cuba	81°25'45.94W	23°2'28.61N
35	Coastal installation, Campeche, Mexico	92°9'53.61W	18°38'24.38N
36	Chemical plant, Campeche, Mexico	92°15'49.52W	18°36'35.77N
37	El Naranjo airport, Peten, Guatemala	90°48'36.87W	17°13'41.47N
54	Mine, Sonora, Mexico	109°32'56.68"W	30°19'41.02"N
91	Oil/gas installation, Tamaulipas, Mexico	97°58'4.48"W	22°34'40.63"N
92	Plant, Tabasco, Mexico	92°26'38.04"W	17°39'2.80"N
93	Mine, Texas, U.S.	102°19'2.87"W	35°24'31.92"N
94	Oil/gas installation, Oklahoma, U.S.	99°18'3.14"W	35°18'36.22"N
96	Mine, Sonora, Mexico	109°1'12.21"W	27°9'39.86"N
97	Mine, Jalisco, Mexico	103°50'6.61"W	20°35'16.57"N
98	Mine, California, U.S.	117°41'58.83"W	35°2'5.56"N
121	Oil platform Gulf of Mexico	92°45'41.92"W	29°31'53.97"N
122	Oil platform Gulf of Mexico	93°17'2.16"W	29°40'54.91"N
Other sites			
30	Cocos Islands airport	96°49'47.97"E	12°11'20.98"S
31	Ascension Island RAF	14°24'9.84"W	7°58'5.76"S
41	Olympic corrections center, Washington, U.S.	124°8'8.80W	47°42'59.96N
101	Cocos Islands village	96°53'43.54"E	12°7'2.56"S
102	Christmas Island detention center	105°34'31.35"E	10°28'17.80"S
Pacific islands			
24	Tuvalu	179°11'55.98E	8°31'14.49S
25	Mururoa	138°47'3.91W	21°49'36.77S
26	Kwajalein airfield	167°44'21.16E	8°43'40.01N
27	Kwajalein airfield	167°28'29.62E	9°23'45.86N
99	Ebeye island, Kwajalein	167°44'14.30"E	8°46'56.62"N
Southern Africa			
44	Mine, South Africa	27°36'59.51"E	25°41'33.27"S
45	Mine, Eastern Cape, South Africa	22°59'27.90"E	27°23'21.27"S
47	Mine, Phalaborwa, South Africa	31°7'1.30"E	23°59'6.84"S

48	Mine, Eastern Cape, South Africa	22°58'29.78"E	27°51'40.49"S
49	Mine, South Africa	27°35'46.47"E	26°24'29.84"S
50	Mine, South Africa	28°39'46.37"E	24°28'46.73"S
53	Mine, Rossing, Namibia	15°2'52.97"E	22°28'21.57"S
60	Mine, Botswana	25°41'25.58"E	21°30'45.74"S
61	Mine, Botswana	25°22'56.53"E	21°18'39.56"S
105	Oryx mine, South Africa	26°43'18.77"E	28°11'6.70"S
106	Mine, Northern Cape, South Africa	23°26'47.29"E	28°22'47.78"S
107	Mine, South Africa	27°39'40.78"E	26°25'5.86"S
108	Mine, South Africa	27°18'16.35"E	26°28'0.02"S
117	Mine, Botswana	26°3'39.27"E	20°31'30.23"S
118	Mine, Botswana	29°18'59.45"E	22°26'49.74"S
119	Mine, Botswana	24°42'30.44"E	24°32'5.39"S
120	Maparangwane airbase, Botswana	25°19'56.40"E	24°14'28.00"S

Bibliography

Abrahams, A. (2015). Hard traveling: Commuting costs and welfare in the second Palestinian uprising. Working paper.

Abrahams, A., Lozano-Gracia, N., and Oram, C. (2015). Correcting overglow in nighttime lights data. Working paper.

Acemoglu, D., Finkelstein, A., and Notowidigdo, M. J. (2013). Income and health spending: evidence from oil price shocks. *Review of Economics and Statistics*, 95(4):1079–1095.

Agnew, J., Gillespie, T., Gonzalez, J., and Min, B. (2008). Baghdad nights: evaluating the US military 'surge' using nighttime light signatures. *Environment and Planning A*, 40(10):2285–2295.

Ahmed, M., Ahuja, S., Alauddin, M., Hug, S., Lloyd, J., Pfaff, A., Pichler, T., Saltikov, C., Stute, M., and Van Geen, A. (2006). Ensuring safe drinking water in Bangladesh. *Science*, 314(5806):1687.

Ahuja, A., Kremer, M., and Zwane, A. P. (2010). Providing safe water: Evidence from randomized evaluations. *Annu. Rev. Resour. Econ.*, 2(1):237–256.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (Forthcoming). Ethnic inequality. *Journal of Political Economy*.

Alloway, B. J. (2013). Sources of heavy metals and metalloids in soils. In *Heavy Metals in Soils*, pages 11–50. Springer.

Anderson, S. (2011). Caste as an impediment to trade. *American Economic Journal: Applied Economics*, 3(1):239–263.

Aragón, F. M. and Rud, J. P. (2013). Natural resources and local communities: evidence from a Peruvian gold mine. *American Economic Journal: Economic Policy*, 5(2):1–25.

Aragón, F. M. and Rud, J. P. (2015). Polluting industries and agricultural productivity: Evidence from mining in Ghana. *The Economic Journal*.

Arceo, E., Hanna, R., and Oliva, P. (2015). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *The Economic Journal*.

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

Ashraf, N., Berry, J., and Shapiro, J. M. (2007). Can higher prices stimulate product use? Evidence from a field experiment in Zambia. Working Paper 13247, National Bureau of Economic Research.

ATSDR (2007). Toxicological profile for lead. Technical report, US Department of Health and Human Services. Public Health Service, Atlanta, GA, USA.

Azzarri, C., Carletto, G., Davis, B., and Zezza, A. (2006). Monitoring poverty without consumption data: an application using the Albania panel survey. *Eastern European Economics*, 44(1):59–82.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., and Tong, S.-L. (1992). Environmental exposure to lead and children's intelligence at the age of seven years: the Port Pirie Cohort Study. *New England Journal of Medicine*, 327(18):1279–1284.

Balarajan, Y., Ramakrishnan, U., Özaltin, E., Shankar, A. H., and Subramanian, S. (2012). Anaemia in low-income and middle-income countries. *The Lancet*, 378(9809):2123–2135.

Balen, J., McManus, D. P., Li, Y.-S., Zhao, Z.-Y., Yuan, L.-P., Utzinger, J., Williams, G. M., Li, Y., Ren, M.-Y., Liu, Z.-C., et al. (2010). Comparison of two approaches for measuring household wealth via an asset-based index in rural and peri-urban settings of hunan province, china. *Emerging Themes in Epidemiology*, 7(7):1–17.

Bates, M. A., Glennerster, R., Gumede, K., and Duflo, E. (2012). The price is wrong. *Field Actions Science Reports*, (Special Issue 4).

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2012). Roads, railroads and decentralization of Chinese cities. Working paper.

Bellinger, D., Leviton, A., Rabinowitz, M., Allred, E., Needleman, H., and Schoenbaum, S. (1991). Weight gain and maturity in fetuses exposed to low levels of lead. *Environmental Research*, 54(2):151–158.

Bellinger, D. C. (2004). Lead. *Pediatrics*, 113(4 Suppl):1016–1022.

Benin, A. L., Sargent, J. D., Dalton, M., and Roda, S. (1999). High concentrations of heavy metals in neighborhoods near ore smelters in northern Mexico. *Environmental Health Perspectives*, 107(4):279.

Bennear, L., Tarozzi, A., Pfaff, A., Balasubramanya, S., Ahmed, K. M., and Van Geen, A. (2013). Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in Bangladesh. *Journal of Environmental Economics and Management*, 65(2):225–240.

Berkman, D. S., Lescano, A. G., Gilman, R. H., Lopez, S. L., and Black, M. M. (2002). Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. *The Lancet*, 359(9306):564–571.

Berry, J., Fischer, G., and Guiteras, R. Eliciting and utilizing willingness to pay: evidence from field trials in Northern Ghana. Working paper.

Bobonis, G. J., Miguel, E., and Puri-Sharma, C. (2006). Anemia and school participation. *Journal of Human Resources*, 41(4):692–721.

Boyle, K. J., Kuminoff, N. V., Zhang, C., Devanney, M., and Bell, K. P. (2010). Does a property-specific environmental health risk create a 'neighborhood' housing price stigma? Arsenic in private well water. *Water Resources Research*, 46(3).

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.

Carson, R. T., Koundouri, P., and Nauges, C. (2011). Arsenic mitigation in Bangladesh: A household labor market approach. *American Journal of Agricultural Economics*, 93(2):407–414.

CDC (2012). Low level lead exposure harms children: A renewed call of primary prevention. *Atlanta, GA: Centers for Disease Control and Prevention*.

Chakraborti, D., Mukherjee, S. C., Pati, S., Sengupta, M. K., Rahman, M. M., Chowdhury, U. K., Lodh, D., Chanda, C. R., Chakraborti, A. K., and Basu, G. K. (2003). Arsenic groundwater contamination in Middle Ganga Plain, Bihar, India: a future danger? *Environmental Health Perspectives*, 111(9):1194.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Chen, Y., Ebenstein, A., Greenstone, M., and Li, H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences*, 110(32):12936–12941.

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

Cohen, J. and Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, 125(1):1–45.

Consejo Nacional de Población, México (CONAPO) (2000). Índices de marginación a nivel localidad 2000. Technical report.

Corno, L. and De Walque, D. (2012). Mines, migration and HIV/AIDS in Southern Africa. *Journal of African Economies*, 21(3):465–498.

Currie, J., Davis, L., Greenstone, M., and Walker, R. (2013). Do housing prices reflect environmental health risks? Evidence from more than 1600 toxic plant openings and closings. Working Paper 18700, National Bureau of Economic Research.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications.

Dell, M. (2010). The persistent effects of Peru's mining *mita*. *Econometrica*, 78(6):1863–1903.

Dewey, K. G. and Begum, K. (2011). Long-term consequences of stunting in early life. *Maternal and Child Nutrition*, 7(s3):5–18.

Doll, C. N., Muller, J.-P., and Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92.

Dube, O. and Vargas, J. (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *The Review of Economic Studies*, 80(4):1384–1421.

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.

Dupas, P. (2014a). Getting essential health products to their end users: Subsidize, but how much? *Science*, 345(6202):1279–1281.

Dupas, P. (2014b). Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1):197–228.

Ebenstein, A. (2012). The consequences of industrialization: Evidence from water pollution and digestive cancers in China. *Review of Economics and Statistics*, 94(1):186–201.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.

Elvidge, C., Safran, J., Nelson, I., Tuttle, B., Ruth Hobson, V., Baugh, K., Dietz, J., Erwin, E., Lunetta, R., and Lyon, J. (2004). *Area and positional accuracy of DMSP nighttime lights data*. CRC Press: Boca Raton, FL, USA.

Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., Erwin, E. H., and Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3):595–622.

Factor-Litvak, P., Slavkovich, V., Liu, X., Popovac, D., Preteni, E., Capuni-Paracka, S., Hadzialjevic, S., Lekic, V., LoIacono, N., Kline, J., et al. (1998). Hyperproduction of erythropoietin in nonanemic lead-exposed children. *Environmental Health Perspectives*, 106(6):361–364.

Factor-Litvak, P., Wasserman, G., Kline, J. K., and Graziano, J. (1999). The Yugoslavia Prospective Study of environmental lead exposure. *Environmental Health Perspectives*, 107(1):9–15.

Fendorf, S., Michael, H. A., and Van Geen, A. (2010). Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science*, 328(5982):1123–1127.

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.

Filmer, D. and Scott, K. (2012). Assessing asset indices. *Demography*, 49(1):359–392.

Frankel, J. A. (2010). The natural resource curse: a survey. Working Paper 15836, National Bureau of Economic Research.

Gallego, F., Montero, J.-P., and Salas, C. (2013). The effect of transport policies on car use: Evidence from Latin American cities. *Journal of Public Economics*, 107(0):47–62.

Gennaioli, N., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2013). Human capital and regional development. *The Quarterly Journal of Economics*, 128(1):105–164.

George, C. M., Inauen, J., Rahman, S. M., and Zheng, Y. (2013). The effectiveness of educational interventions to enhance the adoption of fee-based arsenic testing in Bangladesh: A cluster randomized controlled trial. *The American Journal of Tropical Medicine and Hygiene*, 89(1):138–144.

George, C. M., van Geen, A., Slavkovich, V., Singha, A., Levy, D., Islam, T., Ahmed, K. M., Moon-Howard, J., Tarozzi, A., Liu, X., et al. (2012). A cluster-based randomized controlled trial promoting community participation in arsenic mitigation efforts in Bangladesh. *Environmental Health*, 11(1):1–10.

Ghosh, T., Sutton, P., Powell, R., Anderson, S., and Elvidge, C. D. (2009). Estimation of Mexico's informal economy and remittances using nighttime imagery. *Remote Sensing*, 1(3):418–444.

Grandjean, P., Jensen, B. M., Sandø, S., Jøgensen, P., and Antonsen, S. (1989). Delayed blood regeneration in lead exposure: an effect on reserve capacity. *American Journal of Public Health*, 79(10):1385–1388.

Greenstone, M. and Hanna, R. (2011). Environmental regulations, air and water pollution, and infant mortality in India. Working Paper 17210, National Bureau of Economic Research.

Greenstone, M. and Jack, B. K. (2013). Envirodevonomics: A research agenda for a young field. Working Paper 19426, National Bureau of Economic Research.

Hanna, R. and Oliva, P. (2011). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. Working Paper 17302, National Bureau of Economic Research.

Hartwick, J. M., Olewiler, N. D., et al. (1986). *The Economics of Natural Resource Use*. Harper & Row.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *The American Economic Review*, 102(2):994–1028.

Hendryx, M. and Ahern, M. M. (2008). Relations between health indicators and residential proximity to coal mining in West Virginia. *American Journal of Public Health*, 98(4):669–671.

Hernandez-Avila, M., Peterson, K. E., Gonzalez-Cossio, T., Sanin, L. H., Aro, A., Schnaas, L., and Hu, H. (2002). Effect of maternal bone lead on length and head circumference of newborns and 1-month-old infants. *Archives of Environmental Health: An International Journal*, 57(5):482–488.

Hoddinott, J., Maluccio, J., Behrman, J. R., Martorell, R., Melgar, P., Quisumbing, A. R., Ramirez-Zea, M., Stein, A. D., and Yount, K. M. (2011). The consequences of early childhood growth failure over the life course. Technical report, International Food Policy Research Institute (IFPRI).

Horton, S. and Ross, J. (2003). The economics of iron deficiency. *Food Policy*, 28(1):51–75.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Infomine (2013). Company and property mining intelligence.

International Monetary Fund (2013). Primary commodity prices.

IntierraRMG (2013). Raw materials data.

Johnson Matthey (2013). Price charts.

Kelly, T. and Matos, G. (2013). Historical statistics for mineral and material commodities in the United States (2013 version).

Kotsadam, A. and Tolonen, A. (2013). Mineral mining and female employment. *Oxcarre Research Papers*, 114.

Kremer, M. and Miguel, E. (2007). The illusion of sustainability. *The Quarterly Journal of Economics*, 122(3):1007–1065.

Landrigan, P. J. and Baker, E. L. (1981). Exposure of children to heavy metals from smelters: epidemiology and toxic consequences. *Environmental Research*, 25(1):204–224.

Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., Canfield, R. L., Dietrich, K. N., Bornschein, R., Greene, T., et al. (2005). Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environmental Health Perspectives*, 113(7):894.

Lauwerys, R., Buchet, J., Roels, H., and Hubermont, G. (1978). Placental transfer of lead, mercury, cadmium, and carbon monoxide in women: I. comparison of the frequency distributions of the biological indices in maternal and umbilical cord blood. *Environmental research*, 15(2):278–289.

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

Mathers, C., Fat, D. M., and Boerma, J. (2008). *The global burden of disease: 2004 update*. World Health Organization.

McCann, J. C. and Ames, B. N. (2007). An overview of evidence for a causal relation between iron deficiency during development and deficits in cognitive or behavioral function. *The American Journal of Clinical Nutrition*, 85(4):931–945.

Meredith, J., Robinson, J., Walker, S., and Wydick, B. (2013). Keeping the doctor away: Experimental evidence on investment in preventative health products. *Journal of Development Economics*, 105:196–210.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, 13(2):151–161.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, 81(1):113–152.

Min, B. and Gaba, K. M. (2014). Tracking electrification in Vietnam using nighttime lights. *Remote Sensing*, 6(10):9511–9529.

Min, B., Gaba, K. M., Sarr, O. F., and Agalassou, A. (2013). Detection of rural electrification in Africa using DMSP-OLS night lights imagery. *International Journal of Remote Sensing*, 34(22):8118–8141.

Mining Atlas (2014). Mining industry data and maps.

Monteiro, J. and Ferraz, C. (2009). Does oil make leaders unaccountable? Mimeo, Pontificia Universidade Catolica do Rio de Janeiro.

Montgomery, M. R., Gragnolati, M., Burke, K. A., and Paredes, E. (2000). Measuring living standards with proxy variables. *Demography*, 37(2):155–174.

Natural Resources Canada (2009). *Canadian Minerals Yearbook 2009*.

Nickson, R., Sengupta, C., Mitra, P., Dave, S., Banerjee, A., Bhattacharya, A., Basu, S., Kakoti, N., Moorthy, N., Wasuja, M., et al. (2007). Current knowledge on the distribution of arsenic in groundwater in five states of India. *Journal of Environmental Science and Health Part A*, 42(12):1707–1718.

Opar, A., Pfaff, A., Seddique, A., Ahmed, K., Graziano, J., and Van Geen, A. (2007). Responses of

6,500 households to arsenic mitigation in Araihazar, Bangladesh. *Health & Place*, 13(1):164–172.

Parvez, F., Wasserman, G. A., Factor-Litvak, P., Liu, X., Slavkovich, V., Siddique, A. B., Sultana, R., Sultana, R., Islam, T., Levy, D., et al. (2011). Arsenic exposure and motor function among children in Bangladesh. *Environmental Health Perspectives*, 119(11):1665.

Pitt, M., Rosenzweig, M. R., and Hassan, N. Identifying the hidden costs of a public health success: arsenic well water contamination and productivity in Bangladesh. Working Paper 21741, National Bureau of Economic Research.

Rau, T., Reyes, L., and Urzúa, S. S. (2013). The long-term effects of early lead exposure: Evidence from a case of environmental negligence. Working Paper 18915, National Bureau of Economic Research.

Ricci, K. A., Girosi, F., Tarr, P. I., Lim, Y.-W., Mason, C., Miller, M., Hughes, J., Von Seidlein, L., Agosti, J. M., and Guerrant, R. L. (2006). Reducing stunting among children: the potential contribution of diagnostics. *Nature*, 444:29–38.

Ripley, E. A., Redmann, R. E., Crowder, A. A., et al. (1996). *Environmental Effects of Mining*. St. Lucie Press Delray Beach, FL.

Roels, H., Bucket, J.-P., Lauwerys, R., Hubermont, G., Bruaux, P., Claeys-ThOreau, F., Lafontaine, A., and Overschelde, J. V. (1976). Impact of air pollution by lead on the heme biosynthetic pathway in school-age children. *Archives of Environmental Health: An International Journal*, 31(6):310–316.

Roper, W. L., Houk, V., Falk, H., and Binder, S. (1991). Preventing lead poisoning in young children: A statement by the Centers for Disease Control, October 1991. Technical report, Centers for Disease Control, Atlanta, GA (United States).

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680–685.

Sahn, D. E. and Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Review of Income and Wealth*, 49(4):463–489.

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., Begum, N., Shah, R., Karyana, M., Kosen, S., et al. (2013). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859):2129–2143.

Salomons, W. (1995). Environmental impact of metals derived from mining activities: processes, predictions, prevention. *Journal of Geochemical Exploration*, 52(1):5–23.

Sanín, L. H., González-Cossío, T., Romieu, I., Peterson, K. E., Ruíz, S., Palazuelos, E., Hernández-Avila, M., and Hu, H. (2001). Effect of maternal lead burden on infant weight and weight gain at one month of age among breastfed infants. *Pediatrics*, 107(5):1016–1023.

Schulz, L. C. (2010). The Dutch Hunger Winter and the developmental origins of health and disease. *Proceedings of the National Academy of Sciences*, 107(39):16757–16758.

Schwartz, J., Landrigan, P. J., Baker Jr, E. L., Orenstein, W. A., and Von Lindern, I. (1990). Lead-induced anemia: dose-response relationships and evidence for a threshold. *American Journal of Public Health*, 80(2):165–168.

Secretaría de Desarollo Social (SEDESOL) (2000). Encuesta de evaluación de los hogares 2000. http://evaluacion.oportunidades.gob.mx:8010/index1.php.

Shampanier, K., Mazar, N., and Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science*, 26(6):742–757.

Shukla, R., Dietrich, K. N., Bornschein, R. L., Berger, O., and Hammond, P. B. (1991). Lead exposure and growth in the early preschool child: a follow-up report from the Cincinnati lead study. *Pediatrics*, 88(5):886–892.

Sloan, N. L., Jordan, E., and Winikoff, B. (2002). Effects of iron supplementation on maternal hematologic status in pregnancy. *American Journal of Public Health*, 92(2):288–293.

Small, C., Elvidge, C. D., Balk, D., and Montgomery, M. (2011). Spatial scaling of stable night lights. *Remote Sensing of Environment*, 115(2):269–280.

Stoltzfus, R., Mullany, L., and Black, R. E. (2004). Iron deficiency anaemia. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, 1:163–209.

Storeygard, A. (Forthcoming). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *Review of Economic Studies*.

Tarozzi, A., Mahajan, A., Blackburn, B., Kopf, D., Krishnan, L., and Yoong, J. (2014). Micro-loans, insecticide-treated bednets, and malaria: Evidence from a randomized controlled trial in Orissa, India. *American Economic Review*, 104(7):1909–41.

UNCTAD (2013). Annual free market commodity prices.

United States Geological Survey (2005). Mineral Resources Data System.

van Geen, A., Ahmed, E. B., Pitcher, L., Mey, J. L., Ahsan, H., Graziano, J. H., and Ahmed, K. M. (2014). Comparison of two blanket surveys of arsenic in tubewells conducted 12 years apart in a 25km^2 area of Bangladesh. *Science of the Total Environment*, 488:484–492.

Van Geen, A., Ahsan, H., Horneman, A. H., Dhar, R. K., Zheng, Y., Hussain, I., Ahmed, K. M., Gelman, A., Stute, M., Simpson, H. J., et al. (2002). Promotion of well-switching to mitigate the current arsenic crisis in Bangladesh. *Bulletin of the World Health Organization*, 80(9):732–737.

van Geen, A., Bravo, C., Gil, V., Sherpa, S., and Jack, D. (2012). Lead exposure from soil in Peruvian mining towns: a national assessment supported by two contrasting examples. *Bulletin of the World Health Organization*, 90(12):878–886.

Wasserman, G., Graziano, J., Factor-Litvak, P., Popovac, D., Morina, N., Musabegovic, A., Vrenezi, N., Capuni-Paracka, S., Lekic, V., Preteni-Redjepi, E., others, et al. (1992). Independent effects of lead exposure and iron deficiency anemia on developmental outcome at age 2 years. *The Journal of Pediatrics*, 121(5):695–703.

Wasserman, G. A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., van Geen, A., Slavkovich, V., Lolacono, N. J., Cheng, Z., Hussain, I., et al. (2004). Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh. *Environmental Health Perspectives*, 112(13):1329–1333.

Wilson, N. (2012). Economic booms and risky sexual behavior: Evidence from Zambian copper mining cities. *Journal of Health Economics*, 31(6):797–812.

World Bank (2013). Global economic monitor.

World Health Organization (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.

Wright, D. A. and Welbourn, P. (2002). *Environmental Toxicology*, volume 11. Cambridge University Press.

Zhu, M., Fitzgerald, E. F., Gelberg, K. H., Lin, S., and Druschel, C. M. (2010). Maternal low-level lead exposure and fetal growth. *Environmental Health Perspectives*, 118(10):1471.