MiningLeaks: Water Pollution and Child Mortality in Africa

Mélanie Gittard [1] and Irène Hu [2] [3]

Draft Version

May 9^{th}, 2022

Abstract

Industrial mining can be a boon or a bane for communities living in the vicinity of production sites. We assess the environmental and health impacts of industrial mining in Africa using micro-data from 1986 to 2018 in 26 countries matched with geocoded data of industrial mining sites. Through a staggered difference-in-difference strategy, we exploit the variation of the opening of a mine and the relative topographic position of surrounding villages. Relying on an upstream-downstream topographic treatment, we show that opening an industrial mining site increases infant mortality, and that this effect is driven by mining-induced water pollution. We find that being downstream when a mine opens increases by 2.3 percentage points the 24-month mortality rate, which corresponds to an increase of 27%. To our knowledge, we are the first to isolate the channel of water pollution by comparing households living upstream and downstream water subbasins from industrial mining sites in Africa and contribute to the literature by exhibiting the heterogeneous effects of industrial mining activity by taking into account the topography and not only the geographic distance of exposure.

Keywords : Africa, Industrial Mining, Health, Pollution, Natural Resources, Water, Environmental Degradation

JEL codes : I1, L72, Q53, Q32, Q33,O12, O13, O15, R11

[1] Paris School of Economics (PSE), Centre International de Recherche sur l'Environnement et le Développement (CIRED) and Ecole Nationale des Ponts et Chaussées (ENPC), France. Contact author's email : melanie.gittard@psemail.eu

[2] Paris School of Economics (PSE) and Université Paris 1 Panthéon-Sorbonne, France. Contact author's email : irene.hu@psemail.eu

[3] We are grateful to our advisors Denis Cogneau, Fabrice Etilé, François Libois and Philippe Quirion for their support and extensive feedbacks on this project. We also would like to thank Liam Wren-Lewis, Thierry Verdier, Oliver Vanden-Eynde, David Margolis, Luc Behagel, Sylvie Lambert, Victoire Girard, Mathieu Couttenier, Dominic Rohner, Francis Cottard, Aurore Stéphant, Jean-Alain Fleurisson, Laurent Polidori, Eléonore Lèbre, Cédric Rebolho and our many colleagues in Paris School of Economics and the CIRED for helpful comments and discussion. More particularly, we are grateful to the participants at the Casual Development Seminar at PSE (CFDS), the *Atelier Theia* and the CSAE Conference 2022. This research received the support of the CEPREMAP, the EHESS, the EUR-PGSE, and the GPET thematic group of PSE. We thank the support of the EUR grand ANR-17-EURE-0001. The usual disclaimer applies. Declaration of conflicts of interest: none.

1 Introduction

The increase in commodity prices since 2000, especially from the extractive sector, has intensified the investments in areas with abundant resources, from hydrocarbons to minerals. The African geology, which is richly endowed containing 30% of the world's mineral reserves [Chuhan-Pole et al., 2017], remains largely unexplored due to the lack of infrastructure and inhospitable terrains [UN Environment Program , 2022, Taylor et al., 2009] . Thus, the continent has been facing a major mining boom since the 2000s, attracting foreign and domestic investments, which raises questions whether natural resources production has improved living standards in resource-rich regions.

The extraction of natural resources can represent a potential boon for local industrial development, as well as a bane because of the generation of negative externalities. On the one side, mining can improve health and local welfare through income effects, by increasing the demand for local labor, goods and services, as well as market channels by attracting or inducing the construction of roads and infrastructure [Chuhan-Pole et al., 2017]. Through a fiscal channel at the national level, an increase of taxation can raise public spendings on local services and improve the access to basic facilities such as electricity, piped water or health services. On the other side, mining can also generate negative externalities such as increasing rapacity and corruption, and creating opportunity cost effects that trigger insecurity and conflicts ([Berman et al., 2017]). Industrial mining also attracts migrant workers that often live in promiscuity and dire straits ([Corno and de Walque, 2012]) or discourage educational attainment ([Atkin, 2016], [Ahlerup et al., 2020]). In addition, production sites can damage the local population's health through the release of pollutants that contaminate the environment. Determining which of these effects is predominant is still debated in the literature studying the relevance of a natural resource curse ([van der Ploeg, 2011], [Cust and Poelhekke, 2015], [Venables, 2016]), and remains an empirical question. We focus on Africa and investigate the local impacts of industrial mining activity on health. Using geocoded micro-data, we manage to isolate the channel of water pollution through which local populations are directly and indirectly affected by the industrial mining sites.

Throughout each stage of a mine's life cycle, its activity can produce and release chemical and mineral pollutants prone to contaminate the surrounding air, water, and soil [Coelho and Texeira, 2011]. Moreover, the ore extraction processes are water-demanding and need access to a water source that very often competes with the local demand, which is all the more alarming in water-stressed areas. Mining activity mostly consists in extracting small concentrations of minerals from huge volumes of rocks and therefore creates a lot of waste, which leaking is hard to avoid. For industrial mines, these wastes are diluted into water and then stored in retention ponds, where they can leak within the local environment and contaminate soil and water bodies. Mining waste actively pollutes during the whole mine's life cycle, starting from its exploration before the opening, during the production, and also at the closure when the mine can be left without maintenance. In the latter case, retention ponds can be left without coverage, dry up, and be scattered directly in the surrounding air, water, and soil. Our study is restricted to industrial mining because of data limitations, but artisanal and small-scale mining (ASM) can also be responsible for pollution. ASM is associated with a different type of pollution, which is accused of being more severe than the industrial sector, because of the use of mercury and lack of monitoring. While mercury has been officially banned in over 140 countries (Minamata Convention on Mercury, adopted in 2013), industrial mining can abide by this law but many use cyanide instead. Both chemicals being highly toxic pollutants, focusing on industrial mining only is a lower-bound analysis of the impacts of mining activity on local populations' health. Besides, industrial mining differs widely from ASM in terms of production quantities and the size of mining sites: industrial mines are responsible for 80% of the gold production and 75% of the diamond production [Del, 2020]. The industrial sector accounts for the majority of the African production and heavy metals released into the environment. If ASM has severe effects on miners' health due to hazardous working conditions characterized by the lack of protection devices, the rare usage of masks, the lack of preventive infrastructure, and the intense usage and exposure to mercury. Industrial mining has broader and larger effects on surrounding communities and especially on the health of individuals living in their vicinity, mainly via the leaking of heavy metals within the local environment.

The main toxic metals released by mining sites are arsenic, cadmium, copper, lead, mercury, and nickel. If low concentration levels of heavy metals can be essential for human health, the abnormal quantities found in the environment within the mine's vicinity can cause several health problems. Individuals living nearby industrial mining are exposed to high concentrations of heavy metals through ingestion, dermal contact, and inhalation of soil particles. In this paper, we mainly focus on the absorption mechanism as we identify the effects of mining activity through water pollution. High blood metal concentrations are associated with neurological effects [Dike et al., 2020], cardiovascular effects, gastrointestinal hemorrhages [Obasi et al., 2020], organ dysfunction [Briffa et al., 2020], higher probability of cancer development [Madilonga et al., 2021, Obasi et al., 2020], but also higher probability of infertility, miscarriages for women, and malformation of newborns [Briffa et al., 2020]. Thus, exposure to heavy metals plays detrimental effects on human health in general and child health in particular, especially during their first months of development, both in and ex-utero [Coelho and Texeira, 2011]. Children are the most sensitive, even at low concentration, as they are at a stage of rapid biological development [Dike et al., 2020], but also as they are more exposed, through higher blood concentration linked to incidental ingestion of urban soil and dirty water [He et al., 2020].

In this paper, we focus on under 12 months and 24 months mortality as a primary health outcome, as effects on children are the most dramatic, and to capture the effects of heavy metal absorption on early-age biological development. Besides, child mortality is a short-term measure [Greenstone and Hanna, 2014, Do et al., 2018], and is available over a long-time span of four decades and across the majority of African countries. This paper focuses on water pollution and heavy metal ingestion and absorption mechanisms. Human exposure to heavy metals through the consumption of contaminated water is of prior concern in Africa and in Sub-Saharan Africa in particular, where only 24% of the population have access to safe drinking water [Programme, 2019]. We match socio-economic and health data from the Demographic Health Surveys (DHS) with state-of-the-art geolocalized data on industrial mineral resource exploitation from the SNL Metals and Mining database, which provides information on opening dates and mineral types. Our study spans 26 out of the 54 African countries, from 1986 to 2018. Our first contribution lies in the construction of the industrial mines dataset, as we checked over 1,700 mines by hand to complete their opening date [4]. The main research question of this paper is to estimate the impacts of industrial mining activity on local populations' health, mainly in terms of child mortality, to assess their heterogeneity, and to give indirect evidence of the channel of water pollution. We conduct a staggered Difference-in-Difference strategy exploiting the variation of the opening of a mine and the relative topographic position of surrounding villages. We indirectly isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison, which is used as a proxy for the exposure to water pollution linked to mining activity.

Our main result shows that being born in a village located downstream of a mine that just opened increases the mortality rates under 24 months by 2.3 percentage points, which corresponds to an increase by 27% of the mortality rate. We find no significant result for the under 12-month mortality rate, which suggests a lag in the effect of water pollution on early-childhood health and in the absorption of toxic elements. The analysis of the dynamic effects identifies an impact of the development phase, which starts a couple of years before a mine starts its production [Benshaul-Tolonen, 2018].

The major contributions of this paper are twofold. First, we propose an extension of an industrial mine dataset extensively used in the literature, as we did archival handwork to complete the opening dates of 1,700 African mines of the SNL dataset. Thanks to this work, we alleviate the issue linked to small samples and low statistical power and can achieve a heterogeneity analysis and identify a balanced panel. Second, our results using large-scale and systematic micro data embody an important contribution to the literature, as it enters into the debate on the positive and negative effects of industrial mining activity on health. It contradicts the literature that uses geographical distance to a mine as a proxy for exposure to mining activity, which finds a reduction in mortality rates [Benshaul-Tolonen, 2018, Cossa et al., 2022].

[4] Opening dates indicate when production first began. Data available in the SNL database was gathered by SNL from the mining companies' reports, and the hand-check work we made has completed this database by going deeper into archival mining reports

In the Appendix, we propose an extension of the papers using a geographic treatment (Appendix A.4, comparing the health outcomes of individuals living nearby to those living further away from the mining site. We replicate in Appendix 21 the exact analysis undertaken by [Benshaul-Tolonen, 2018] which uses fewer observations, countries, and years (41 mines in 9 countries over 1987-2012). If we manage to find the same results (i.e. a reduction of the under 12-month mortality rate) using this proximity treatment on this restricted sample, we do not find that these results are stable to our extended sample which shows the limitation in terms of external validity of these findings. In contrast, we conduct several robustness checks to assess the stability of our main results and show that the increase in child mortality due to mining-induced water pollution goes through all our tests. We argue that using a geographic treatment based on proximity to mining sites encompasses contradictory effects that happen in mines' vicinity, and average compensated effects (both positive and negative externalities) using this empirical strategy explains the absence of results. Our paper shows the importance of looking at the topographic heterogeneity of the effects of mining activity on health to identify the negative effects induced by water pollution, and how they outweight the positive effects.

The remainder of the paper is organized as follows. Section 2 presents the context in light of the literature. Section 3 describes the context and the data. Section 4 details the methodology and the main empirical strategy. Section 5 introduces the results, including the average, dynamic and heterogeneous effects. Section 6 proposes a list of robustness checks and sensitivity analyses to prove the robustness of our main results, while Section 7 gives further discussion. Section 8 concludes.

2 Literature review

The literature on the local effects of mining on local communities has been recently growing during the past decade, yet the debate remains on the costs and benefits, the positive and negative impacts of industrial mining activity in developing countries. Diverse results have been found on the effects on health, and there is still uncertainty on the direction and the magnitude of the impacts of mines on the local population's health. Besides, if geographical proximity to a mining site is usually used as a proxy for pollution exposure, few papers directly observe the negative externalities on the environment and its consequences on health.

A growing literature has focused on the effects of industrial mining activity on local populations' welfare in developing countries, and the results are still under debate. Mining has been shown to have negative effects on the environment and agricultural productivity. [Aragón and Rud, 2016] find that the expansion of large-scale gold mining in Ghana (1997-2005) is responsible for agricultural total factor productivity decrease in the vicinity of mines, and they argue that pollution is the most likely explanation. They use satellite imagery of NO_2 concentration as a proxy for air pollution but only through a cross-sectional analysis. They also compare households located upstream and downstream of an active mine and find no significant difference, but take the result with caution due to a lack of statistical power. At a broader scale of analysis, [Mamo et al., 2019] look at the effects of the discoveries of industrial mining deposits on living standards in Sub-Saharan Africa, and find an increase of district level night lights emissions but no significant effects on household wealth measured through the dimensions of access to electricity, wealth index, urbanization, mortality and education. They find only temporary positive effects on public services provisions which fits with the idea of some non-durable initial concessions made by mining companies or the state to the locals. More specifically, they show a degradation of the sewerage system and piped water supply as mining activity is water intensive. Finally, they find that mines in Sub-Saharan Africa exhibit enclave characteristics with little spillovers to neighbouring districts. We therefore refine their analysis by looking at water pollution at the surface and underground, to capture negative spillovers to districts sharing the same surface water or groundwater. Last but not least, [Bazillier and Girard, 2020] compare the local spillovers between artisanal and industrial mining sites in Burkina-Faso. They find positive impacts of artisanal mining (labor intensive and managed in common) and an absence of effects of opening industrial mines (capital intensive and privatized) on household consumption.

A growing literature has shown the importance of environmental policy to regulate air and water pollution and has studied its impact on health and infant mortality in developing countries ([Jayachandran, 2009], [Greenstone and Hanna, 2014] and [Do et al., 2018]). Industrial activity can have harmful externalities on pollution and health ([Ebenstein, 2012], [Chen et al., 2013], [Brainerd and Menon, 2014] and [He and Perloff, 2016]), yet few papers have managed to show to what extent industrial mining activity creates negative externalities on the environment. [Balietti et al., 2018] look at the effects of mining industries on deforestation in India, [Von der Goltz and Barnwal, 2019] have suggested the mechanism of water pollution but without strong empirical evidence. Yet, in-situ measurements have shown the contamination of water drinking sources by harmful levels of nitrate, turbidity, iron, cadmium, manganese, and arsenic by industrial mining sites ([Cobbina et al., 2013]). To our knowledge, we are the first to provide indirect systematic and large-scale evidence of the mechanism of water pollution by industrial mining activity.

Few papers have dealt with upstream and downstream at the scale of a continent, since it requires high computational capacity and a complex matching methodology. [Duflo and Pande, 2007] study the productivity and distributional effects of large irrigation dams in India and use river networks and calculate gradients computed from digital elevation maps for India. [Do et al., 2018] use river networks and pollution monitoring stations data in India to conduct their upstream-downstream analysis. Unfortunately, it is impossible in our case study due to the absence of water quality data at the scale of Africa. [Garg et al., 2018] use river networks in Indonesia and re-calculate the upstream-downstream relationship between village pairs using a 30m resolution Digital Elevation Model. Very refined level, but not likely to be undertaken at the scale of the African continent in our study, so we choose secondary data computed by hydrologists (HydroSheds). We use systematic and highly disaggregated data on water sub-basins that enable to encompass a wider set of countries, overcoming the issue of matching a mine or a village to the closest river, since there is uncertainty about whether this point is located above or below in altitute compared to the level of the river. [Strobl and Strobl, 2011] studied the distributional effects of large dams on agricultural productivity at the scale of the African continent, using Pfafstetter level 6 with an average area of 4200 km^2. Our study takes into account subbasins at a finer level (Pfafstetter level 12) with an average area of 100 km^2.

Our approach which consists in studying the effects of industrial mines on health is mainly based on two papers. [Benshaul-Tolonen, 2018] finds that large-scale gold mining in nine countries of Sub-Saharan Africa (Burkina Faso, Ivory Coast, DRC, Ghana, Guinea, Ethiopia, Mali, Senegal and Tanzania) between 1987 and 2012, decreases infant mortality within 10 km during the opening and operating phases. However, she finds no effect on farther communities (10-100km). Our project encompasses other minerals than gold, a broader set of countries over a longer time window, but above all deals with endogeneity issues that remain in her strategy (discussed in Section 4). [Von der Goltz and Barnwal, 2019] assess the effects of industrial mines in 44 developing countries from 1988 to 2012 and find gains in asset wealth, increased anemia among women and stunting in young children living within 5 km. As anemia and growth deficits are argued to be mainly the consequences of exposure to lead, the observed effects on health are interpreted to be the results of pollution due to metal contamination and lead toxicity. They find that women in mining communities show depressed blood hemoglobin, recover more slowly from blood loss during pregnancy and delivery, and that children in mining communities suffer some important adverse growth outcome from *in utero* exposure (stunting). Yet, their empirical strategy is also to be improved for reasons extensively discussed in Section 4. A replication analysis of [Benshaul-Tolonen, 2018] in Appendix 21 finds no significant results when applied to our more comprehensive sample. [Cossa et al., 2022] also used [Benshaul-Tolonen, 2018] methodology, and studied a broader set of countries. They find comparable results of decrease in child mortality. This recent work emphasizes the crucialty of shedding light on the actual health impacts of industrial mining.

[Romero and Saavedra,] (unpublished) find negative effects of industrial mining on newborns in Columbia. In particular, they find low Appearance, Pulse, Grimace, Activity and Respiration (APGAR) score only among individuals living downstream of mines. They apply an upstream-downstream analysis using river segmentsand put forward the mechanism of fish consumption as a vector of intoxication and health hazards. They match the closest river to each mine, and calculate the total area of the active mine through the mine's title information. For each municipality they aggregate the total area upstream up to 25 km that is mined. Our study combines a larger set of countries and a longer time span, and therefore claims stronger statistical power and external validity.

3 Data and Context

3.1 Data

In this paper, we match socio-economic data from the Demographic Health Surveys to an industrial mining database provided by SNL Mining and Metals.

3.1.1 Health and socio-economic data

We use all available survey rounds from the Demographic Health Surveys that contain GPS coordinates, from 1986 to 2018, covering 36 out of the 54 African countries. We then select all the countries which have at least two survey waves to be able to implement our Difference-in-Difference strategy with a sufficient time length before and after the opening of a mine [5] and end up with 26 countries, 8,087 clusters and 199,934 children under the age of five [6]. Table 13 in the Appendix displays the DHS survey years and countries that we use for the analysis. We construct the variables of child mortality based on the DHS child recode database which has information on the age and death of children under five years old, whose mothers are aged between 15 and 49 years old. Our dependent variable is the probability of under 12 months and under 24 months mortality for each DHS cluster (i.e. for each child, we build a dummy equal to 1 if he/she is alive and 0 if not, conditional on having reached 12 and 24 months respectively). We will further estimate the effects of mining activity on biomarker variables (weight, height, anemia), and other indicators of occurrences of diarrhea, fever, and cough within two weeks preceding the day of the interview among young children. As we intend to observe the impacts of mining activity on the most vulnerable population, we will extend our analysis to pregnant women and check for indicators such as complications during pregnancy, and *in-utero* exposure, well as fertility rate [WORK IN PROGRESS]. Finally, as the aim of this article is to isolate the mechanism of water pollution, we use the questions from the DHS on the main source of drinking water and the access to health facilities to control for households' sanitary environment. As our empirical strategy relies on a staggered Difference-in-Difference, we emphasize that the countries driving our results are those with at least two DHS waves to capture the variations in exposure to mining activity. Thus, we drop in our analysis all the countries that have only one wave, and our final sample contains 26 countries (cf Tables 12 and 13 to see the dropped countries [7]).

3.1.2 Mineral resource exploitation data

The industrial mining variables come from the SNL Metals and Mining database, which is privately owned by *S&PGlobal* and on license [8]. The SNL database is the best existing panel of mine production, providing information on the location, the dates of opening and closure, the commodity type,and the yearly production (for some mines). This is a non-exhaustive panel of industrial mines in Africa, yet to our knowledge, constitutes the most complete sample of mines giving the timing of the industrial activity. This dataset has been intensively used in the literature [Aragón and Rud, 2016, Berman et al., 2017, Kotsadam and Tolonen, 2016, Benshaul-Tolonen, 2018, Von der Goltz and Barnwal, 2019, Mamo et al., 2019] and argued to be the best product available. We emphasize here that this paper focuses on the effects of industrial mining, and that we do not include artisanal mining (ASM) that are not available in the SNL database.

Overall, the SNL database gathers 3,815 industrial mines in Africa from 1981 to 2021, and 2,016 were located within 100 km of a DHS cluster from a country with at least two surveys. For our difference-in-difference strategy, we need information on the timing of the beginning (and closing [WORK IN PROGRESS]) of the mining production. The SNL database gives this information for 278 mines and

[5] Indeed, we consider that doing a Difference-in-Difference strategy on the sample of countries which have only one round of survey, hence a maximum of five years period, will not enable to capture the longer-term effects of mining activity.

[6] Please note that our final sample does not include Egyp which has 7 DHS waves and is a well-known mining country. This is explained by the fact that the SNL database characterized Egypt within the Middle East rather than in Africa and thus was dropped from our sample. [WORK IN PROGRESS] to include Egypt within the analysis.

[7] We drop Angola, Central African Republic, Gabon, Comores, Morocco, Mozambique, Swaziland, Chad, South Africa

[8] We are grateful to CEPREMAP, PjSE, EHESS, and the GPET thematic group of PSE, for their financial support and their help in purchasing the access to the data.

we retrieved from a hand-work the start up year for the 1,738 remaining mines. The hand-check was realized using archival information on the mining history available in the SNL database (cross-checked with Google maps and aerial images) and we describe this handwork more extensively in the Appendix A.1.2.

We build three main variables from the SNL Mining and Metals database, relying on the geocoded information and the time of opening. According to the estimation strategy, we will use a variable of proximity (distance to the closest mine), position (whether individual i is upstream or downstream), and a dummy for being open or not. Opening dates that were available in the SNL database were computed by the SNL team, and indicate the actual start-up year of the mine, i.e when production first began. We used the same criteria for our handwork that consisted of reading mining sites' archival reports [9]. For the moment, we do not use the closing year variable, as it was harder to retrieve from reports reading.

3.1.3 Water basins

We use the HydroBASINS subbasins geographic information provided by HydroSHEDS, which subdivides subbasins into multiple tributary basins and shows the network of nested subbasins at different scales. Following the topological concept of the Pfafstetter coding system, each polygon of subbasin has information on the up- and down-stream connectivity. We take the finest Pfafstetter level (12 out of 12) that breaks down subbasins at an average area of 100 km^2. See Figure 6a for an example. [10]

3.1.4 Administrative district boundaries

We extract district-level data (sub-national level 2) from GADM database of Global Administrative Areas[11] and match the corresponding polygon shapefiles with DHS clusters. As there is a random reshuffling of DHS villages (from 2 to 10 km) [12], we are well aware that there is some random error in the assignment of districts to observed individuals, and that some may have been assigned to a neighboring district.

3.2 Context

3.2.1 Mining in Africa

Temporal and spatial variation

Africa has a long history of mining, dating back to the period of the Pharaohs in Egypt. The African continent is richly endowed with mineral reserves, containing around 30% of the world's mineral reserves, and is estimated to have the largest reserves for strategically important metals such as cobalt, diamonds, and uranium [UN Environment Program , 2022]. However, as the continent is made of inhospitable terrains and lacks infrastructure, the continent's geology remains largely unexplored [Taylor et al., 2009]. Despite its rich geology, Africa's production accounts for around 8 % of global mineral production (in 2012) [Africa Bank, 2022], and represents an opportunity for mining investors. Thus, Africa is facing a mining boom since the 2000s, attracting foreign investment mainly from China, Canada, Australia, Brazil and Russia, which raises concerns about environmental degradation on the

[9] Please note that for the reports where the 'start of production' during a specific year was not explicitely written, we reported a year if the information suggested a start of production, and completed with satellite images. This might introduce some noise in the variable.

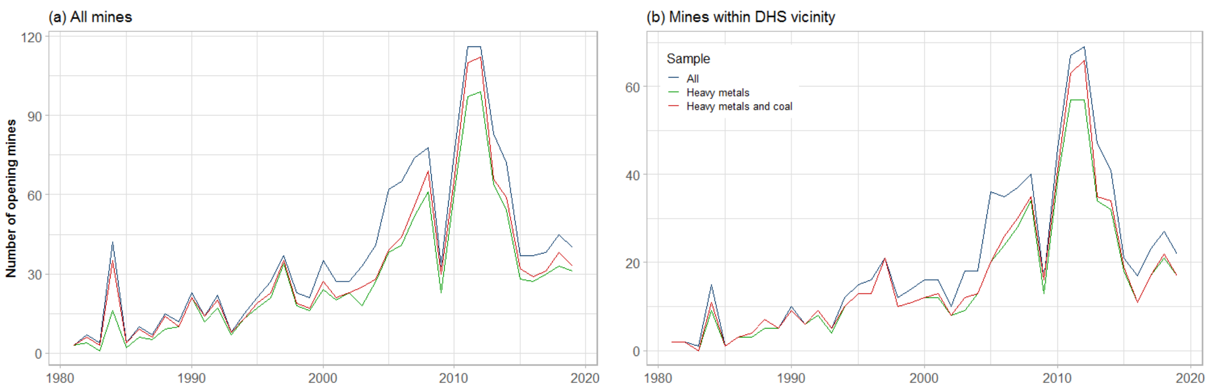
[10] We conduct our analysis taking into consideration the three closest subbasins to each industrial mine, meaning that we take each mine's subbasin A and tag the one just downstream that we call B, the one just downstream of B that we call C, and then the one just downstream of C that we call D. Thus, B,C, and D are the three closest subbasins of A.

[11] Accessed on the website: gadm.org.

[12] "In DHS household surveys the GPS coordinate displacement process is carried out as follows: urban clusters are displaced a distance up to two kilometers (0-2 km) and rural clusters are displaced a distance up to five kilometers (0-5 km), with a further, randomly-selected 1% (every 100th) of rural clusters displaced a distance up to 10 kilometers (0-10 km)", as mentionned in [USAID, 2013]

continent [Taylor et al., 2009, Edwards et al., 2013].

Figure 1: Temporal evolution of mine opening



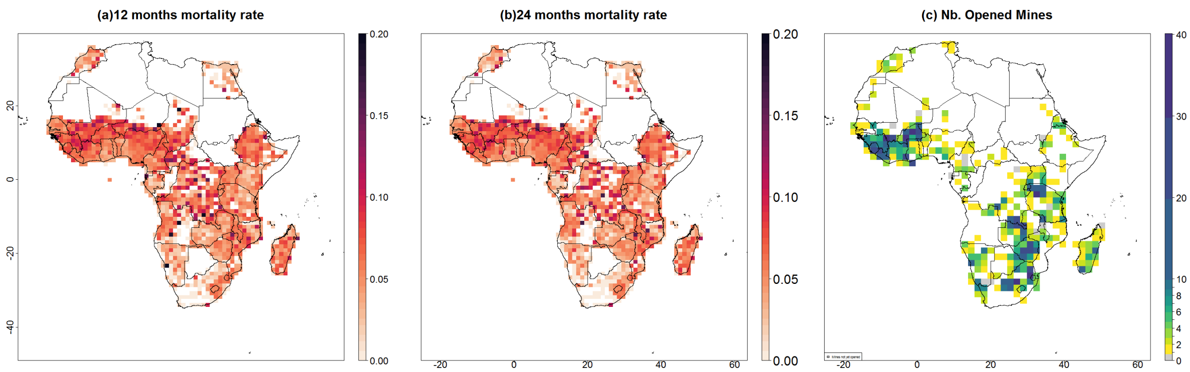
Notes : The Figures plot the number of mines opening each year over the 1981-2019 period, for all mines, heavy metal mines including coal (sample of the main analysis), and only heavy metal mines. Figure (a) displays the temporal evolution of the total mine sample, while Figure (b) of mines that are within the sample of the main analysis, meaning mines that have DHS clusters upstream at 100km at least and DHS clusters downstream within the three closest sub-basins.
Sources : authors' elaboration on DHS and SNL data.

Figure 1 shows the evolution of the yearly number of mines that opened over the whole continent of Africa over the 1981-2019 period, Figure 1 (a) for the entire mining sample while Figure 1 (b) for the mines that are in the sample of the main analysis, meaning mines that have DHS clusters upstream within 100km and DHS clusters downstream within the three closest sub-basins. The mining boom since 2000 is captured in the Figures, with the first peak in 2007, in line with the peak in exploration activity that occurred in 2003 [Taylor et al., 2009] (as the exploration phase is on average a couple of years before a mine opens), and the second one in 2012. What is striking in Figure 1 is that the evolution of mine openings follows the exact same pattern as the evolution of industrial metal prices. The mining boom since 2000 follows the increase in real prices of Copper, Tin, Lead, Aluminim, Zinc, Nickel and other heavy metals, while the sharp fall around 2008/2009 corresponds to the financial crisis. Again, the local minimum around 2016 corresponds to the drop in commodity prices in June 2014 [Khan et al., 2016, Glöser et al., 2017]. This similar evolution suggest that heavy metal prices are good Instrument Variables for the variable year of mine opening, such as [Berman et al., 2017, Bazillier and Girard, 2020] use in their analysis.

In Africa, around 120 industrial mines opened in 2012, based on the non-exhaustive SNL database. The Figures also distinguish the evolution according to the mines' characteristics: it distinguishes the pattern for all mines, heavy metal mines, and heavy metals including coal mines (mines that are in the main regression). The list and chemical characteristics of heavy metal mines are displayed in Table 19. We observe no differences in timing patterns between Figure 1 (a) and (b), neither between mine types.

Figure 2 (c) shows the map of the number of mines that have opened before 2019, including mines that opened before 1986, averaged at the cell level (160km cells). Cells in grey represent areas where no mine opened before 2019, but where at least one will open in the future (whether we know from the data that it has opened between 2019-2021, or if the opening is planned further). The main mining countries in the SNL database are Guinea, Sierra Leone, Ivory Coast, Ghana, Niger, Burkina Faso, Zimbabwe, Tanzania, Zambia, and the north of South Africa. Please note that, as we exclude countries with only one DHS wave in our main analysis' sample (cf Tables 12 and 13), to avoid comparing areas with too many differences in terms of temporal variations, we did not undertake the hand work for these countries, which explains why South Africa (which is not in the final sample) does not appear as a major mining

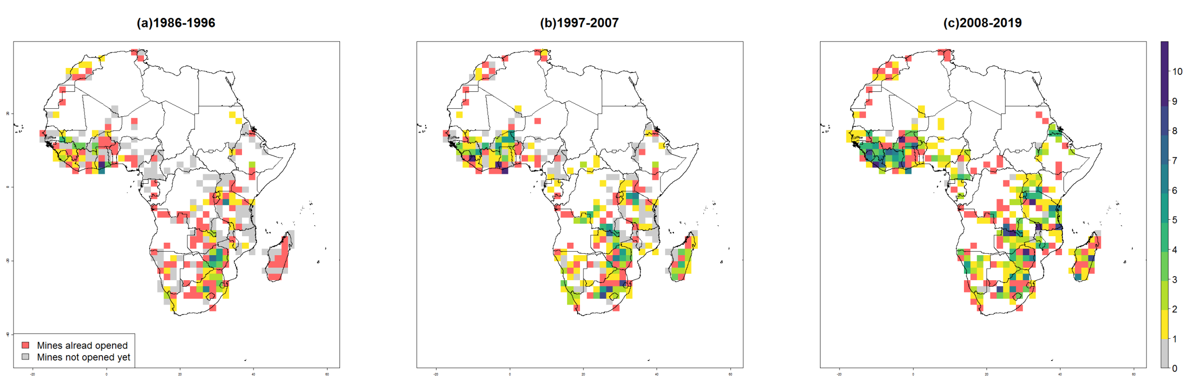
Figure 2: Outcomes spatial distribution



Notes : Figures (a) and (b) represent the means of 12- and 24-month mortality rates for each DHS waves available (listed in table 13), from 1986 to 2019. Means are computed at the grid level (100km mean size). The mortality rates are estimated without the children that did not reach 12/24 months at the time of the survey. Figure (c) displays the stock of mines that have opened before 2019 (including mines that opened before 1986). Means are computed at the grid level (100km mean size).
Sources : authors' elaboration on DHS and SNL data.

country in Figure 2 (c). Figure 3 shows both the temporal and spatial variation of mine opening in Africa (for all the mines sample, and not the restricted one for our main analysis), as it plots the number of mines that opened over different periods of our analysis per grid cell. The cells in red are areas where no mines opened during the period, but where at least one mine has opened before, whereas cells in grey are areas where no mines have ever opened while at least one will open in the future. We observe that the increase in mine opening was higher during the third period 2008-2019 (which is in coherence with Figure 1), and was particularly important in West Africa.

Figure 3: Spatial variation of mine opening per periods



Notes : The figures represent the number of mines that opened during the periods over the grid area (160km on average). A red grid represents an area where no mine opened over the period, but where at least one mine has opened before the period. A grey cell represents an area where no mine opened over the period, but where at least one mine will open in the future.
Sources : authors' elaboration on SNL data.

Mine's life cycle

Throughout each stage of a mine's life cycle, its activity can produce and release chemical and mineral pollutants prone to contaminate the surrounding air, water and soil [Coelho and Texeira, 2011]. During the exploration and prospecting stage that can last several years before a mine is considered economically viable and worthwhile to open, mining companies conduct mapping and sampling, as well as drilling, boreholes and excavation that require both physical and chemical measurements methods likely to pollute at the surface and underground, depending on the nature of the deposit in the targeted area. If found financially viable, the company launches the discovery phase where the design and planning of the construction is undertaken, and the feasibility of the project which requires further exploration and engineering studies. Subsequently, the development stage takes place and the mine's infrastructures and processing facilities are constructed. It is only after all these stages that production can start. Once the deposit is exhausted comes the closure and reclamation stage, where the company is supposed to clean, stabilize and rehabilitate the land and isolate contaminated material. Yet, it is common that waste, tailings or retention dams are just left abandoned without care and maintenance, and this constitutes a potential disaster if the hazardous materials are leaked and discharged into the environment. Figure 20 in Appendix proposes a scheme to explain the life cycle of a mine. Figure 19 displays satellite images of the different stages of a rutile mine in Sierra Leone.

Throughout all these stages, different types of pollution can be engendered: air pollutants which can be carried by dust over long distances by the wind, can damage surrounding soils and crops and be inhaled by the local population. The leakage of pollutants in the air can also affect water through acid mine drainage that ends up polluting the surface and then groundwater. During the digging and processing in order to extract the targeted ore from waste rocks, rocks are crushed and then go through either heap leaching, froth flotation or smelting. These techniques require the addition of chemicals such as cyanide or acid, that are able to separate the targeted minerals from waste. Moreover, these processes are water-demanding and need access to a water source that very often competes with the local demand. Last but not least, even without the use of these chemicals, leaching happens through the contact of water and oxygen with sulfide minerals contained in the extracted rocks [13] accelerates the acidification process and modifies the pH levels of water bodies. Pollutants can be released into the environment during the process by spills or after by leaks of humid waste stored in retention dams but also through the erosion and the sedimentation of solid waste that are piled in the tailings around the mining site and that drain to the soil with rain. The wastes actively pollute during the whole life cycle of the mine, starting from its opening and during the production, but also can continue to pollute when a mine closes and is left without maintenance (when retention ponds are not covered and dry, letting these wastes go directly through the local environment).

3.2.2 Health risks

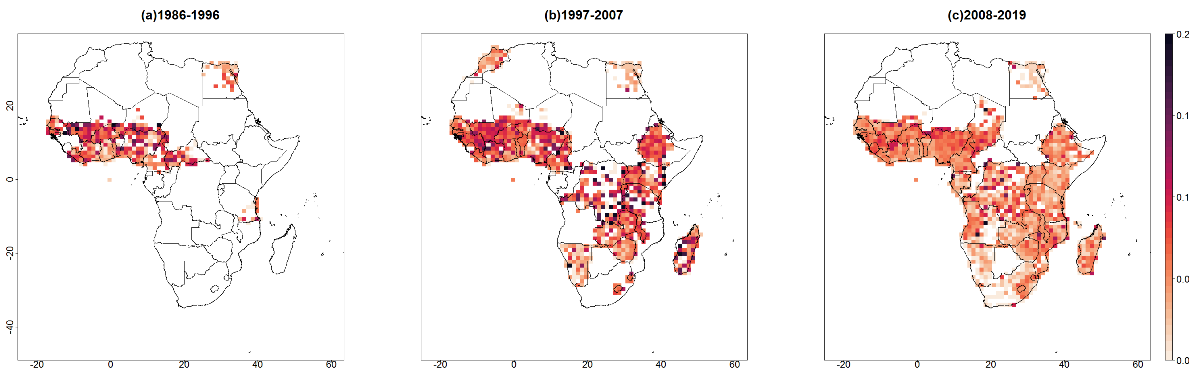
Africa faces high infantile mortality rates, as the average 12 months mortality rate is 6.5 % and the average 24 months mortality rate is 8.5% according to DHS data (cf Table 1). Figures 2 (a) and (b) plot the average mortality rates for all DHS from 1986-2019 averaged at the grid level, and show the spatial variation of mortality rates [14]. Figures 4 and 5 map both spatial and temporal variation of mortality rates as it shows the average mortality rates for the three main periods of our DHS sample. We can observe the global reduction of mortality over the period and also the DHS cluster distribution.

The main toxic metals released by mining sites are arsenic, cadmium, copper, lead, mercury, and nickel. Depending on their blood level concentration, they can be essential or non-essential for human health [El-Kady and Abdel-Wahhab, 2018]. However, heavy metals released by mining activity are non-biodegradable, have long-term impacts on the environment, and are found at abnormally high concentrations in the vicinity of mines, within the soil, water resources, vegetation, and crops [Oje et al., 2010,

[13] such as arsenic, cobalt, copper, cadmium, lead, silver, zinc

[14] Please note that the higher the DHS cluster density, the more accurate the average. The spatial variation is endogeneous to the DHS sample.

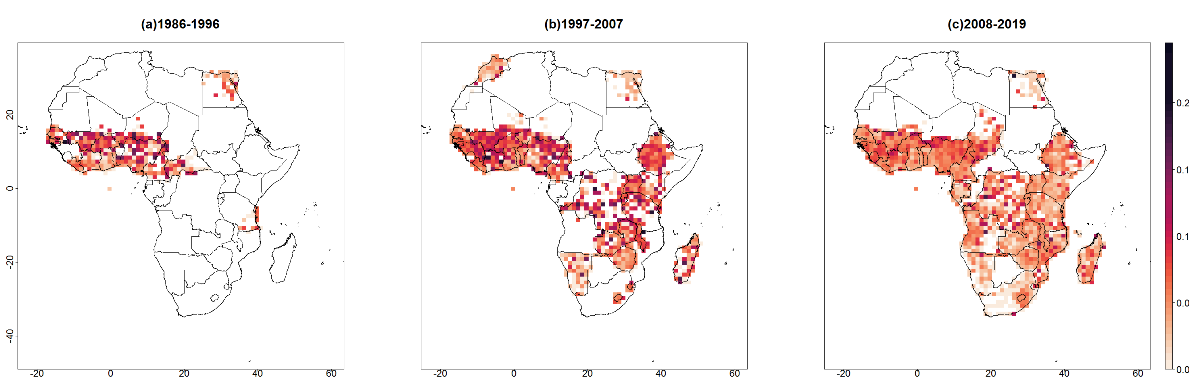
Figure 4: Spatial variation of 12-month mortality rates per period



Notes : The figures represent the means of 12-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008 and (c) 2008-2019. The mortality rates are estimated without the children that did not reach 12 months at the time of the survey.
Sources : authors' elaboration on DHS data.

Dike et al., 2020]. People living in that environment are exposed to high quantities of heavy metals through ingestion, dermal contact, and inhalation of soil particles, which can cause several implications for their health. Heavy metals can be absorbed through the gastrointestinal tract as well as through the respiratory tract. In this paper, as we focus on water pollution, we will mainly identify the health consequences of absorption of heavy metals rather than inhalation. High blood metal concentrations are associated with neurological effects (which induce behavioral problems, learning deficits, memory losses, especially among children) [Dike et al., 2020], neurodegenerative diseases, cardiovascular effects, gastrointestinal hemorrhages [Obasi et al., 2020], organ dysfunction (kidney, decrease the production of red and white blood cells, lung irritation) [Briffa et al., 2020], higher probability of cancer development [Madilonga et al., 2021, Obasi et al., 2020], but also a higher probability of infertility, miscarriages for women, and malformation of newborns [Briffa et al., 2020]. Thus, exposure to heavy metals plays detrimental effects on human health in general and child health in particular, especially during their first months of development, both in and ex-utero [Coelho and Texeira, 2011]. Besides, children in early age are the most sensitive, even at low concentration, as they are at a stage of rapid biological development

Figure 5: Spatial variation of 12-month mortality rates per period



Notes : The figures represent the means of 24-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008 and (c) 2008-2019. The mortality rates are estimated without the children that did not reach 24 months at the time of the survey.
Sources : authors' elaboration on DHS data.

[Dike et al., 2020], but also as they are more exposed, through higher blood concentration linked to incidental ingestion of urban soil and dirty water (less conscious of their environment and danger, playing with polluted soil/eat drink without carefulness)[He et al., 2020].

4 Empirical strategy

The main empirical strategy of this paper uses the relative topographical position of subbasins as a proxy for exposure to mining activity pollution. It compares the effects on the health of individuals living downstream to those living upstream of a mine, before and after the opening of at least one site. It is a staggered design difference-in-difference analysis with two-way fixed effects at the mine's subbasin and birth year level. This upstream-downstream strategy intends to identify the mechanism of water pollution, as well as to correct for the bias linked to geographic treatments commonly applied in the literature, that proxies exposure with proximity to an industrial site, rather than a topographic position. It resolves the challenges of matching DHS clusters to mining sites, as geographic treatments have unbalanced samples due to an endogenous matching. Second, it breaks the average effects based on distance buffers and highlights the heterogeneity of the effects of mining activity on health, and isolates the negative externalities linked to water degradation. As one of the contributions of this paper is the completion of the SNL dataset with hand work on the timing of the opening of mines, we propose in the Appendix A.1.2 a faithful replication of the geographic treatment from [Benshaul-Tolonen, 2018] to complete and compare with the results on infantile mortality, and to highlight the differences between the geographic and topographic analyses.

4.1 Matching DHS clusters and mining sites

4.1.1 Challenges

Several econometric estimations can be found in the literature, each raising different endogeneity issues. The differences in terms of empirical strategies could explain the disparities in the results of mining activity on health. For instance, using different empirical strategies, [Von der Goltz and Barnwal, 2019] find important effects on stunting in young children while [Benshaul-Tolonen, 2018] finds a reduction in infant mortality rates. If these results are not contradictory, they raise the question of the direction, the magnitude, and the heterogeneity of the health-wealth trade-off implied by industrial mining activities. We discuss the possible empirical strategies in this section and the challenge of matching the surveyed villages to the mining sites.

Numerous papers use the geographic proximity to an active/open mine as a proxy for exposure to pollution ([Von der Goltz and Barnwal, 2019, Kotsadam and Tolonen, 2016, Aragón and Rud, 2016]). Following a Difference-in-Difference strategy, [Von der Goltz and Barnwal, 2019] restrict their sample to households living within 20 km of a mine, and compare households living within 0 to 5 km to households living between 5 and 20 km (spatial difference), before and after the opening of a mine (temporal difference). The coefficient of interest is therefore the interaction between being close to a mine and the mine's activity status during the child's year of birth. They implement a mine-level panel analysis, which main limitation is to build unbalanced treatment and control groups. This imbalance might be endogenous to socio-economic outcomes or polluting behaviors. Indeed, as each DHS cluster is paired to the closest mine, this mechanically excludes from the control group DHS clusters that are in both distance categories (within 5 km of mine A but within 5-20 km of mine B). This is more likely to happen in areas of high mining activity density (for instance Ghana). Let us consider the two distance categories as the treatment/control groups form a usual Difference-in-Difference (being close to a mine being the treatment). This empirical strategy increases the size of the treatment group (and decreases the size of the control group) in areas with high mining activity density, as DHS clusters that are in both categories are mechanically put in the treated areas. It is hard to believe that the mining density is a random allocation and that areas with low density will be comparable to the ones with high density, in terms of health, wealth, public access, and pollution. This matters as the mine fixed-effect empirical strategies rely on a within mine buffer area comparison. The strategy compares the infant mortality of the treated area (within 5 km) of mine A to the mortality of the control area (5-20 km) of mine A, before and after activity: the mines that matter are the ones that are paired to DHS clusters in both distance categories. This means that potentially, with this estimation, the authors endogenously select mines that are in places with low mine activity density, which might be correlated with the intensity or type of pollution of the mine (no peer effect because only one mine around), but also some socio-characteristics of the population in the neighbourhood (access to public services, wealth) changing how the health will be affected by the pollution. This might bias the estimate downward (mining activity affecting more the mortality because of fewer pollution behaviors) or on the contrary upward (because of major industrial areas). To reassure the reader, [Von der Goltz and Barnwal, 2019] use an Instrumental Variable (IV) strategy, instrumenting the mine location with mineral deposit. They use mineral deposit information from S&P data, proxied by *deposits that are being explored or prepared for exploitation*. We argue here that mining exploration is also highly correlated to socio-demographic characteristics, such as density (more likely to search where the workforce exists) and it is far from random: pairing DHS to the nearest mineral exploration raises the same issues than directly with the mining site.

To avoid this limitation, [Benshaul-Tolonen, 2018] uses an administrative district fixed-effect panel, and extends the distances : treated households are living within 10 km from a mine, and are compared to control households living between 10 and 100 km from a mine. However, even without using a mine-fixed effect strategy, defining the treatment areas according to the geographical proximity to the mining site is also highly endogenous. Indeed, it is commonly argued that the conditions for an industrial mine to settle are the presence of mineral deposits, which is considered random. However, this is a necessary condition, but not sufficient, and the presence of a mine (and the declared presence of a mineral deposit) is highly correlated to the population density. For instance, if mining exploration needs labour force, mines are more likely to open in highly dense areas, where mechanically DHS is more likely to have surveyed individuals. Thus, treatment based on geographic proximity to the mine might be endogenous to the initial density of the area, again. The likelihood that a village has been surveyed by DHS is correlated with the presence of a mine within 10km (because being an area of high density), and thus its proximity to the mining site. As the district fixed effect panel relies on a within-district comparison (mortality in treatment areas in district A to mortality in control areas in district A), the fixed-effect only considers districts with DHS clusters in treatment and control zones. This rejects mechanically and endogenously from the regression districts with high heterogeneity in terms of density, and that become spatially heterogeneous because of the development of a mine (very close to mines located in highly dense areas, maybe linked to the arrival of the workforce, but not in control zones with low density), and selects districts that remained homogeneous in terms of density after the development of mines. Again, this might select only places that are more stable, well-off, and with less detrimental behavior in terms of pollution. This might bias the estimation upward (i.e less mortality linked to mining activity), and explain the positive effect of mines that [Benshaul-Tolonen, 2018] find on mortality in Africa.

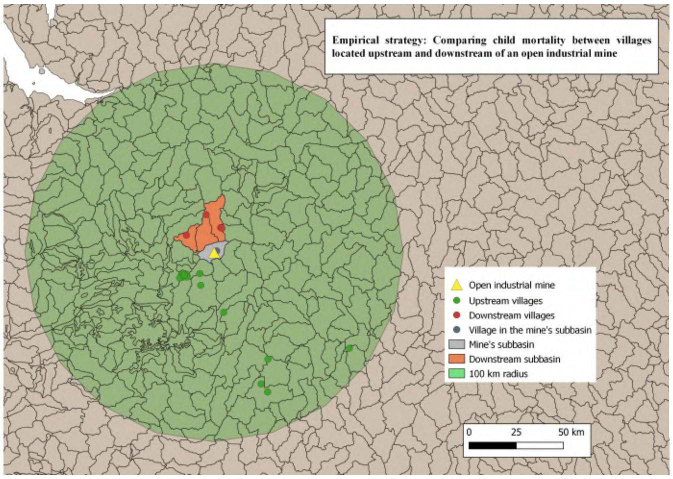
Controlling for district fixed effects might still not be enough if the whole district has a tradition of industrial mining activity. Therefore, it is all the more important to include the upstream-downstream analysis, since it will enable us to introduce a topographic difference within the same distance from a mine. Distance no longer is a proxy for exposure, as we use each household's relative topographic position to a mine.

4.1.2 Matching strategy

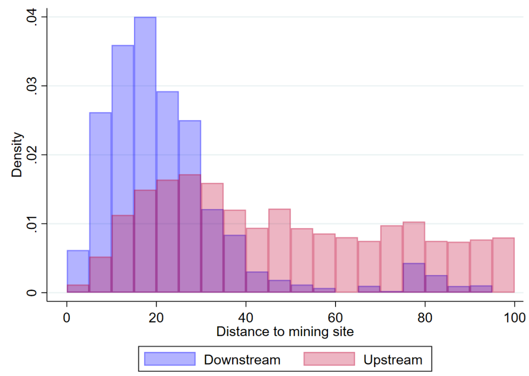
The matching of DHS clusters to mines represents a significant challenge, as each DHS cluster can be downstream of and close to several industrial sites in major mining areas. It introduces endogeneity in the sample selection and raises the issue of imbalanced samples. In this analysis, we propose the following matching to overcome this issue.

First, we construct a 100 km buffer around each DHS cluster, and register all mines within this buffer (independently of their activity status). We then categorize the topographic position of the DHS cluster relative to the industrial site (i.e we create a dummy equal to 1 if the cluster is downstream of the mine

Figure 6: Matching Strategy



(a) Scheme for matching strategy



(b) Density of distance to the mine

Notes : Figure (a) is a scheme that illustrates the matching strategy, giving the example of a mine, its main sub-basin, the three closest downstream sub-basins, and DHS clusters that are in the treatment and control areas. Figure (b) plots the density of the distance to the mining site for DHS clusters across their upstream-downstream position.
Sources : authors' elaboration on DHS, SNL and HydroSheds data.

and 0 if it is located upstream). This topographic position (upstream or downstream) is defined using the position of each sub-basin. As each cluster and sites have GPS coordinates, they lie in a specific sub-basin, and we used the relative position of each sub-basin to classify the DHS according to the paired mine. Through such a process, we also have pairs that are located in the same sub-basin, and for which it is impossible to say exactly whether the cluster is downstream or upstream of the mine. At this stage, for these specific couples, the topographic dummy is assumed to be equal to 1 (i.e for the moment, we consider that the DHS is downstream). Please note that, as explained in section 3.1.3, we used the finest Pfasftetter level 12 that breaks down subbasins at an average area of $100km^2$ (the size of the sub-basin varies according to their shape, cf Figure 6a). At this stage, some villages can be paired with several mines and can have more than one occurrence in the sample. The difficulty of the strategy lies in choosing the mine that will be paired to the cluster.

Second, we restrict the group of downstream DHS clusters to the ones that lie within one of the three closest sub-basins downstream of the mine's sub-basin, to focus on the potentially most contaminated areas if pollution follows the water flow. We focus on the definition of sub-basins and not on river flows as mining activity contaminates surface as well as groundwater resources, as we intend to capture both type of pollution. Section 5.3.3 looks at the heterogeneity of the main results according to geographical characteristics of sub-basins, including the presence of rivers, their importance, coverage area, or direction and magnitude of the flow for instance.

Third, to pair each cluster with only one mining site, we proceed as follows, based on the upstream and downstream concepts explained. If a DHS was in both groups (i.e downstream a mine A and upstream a mine B), then it is automatically assigned to the downstream group, and it is paired to the mine from which it is downstream (i.e it is paired to mine A), regardless of its activity status. At this stage, some clusters may still be counted twice, as they can be upstream of several mines, or in the three closest sub-basins downstream of several mines. To complete the uniqueness of the coupling, we paired each cluster to the nearest mine, regardless of its activity status as well.

Now, the coupling is terminated. In conclusion, the DHS clusters are attached to the nearest mine from which they are downstream up to the third sub-basin level, or else attached to the nearest mine upstream up to a radius of 100km. The final remaining problem relates to the clusters that are in the mine's same sub-basin, which we have so far identified as being downstream. We eliminated from the main analysis all DHS villages which are located in the same sub-basin of the mines from which they were matched. Also, this reduces the noise linked to the random displacement of DHS villages [15], and avoids to allocate villages as being downstream whereas they are upstream due to the displacement, as it drops the closest areas around the mine.

The matching is illustrated in Figure 6a which represents a sample of sub-basins, a mine, the main sub-basin of the mine, the three closest downstream sub-basins and the DHS villages upstream at 100km. The figure points to the fact that for each mine, the upstream treatment area is larger than the treated area. This mechanically balances the group of control and treated for statistical analysis, but also creates a difference in the distance to the mine distribution between the upstream and downstream villages, as shown in Figure 6b. We observe that the DHS in the downstream group follow a left-centered distribution, with the highest density around 10-15km, while the villages upstream are rather evenly distributed up to 100km. This bias mechanically introduced by the matching is discussed in the robustness section **??**, which shows that the main result is maintained when we restrict the total sample to different distances, in order to convince that we are isolating an effect linked to the position downstream rather than from a distance to the mine.

4.2 Identification Strategy

The main analysis relies on a Difference-in-differences strategy using the topographic position (upstream-downstream) of a DHS cluster relative to a mine deposit in order to identify the channel of water pollution. We propose a staggered Difference-in-Difference specification, with a subbasin fixed effect panel for each mine. We isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison. We use the definition of a subbasin and the level 12 of the Pfafstetter classification (the finest) to define for each subbasin the subbasins that are located upstream and downstream. We restrict our analysis using the matching strategy explained in previous section 4.1.2. We compare health outcomes in upstream-downstream areas, both before and after the opening the paired mine. The empirical strategy can be formally written as follows:

$$\begin{aligned} Death_{i,v,c,m,SB} = &\alpha_0 + \alpha_1 Opened_{birthyear,i,v} + \alpha_2 Downstream_{v,SB} \\ &+ \alpha_3 Opened_{birthyear,i,v} \times Downstream_{v,SB} + \alpha_4 X_i \\ &+ \gamma_S B + \gamma_{SB-trend} + \gamma_{c,birthyear} + \epsilon_v \end{aligned} \tag{1}$$

With $Death_{i,v,c,m,SB}$ a dummy equal to one if child i from DHS village v of country c, has reached the n^{th} month and has died (n being 12 for the 12-month old mortality, same for 24 months). $Opened_{birthyear,i,v}$ is a dummy equal to 1 if the mine, which is located in subbasin SB, has opened before child i's year of birth. $Downstream_{v,SB}$ is a dummy of relative position (equal to 1 if village DHS v is located in the downstream subbasin SB, and 0 if it is upstream), X_i a vector of child and mother level controls (mother's age, age square, years of education, urban residency). Finally, γ_{SB} is a subbasin fixed effect, $\gamma_{SB-trend}$ a subbasin linear birthyear trend and $\gamma_{c,birthyear}$ a country-birthyear fixed effect. This analysis is a staggered design as the treatment shock (mine opening) does not occur at the same time for each DHS cluster.

The main regression is run without the DHS clusters that lie within the same sub-basin as the mine they are coupled with, as discussed in previous section. The list of countries and survey years used in the main regression are given in Table **??**, they are the countries that had at least more than one DHS round

[15]"In DHS household surveys the GPS coordinate displacement process is carried out as follows: urban clusters are displaced a distance up to two kilometers (0-2 km) and rural clusters are displaced a distance up to five kilometers (0-5 km), with a further, randomly-selected 1% (every 100th) of rural clusters displaced a distance up to 10 kilometers (0-10 km)", as mentionned in [USAID, 2013]

(to limit differences in terms of temporal variation) [16]. Finally, the main regression is made for mine that are associated with heavy metal mines (metals with density higher than $5gcm^{-3}$ [Briffa et al., 2020], which are the metals listed in Table 19 in Section A.2.2 of the Appendix [17]. We also include coal mines, as their extraction is associated with mercury and arsenic that are highly toxic heavy metals. Section 6.1.3 shows that the main results are stable according to the list of metals and Section A.5 when dropping one country one by one.

4.3 Descriptive statistics and parallel trends

In this section, we describe the balance tables that play a key role in our analysis, out of parsimony. Sections A.1 and A.3.2 in the Appendix displays extensive descriptive statistics.

Table 1: Balance Table - Double Difference with Topographic Treatment - Descriptive Statistics

	Before Mine Opening					After Mine Opening					Within	Within	Within
	Upstream		Downstream		Diff	Upstream		Downstream.		Diff			
	N	Mean /(SD)	N	Mean /(SD)	(4-2) /(p.v)	N	Mean /(SD)	N	Mean /(SD)	(9-7) /(p.v)	(7-2) /(p.v)	(9-4) /(p.v)	(12-11) /(p.v)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dth<12													
All	44725	0.071	8102	0.073	0.003	24697	0.055	5085	0.054	-0.001	-0.015	-0.019	-0.004
		(0.256)		(0.26)	(0.398)		(0.228)		(0.226)	(0.717)	(0)	(0)	(0.511)
Mines	279		235			204		186					
Dth<24													
All	33688	0.092	6114	0.097	0.005	17465	0.069	3582	0.075	0.006	-0.023	-0.022	0.001
		(0.289)		(0.296)	(0.271)		(0.254)		(0.263)	(0.239)	(0)	(0)	(0.494)
Mines	279		234			192		172					

Notes: Standard errors and p-values in parentheses. Outcomes descriptive statistics of under 12 and 24 months mortality,for villages Upstream and Downstream mining sites, for individuals born before and after the opening of the mine.

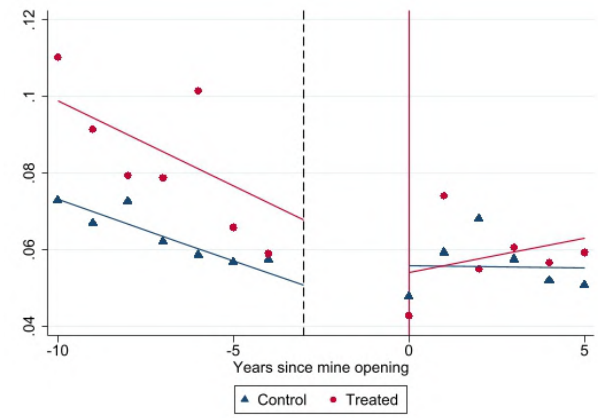
Balance Table 1 displays descriptive statistics of under 12 and 24 months mortality, for DHS villages surveyed before any mine opening, and those surveyed after the beginning of industrial activity. Table 1 compares the changes in infantile mortality before and after the opening of a mine, for places upstream *vs* downstream of the mining site, following the matching strategy. It displays also the number of individuals and paired mines in each group of the analysis. Maps from Figures 21 22 23 24 in Appendix display the spatial and temporal distribution of mining sites and DHS clusters, as well as main outcomes (infantile mortality rates and mine opening), restricted to the sample of the main analysis. Figure 25 in the Appendix identifies the country with the biggest stock of open mines in our sample (Ghana, Zimbabwe, Tanzania with the highest density of open mines nearby DHS), as well as insights on the variation in mine opening over the period per country.

On average, Upstream and Downstream areas have non-significant differences in terms of under 12 and 24 months mortality (columns 5 and 10). For both Upstream and Downstream clusters, the opening of a

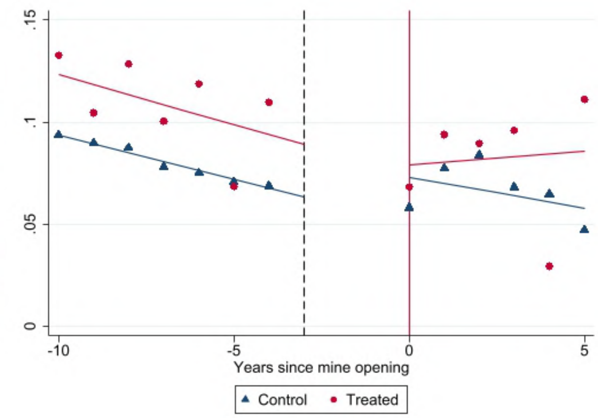
[16] The list of countries within our sample are : Benin, Burkina Faso, RDC, Burundi, Cote d'Ivoire, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Liberia, Lesotho, Madagascar, Mali, Malawi, Nigeria, Niger, Namibia, Rwanda, Sierra Leone, Senegal, Togo, Tanzania Zambia and Zimbabwe

[17] The list of metals of the main sample are : Gold, Copper, Iron Ore, U308, Nickel, Platinum, Zinc, Chromite, Illmenite, Lanthanides, Manganese, Tin, Cobalt, Tungsten, Tantalum, Vanadium, Niobium, Heavy mineral Sands, Silver, Lead and Coal

Figure 7: Linear Trends dropping investment phase



(a) Infant mortality Rate - 12 months



(b) Infant mortality Rate - 24 months

Notes : The Figure plots the trends of (a) 12-month and (b) 24-month mortality rates. Years since the opening of mine are in abscissa. The control group is DHS clusters upstream of the matched mine, while the treated are DHS clusters downstream. The mortality rates are computed on individuals that reached the age of 12 and 24 months. Figures (a) and (b) allow for a break at year -3, which corresponds to the exploration phase.
Sources : authors' elaboration on DHS, SNL and HydroSheds data.

mine significantly decreases the mortality probability (columns 11 and 12), which is in line with the result of [Benshaul-Tolonen, 2018], and with the fact that mortality rates decrease over time in Africa (Figures 22 23). Table 1 shows that this reduction is overall slightly more important in upstream areas than in downstream areas for under 24-month mortality (column 13), while it is the contrary for under 12 months mortality. This is suggestive evidence that positive effects (in the sense of reduction of mortality) found by [Benshaul-Tolonen, 2018] of mining activity are weakened due to pollution exposition through water contamination. Of course, Table 1 does not include any controls, and is mainly used to test whether control and treated (i.e upstream and downstream areas) are statistically different, which is not the case. Table 20 replicates this exercise for control variables, and we observe some statistical differences relative to the topographic position as well as the opening of the mine, showing the importance to test for parallel trends.

The key assumption of the DiD strategy is that the outcome, i.e infantile mortality rates, in upstream and downstream areas would follow the same time trend in the absence of the mine opening. The common trends assumption cannot be tested, however, we can observe the pre-treatment data and the evolution of infantile mortality before each mine opening according to the topographic position. Figure 7 plots the linear trends of the 12 and 24-month mortality rates, and distinguishes between control (upstream) and treated (downstream) DHS cluster, before and after the opening of the coupled mine. Figure 7a and 7b allow for a break at year -3, which corresponds to the exploration phase, and is shown in the event study (Figure 8b section 5.2) to have an impact on the outcome. This reduce also noise around the timing of opening. The linear trends show that upstream and downstream communities follow similar trends in both outcomes before the mine openings, which suggests the parallel trend assumption.

5 Results

This section displays the results of our main analysis. The first section displays the overall effects of mining activity on local population's health according to the topographic treatment. The second section looks at the dynamic impact of the staggered DiD strategy and describes the event study of the effects. Third, we display the heterogeneity of the effect according to households', mines' and geographic characteristics.

5.1 Average effects of mine opening

This section displays the main results of this paper from equation 1. Table 2 gathers our main results with mine sub-basin and country-birth-year fixed effects. We also include mine sub-basin and birthyear linear trends, adjusting for spatial and period-specific cofounders and trends, and commodity fixed effects (Section 6.1.5 shows that the results are stable when dropping fixed effects one by one). Columns (1) and (2) give the results for the under 12 months mortality, while columns (3) and (4) for the under 24 months mortality. Columns (1) and (3) show the results for the whole main sample, which is defined in the previous Section 4. We remind that the main sample drops all the DHS clusters that lie within the same sub-basin as the mining site, and keep only mines where the primary commodity is linked with the use of heavy metals (so heavy metals from Table 19 plus coal). The countries used in the regressions are listed in Table 13. Section A.2.2 shows that the main results are stable according to the list of metals and Section A.5 when dropping countries one by one. Columns (2) and (4) give the estimators for the same sample while dropping the three years previous to the opening of the mine that can be identified as a pre-opening period of investment and exploration phase [Benshaul-Tolonen, 2018], as we can interpret in the event study (Figure 8b). This also enables to reduce the noise linked to the timing of the opening.

The results show that being downstream when a mine opens increases by 2.3 percentage points (p.p) the 24-month mortality rate. This corresponds to an increase by 27% as the average 24-month mortality increases from 8.5% to 10.8%. When dropping the investment phase, the mortality rates at 24 monhts increase by 2.7 p.p, which represents an increase of 31%. The higher coefficient in Column (4) is coherent as it reduces the noise around the timing of the mine opening, and as it is likely that small production had began also during the investment phase. The results are not significant concerning the 12-month mortality rate, and are very close to zero, showing no difference between individuals leaving upstream to those leaving downstream. This could suggest a lag in the effect of water pollution on children's health, or be the consequences of the compensated effects between the negative externalities (such as pollution) and positive ones (such as the construction of health facilities).

Table 2: Effects of industrial mining activity on under 12, 24 mortality - Topographic Treatment - All Households

	Death <12m		Death < 24m	
	All	Drop investment phase [t-1;t-3]	All	Drop investment phase [t-1;t-3]
	(1)	(2)	(3)	(4)
Downstream×Open	0.000727 [0.00756]	0.00967 [0.00866]	0.0229** [0.00985]	0.0273** [0.0109]
Downstream	-0.0140** [0.00612]	-0.0201*** [0.00668]	-0.0174*** [0.00673]	-0.0197*** [0.00736]
Open	0.00707 [0.00526]	0.00171 [0.00647]	0.00213 [0.00715]	-0.00727 [0.00905]
Birth order number	0.00371*** [0.000745]	0.00357*** [0.000788]	0.00488*** [0.000918]	0.00477*** [0.000972]
Mother's age	-0.0108*** [0.00117]	-0.0108*** [0.00125]	-0.0126*** [0.00152]	-0.0125*** [0.00163]
Mother's age square	0.000151*** [0.0000185]	0.000151*** [0.0000196]	0.000167*** [0.0000237]	0.000164*** [0.0000252]
Years edu.	-0.00134*** [0.000287]	-0.00128*** [0.000307]	-0.00174*** [0.000365]	-0.00183*** [0.000390]
Urban	-0.00628** [0.00285]	-0.00696** [0.00307]	-0.0121*** [0.00356]	-0.0142*** [0.00381]
Constant	0.237*** [0.0180]	0.242*** [0.0193]	0.296*** [0.0238]	0.304*** [0.0255]
Birthmonth FE	Yes	Yes	Yes	Yes
Ctry-Bthyear FE	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes
Drop [t-1,t-3]	No	Yes	No	Yes
N	82571	75076	60814	55218
R2	0.0264	0.0278	0.0365	0.0385
Outcome Mean	0.0652	0.0666	0.0851	0.0873
Outcome Mean - Downstream	0.0657	0.0662	0.0887	0.090
Outcome Mean - Upstream	0.0650	0.0666	0.0844	0.0868

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Columns (1) and (2) give the results for the 12 months mortality rate, while (3) and (4) for the 24 months mortality rate. Columns (2) and (4) dipslay the results without the investment phase, meaning without the three years preceding the opening of the mine.

5.2 Dynamic impact of mine opening

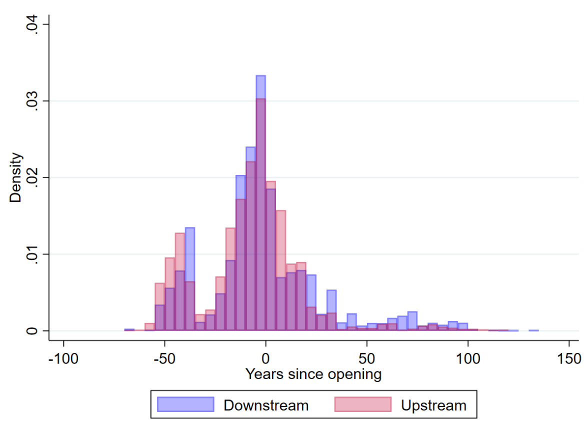
We extend our results by analyzing the dynamic effects of the opening of an industrial mine. We run a staggered difference-in-difference and estimate the following equation:

$$
\begin{aligned}
Death_{i,v,c,m,SB} = & \sum_{t=-10}^{10} \beta_t 1_{[birthyear=t]} Opened_{birthyear,i,v} \times Downstream_{v,SB} \\
& + \alpha_4 X_i + \gamma_S B + \gamma_{SB-trend} + \gamma_{c,birthyear} + \epsilon_v
\end{aligned}
\tag{2}
$$

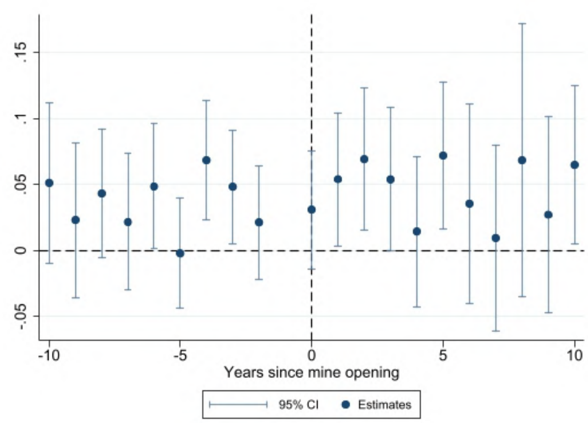
We plot in figure 8a the distribution of the years before and after the mine opening across downstream and upstream villages to make sure of the comparability of our two groups. The results of the event study are plotted in figure 8b. The distribution gives the proportion of mines that are pure controls, i.e those that have not been opened yet (the negative part of the distribution), as well as those that are historical mines (as we have mines opened since 100 years). As the goal of the analysis is to estimate the effect of the opening of the mine, and that the Distribution from Figure 8a shows that the majority of the sample is within [-10,10], we plot the event study using this time window.

We find a positive and significant effect of mine's opening on 24 month mortality rate up to three years after: the opening of a mine increase the 24 months mortality rate by 0.05 percentage points. The effect is persistent during at least three years . More interestingly, we find similar positive and significant effect three years before the opening, suggesting that the investment phase also produce detrimental effects on the local environment and on child mortality.

Figure 8: Event study graphs - 24 months mortality rates



(a) Distribution of years since opening



(b) Event Study

Notes : Figure (a) plots the distribution of the years before and after the mine opening across downstream and upstream villages, for the 24 months mortality rates sample. Negative values correspond to mines that did not open yet at the birth year of the child. Figure (b) plots the coefficient of our treatment before and after the opening of a mine on the 24-month mortality rates. It gives the average treatment effects of mine opening on 24-month mortality, 10 years before the mine opening and 10 years after.
Sources : authors' elaboration on DHS and SNL data.

5.3 Heterogeneity analysis

This section explores the heterogeneity of the main result, focusing on under 24 months mortality, according to some household, mine, and geographic-specific characteristics.

5.3.1 Household Characteristics

Table 3: Effects of industrial mining activity on under 24 mortality - Topographic Treatment - All Households, rural households, and control for migration

	Death < 24m			
	All	All	Rural Households	Rural Households
	(1)	(2)	(3)	(4)
Downstream×Open	0.0229**	0.0235*	0.0326***	0.0290*
	[0.00985]	[0.0133]	[0.0121]	[0.0163]
Downstream	-0.0174***	-0.0170*	-0.0249***	-0.0262***
	[0.00673]	[0.00958]	[0.00720]	[0.0101]
Open	0.00213	-0.00880	-0.00171	-0.0102
	[0.00715]	[0.0113]	[0.00817]	[0.0129]
Birth order number	0.00488***	0.00327***	0.00629***	0.00447***
	[0.000918]	[0.00125]	[0.00108]	[0.00149]
Mother's age	-0.0126***	-0.0127***	-0.0155***	-0.0159***
	[0.00152]	[0.00203]	[0.00182]	[0.00244]
Mother's age square	0.000167***	0.000171***	0.000202***	0.000211***
	[0.0000237]	[0.0000319]	[0.0000279]	[0.0000376]
Years edu.	-0.00174***	-0.00235***	-0.00139***	-0.00207***
	[0.000365]	[0.000491]	[0.000529]	[0.000724]
Urban	-0.0121***	-0.0147***		
	[0.00356]	[0.00514]		
Migrant		0.00918***		0.00586
Constant	0.296***	0.310***	0.347***	0.366***
	[0.0238]	[0.0314]	[0.0287]	[0.0379]
Birthmonth FE	Yes	Yes	Yes	Yes
Ctry-Bthyear FE	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes
N	60814	36377	44999	26745
R2	0.0365	0.0467	0.0432	0.0539
Outcome Mean	0.0851	0.0925	0.0909	0.0997

Notes: Standard errors clustered at the village level, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Column (1) display the main result for all households, while column (2) show the same results controlling for household where the mother has migrated. Columns (3) and (4) replicate this analysis focusing on rural households.

Table 3 displays the main result (Column 1), and restricts the sample for rural households (Column 2). We find a stronger effect of our treatment among rural households: the opening of a mine increases the mortality under 24 months by 3.3 percentage points within households located downstream of a mine, compared to households living upstream (which corresponds to an increase of the mean mortality rates by 37%). The remoteness and scarcity of facilities can explain the effects, as well as the increased lack of monitoring among industrial mining sites that are far away. Columns (3) and (4) show that our results are robust when controlling for households where the mother has reported to have migrated. Please note the fall in the observation number between Columns (1) and (2): indeed, the sample in Columns (2) and (4) is restricted to individuals that answered the question (which is around 60% Table 18). The variable asked mothers whether they have ever migrated to the village, or if they have always lived there. This variable controls for in-migration, which is an important effect of the opening of a mine that attract new working population (cf Section 6.1.4 for more discussion on bias linked to migration).

Table 4: Effects of industrial mining activity on under 24 mortality - Topographic Treatment - All Households, control for piped water access and visits to health facilities

	Death < 24m			
	(1)	(2)	(3)	(4)
Downstream × Open	0.0229**	0.0230**	0.0187*	0.0188*
	[0.00985]	[0.00984]	[0.0102]	[0.0102]
Downstream	-0.0174***	-0.0175***	-0.0150**	-0.0151**
	[0.00673]	[0.00671]	[0.00692]	[0.00691]
Open	0.00213	0.00196	0.00481	0.00456
	[0.00715]	[0.00716]	[0.00758]	[0.00759]
Birth order number	0.00488***	0.00485***	0.00537***	0.00534***
	[0.000918]	[0.000919]	[0.000967]	[0.000967]
Mother's age	-0.0126***	-0.0126***	-0.0122***	-0.0122***
	[0.00152]	[0.00153]	[0.00160]	[0.00160]
Mother's age square	0.000167***	0.000166***	0.000160***	0.000159***
	[0.0000237]	[0.0000237]	[0.0000250]	[0.0000250]
Years edu.	-0.00174***	-0.00169***	-0.00139***	-0.00132***
	[0.000365]	[0.000368]	[0.000387]	[0.000391]
Urban	-0.0121***	-0.0106***	-0.0100***	-0.00805**
	[0.00356]	[0.00375]	[0.00385]	[0.00406]
Piped Water		-0.00465		-0.00589
		[0.00357]		[0.00381]
Visited Health Facility			-0.00588**	-0.00579**
			[0.00287]	[0.00287]
Constant	0.296***	0.296***	0.288***	0.289***
	[0.0238]	[0.0237]	[0.0249]	[0.0249]
Birthmonth FE	Yes	Yes	Yes	Yes
Country-Bthyear FE	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes
N	60814	60814	54333	54333

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines.

Table 4 shows that our main results hold when we control for households that have access to piped water as a main source of drinking water and for whether mothers have visited a health facility within the past 12 months. The fact that our coefficient remains in the same order of magnitude is reassuring for our estimations, in order to avoid potential omitted variable bias.

5.3.2 Mine's Characteristics

Our results hold when controlling for the nature of the ownership of the mine (Table 5 column 1). A mine was considered as domestic if at least one of the owning company is from the same country as the country of location, and they represent 17.8 percent of a mine sample. We find no effect of mine opening when we restrict the sample to domestically owned mines (column 2) whereas our results hold when we restrict to the foreign-owned only mines (column 3). This could be explained by the less dramatic management of a mine if a national company is involved. We look at the openpit nature of the industrial site, which concern 21.6 percent of our mines sample, and find that our results hold when controlling for the mining type (column 4). We find significant effects that are threefold when we restrict to the openpit mines sample (column 6), and no significant effect when we exclude the openpit mines (column 5).

Table 5: Effects of industrial mining activity on under 24 mortality - Topographic Treatment - All Households mine ownership and type

	Death < 24m					
	(1)	(2)	(3)	(4)	(5)	(6)
	Ownership control	Domestic owner	Foreign owner	Mining type control	Non open-pit mines	Open-pit mines
Downstream×Open	0.0229**	0.00294	0.0259**	0.0228**	0.0235	0.0691**
	[0.00987]	[0.0220]	[0.0119]	[0.00984]	[0.0176]	[0.0287]
Downstream	-0.0173**	0.0185	-0.0213***	-0.0170**	-0.0101	-0.0326
	[0.00673]	[0.0204]	[0.00722]	[0.00670]	[0.0154]	[0.0210]
Open	0.00251	0.0205	0.00299	0.00271	-0.00970	-0.0356*
	[0.00717]	[0.0240]	[0.00793]	[0.00713]	[0.0151]	[0.0200]
Birth order number	0.00487***	0.00624***	0.00454***	0.00486***	0.00606***	0.00604***
	[0.000918]	[0.00199]	[0.00103]	[0.000918]	[0.00143]	[0.00162]
Mother's age	-0.0126***	-0.00526	-0.0143***	-0.0126***	-0.0134***	-0.0139***
	[0.00152]	[0.00327]	[0.00173]	[0.00152]	[0.00237]	[0.00255]
Mother's age square	0.000166***	0.0000454	0.000195***	0.000167***	0.000177***	0.000181***
	[0.0000237]	[0.0000517]	[0.0000267]	[0.0000237]	[0.0000371]	[0.0000398]
Years edu.	-0.00174***	-0.00217***	-0.00155***	-0.00174***	-0.00152***	-0.00141**
	[0.000365]	[0.000696]	[0.000433]	[0.000365]	[0.000543]	[0.000602]
Urban	-0.0121***	-0.0229**	-0.0103***	-0.0121***	-0.00861	-0.0103*
	[0.00356]	[0.00909]	[0.00391]	[0.00355]	[0.00529]	[0.00604]
Domestic ownership	-0.0364					
	[0.0341]					
Mine type				-0.0469**		
				[0.0203]		
Open Pit					-0.0157	
					[0.0365]	
Constant	0.303***	0.173***	0.322***	0.314***	0.311***	0.318***
	[0.0246]	[0.0516]	[0.0270]	[0.0249]	[0.0477]	[0.0407]
Birthmonth FE	Yes	Yes	Yes	Yes	Yes	Yes
Ctry-Bthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes	Yes	Yes
N	60814	11722	49078	60814	23160	16896
R2	0.0366	0.0481	0.0384	0.0366	0.0454	0.0459
Outcome Mean	0.0851	0.0749	0.0876	0.0851	0.0772	0.0755

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. All columns give the results for the 24 months mortality rate.

5.3.3 Geographic Characteristics

In this section, we are looking at the heterogeneity of the main result on 24 months mortality rates, according to the characteristics of the area at the level of the mine (i.e the area including the three closest downstream sub-basins, and upstream areas up to 100km), such as the presence of rivers in the downstream sub-basins, the importance of the water flow and coverage of the area.

[WORK IN PROGRESS]

6 Robustness and Sensitivity analysis

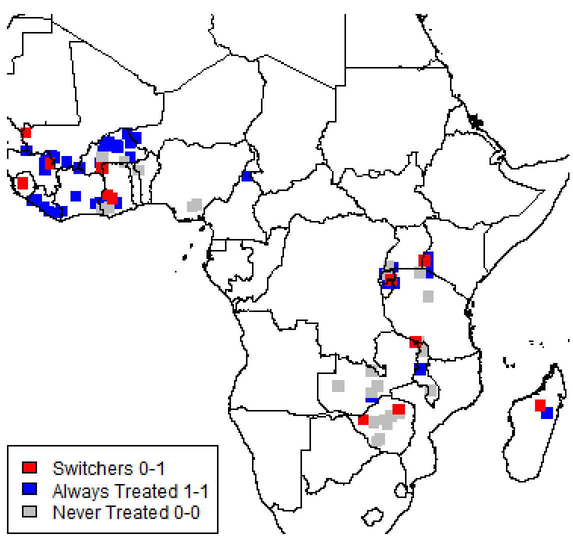
6.1 Robustness

In this section, we propose a list of robustness checks, such as the restriction of the main sample to a balanced sample and a restriction of the analysis according to different distances to the industrial site. We also discuss about the sample selection of mines and about the migrant selection issue.

6.1.1 Balanced Panel

In this section, we replicate the main analysis on a restricted balanced panel of mine sub-basins. As the staggered DiD is driven by the changes in mortality rates of switchers, the goal is both to verify that the results are robust when we restrict to a balanced panel, and also to exactly identify the group of switchers. Figure 9 plots the different groups from the balanced panel, and displays the group of switchers, the always treated and never treated control groups. The balanced panel is built as explained in the following paragraphs.

Figure 9: Balanced Panel - Group identification



Notes : The Figure plots the groups areas accross the three groups of the balanced panel, for the 24-month mortality rate.
Sources : authors' elaboration on DHS and SNL data.

Thanks to the balanced panel, we restrict the sample to pure controls. The setting of this paper verifies the existence of stable groups, meaning that there are groups for which the exposure to the treatment do not change [de Chaisemartin and D'Haultfœuille, 2020]. Please note that groups are made of subgroups, which are the level at which we define the treatment to build the balanced panel. Subgroups have unique mine sub-basin, and gather all individuals living within the three closest sub-basins downstream or upstream within 100km of the mine, for each year. It is easy to identify the subgroups that are the switchers, meaning the subgroups that experienced a change in the exposition to mining activity through the opening of an industrial mine, that are the subgroups playing a key role in the two-way fixed effects estimation.

The design distinguishes three "super groups". First, the subgroup for which a mine has opened between two different years (for which there are DHS observations), Group 1: the independent variable changes from 0 to 1. Second, the subgroup of areas for which the mine has always been opened and are thus always treated (Group 2: the independent variable which is an interaction, is always equal to 1), and third the subgroup of areas where mines have not yet opened (Group 3: the independent variable is equal to 0.). The third group is made of subgroups for which the mine has not opened yet but the opening is planned in the future, and of mines where no DHS cluster was surveyed after it opened. Please note that the way we built the group, make it possible that Group 2 includes open mines that have been surveyed only downstream or only upstream, as well as Group 3. As this is not the treatment that we want to study, the balance table corrects for this.

Table 6: Balanced Sample - Descriptive Statistics

	Group 1 : Switchers 0-1						Groups 2+3		Group 2 : 1-1		Group 3 : 0-0	
	All		Before Opening		After Opening							
N	Mean /(SD)	N	Mean /(SD)	N	Mean /(SD)	N	Mean /(SD)	N	Mean /(SD)	N	Mean /(SD)	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dth<12												
All	2532	0.061	1932	0.064	600	0.052	15462	0.067	4228	0.054	11234	0.072
	(0.24)		(0.245)		(0.222)		(0.25)		(0.226)		(0.258)	
Mines	14		14		14		72		31		41	
Dth<24												
All	1762	0.079	1378	0.081	384	0.07	12097	0.091	3160	0.073	8937	0.098
		(0.27)		(0.273)		(0.256)		(0.288)		(0.26)		(0.297)
Mines	13		13		13		75		31		44	

Notes: Standard errors and p-values are in parentheses. Outcomes' descriptive statistics of under 12 and 24 months mortality, for villages within the Group 1 Switchers for individuals born before and after the opening of the mine, then Group 2 always treated and Group 3 never treated.

The treatment being downstream of an open mine might not be reflected by Groups 1, 2, and 3. The balanced sample is made as follows. First, we only keep subgroups that have individuals both downstream and upstream. For subgroups in Group 2 and 3, meaning that the surveys have occurred only after the opening (Group 2) or before the opening (Group 3), we restrict to subgroups that have observations both upstream and downstream. Accordingly, for Group 1 made of Switchers, we restrict the sample to subgroups that have individuals surveyed both downstream and upstream in both periods, i.e before and after the opening. Table 6 gives the size of the three Groups in the balanced panel, as well as the associated number of mines, for both the 12 and 24 months mortality rates. Please note that the sample of the 12 months mortality rates and 24 months mortality rates are different. This is explained by the fact that we balanced the sample after droppping the individuals who did not reach the age of 12 months, as well as those who did not reach the age of 24 months, which explains the difference. Table 6 shows that the group 1 of Switchers accounts for around 13% of the total balanced sample. We observe that the average mortality rates have decreased over time, before and after the opening of the mining site in the Switcher Group, linked to the decrease of infantile mortality in Africa over time, which is in coherence with Balance Table 1 and Figures 4 and 5, and which highlights the importance of controlling for trends. Figure 9 plots the groups in order to identify where the switchers lie, and what are the areas that are key in the main estimation.

Columns (1) and (3) in Table 7 show the results of the main estimation for the 12 and 24-month mortality rates. Columns (2) and (4) replicate the main analysis for the balanced panel (23% of the main sample). Column (4) shows that when focusing on the balanced panel, being downstream of an opened mine increases the 24-month mortality rate by 3 p.p, which represents an increase by 33% of the mortality. Please note that, despite the small sample, the results remain significant, which is coherent with the fact that the switchers from Group 1 drive the main estimation's results.

Table 7: Effects of industrial mining activity on under 12, 24 mortality - Topographic Treatment - All Households and Balanced Panel Sample

	Death <12m		Death < 24m	
	All	Balanced Panel	All	Balanced Panel
	(1)	(2)	(3)	(4)
Downstream × Open	0.000727	0.0132	0.0229**	0.0305**
	[0.00756]	[0.0123]	[0.00985]	[0.0151]
Downstream	-0.0140**	-0.0170**	-0.0174***	-0.0211**
	[0.00612]	[0.00734]	[0.00673]	[0.00829]
Open	0.00707	0.0136	0.00213	0.0191
	[0.00526]	[0.0222]	[0.00715]	[0.0321]
Birth order number	0.00371***	0.00230	0.00488***	0.00236
	[0.000745]	[0.00157]	[0.000918]	[0.00196]
Mother's age	-0.0108***	-0.0127***	-0.0126***	-0.0150***
	[0.00117]	[0.00276]	[0.00152]	[0.00351]
Mother's age square	0.000151***	0.000187***	0.000167***	0.000209***
	[0.0000185]	[0.00276]	[0.0000237]	[0.0000544]
Years edu.	-0.00134***	-0.00167**	-0.00174***	-0.00188**
	[0.000287]	[0.000729]	[0.000365]	[0.000896]
Urban	-0.00628**	0.00538	-0.0121***	0.00141
	[0.00285]	[0.00557]	[0.00356]	[0.00761]
Constant	0.237***	0.264***	0.296***	0.336***
	[0.0180]	[0.0425]	[0.0238]	[0.0550]
Birthmonth FE	Yes	Yes	Yes	Yes
Country-Bthyear FE	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes
N	82571	17979	60814	13839
R2	0.0264	0.0380	0.0365	0.0482
Outcome Mean	0.0652	0.0661	0.0851	0.0897

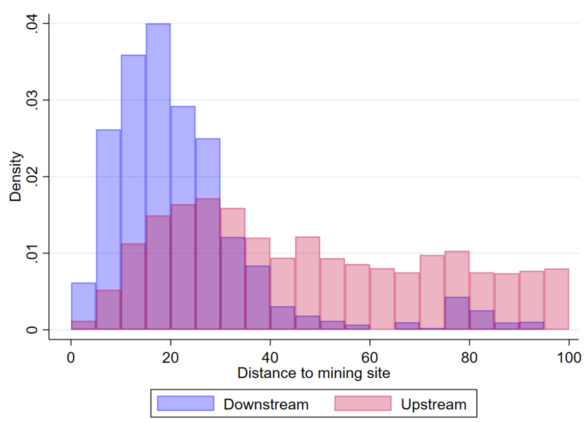
Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Columns (2) and (4) display the main estimation on the balance Panel.

6.1.2 Distinction with geographic treatment

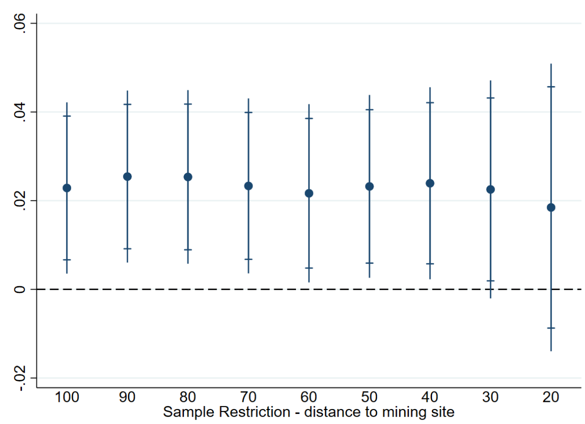
As discussed in Section 4.1.2, the matching strategy creates differences in terms of the distance to the mine between the downstream and upstream Groups. Figure 10a plots the distribution of the distance for both groups, and shows that downstream DHS clusters are mainly within 40 km of the mine, centered around 20km, while upstream areas are homogeneously distributed up to 100km. As the main contribution of this paper is to advocate that using proximity as a proxy for exposure to a mine embraces contradictory phenomena that compensate each other (positive and negative externalities), and that the average does not permit to identify the mechanism of water pollution, it is important to convince that our main estimation is driven by a difference in the topographic position of individuals rather than their distance to the mine. To do so, we replicate the main estimator for samples restricted to specific distance brackets.

Figure 10b plots the main estimate of the interaction variable on the 24-month mortality rates, according to samples based on distance: the first point on the left represents the main result for the whole sample up to 100km, the second one for the sample restricted to 90km, the third one up to 80 km and so on. Figure 10b shows that the main result remains stable and statistically significant, up to 20km. Please note that for the sample restricted to 20km, the number of observations is very low, and the sample is highly imbalanced in favor of downstream areas, explaining why the statistical significance does not hold. Please note that the regressions are made for the main sample, including heavy metal mines plus coal, and without DHS clusters located within the mine's sub-basin.

Figure 10: Impact accross distance to the matched mining site



(a)



(b)

*Notes :*Figure (a) plots the distribution of the distance of DHS clusters to the matched mining site, both for clusters downstream and upstream, for the 24 months mortality rates sample. Figure (b) gives the average treatment effect of being downstream of an open mine on the 24-month mortality for different samples. Please note that Figure (b) plots the results of 9 different regressions, with different sample sizes: the first dot corresponds to the overall sample, as it includes all clusters within 100km, while the second dot restricts the sample to the cluster that are within 90km of the mine, and so on. Please note that the sample size of the regressions decreases, as at the end the sample size of the regression for the clusters within 20km is relatively small.
Sources : authors' elaboration on DHS and SNL data.

6.1.3 Sample Selection of Mines

Table 8 shows that our results hold for mines extracting all types of commodities (column 1), although the coefficient is less precisely estimated and from a small magnitude than heavy metals and coal mines. Column (1) shows that the effect of being downstream of an open mine increase the mortality rates by 1.8 p.p (i.e 21%) for all mines, which is lower than the main result. This is coherent with the fact that heavy metals are environmentally and biologically toxic elements, associated with the mines with the highest impacts in terms of pollution. Our results are also stable when we restrict ourselves to heavy metal mines (listed in Table 19) but exclude coal (column 2). Last but not least, we find no significant effects of industrial gold mines only, which contrasts with the results from the literature.

Table 8: Effects of industrial mining activity on under 12, 24 mortality - Topographic Treatment

	Death < 24m		
	All Mines (1)	Heavy Metals without Coal Mines (2)	Gold Mines (3)
Downstream × Open	0.0179* [0.00947]	0.0208** [0.00966]	0.00789 [0.0168]
Downstream	-0.0120* [0.00673]	-0.0181*** [0.00672]	-0.0171* [0.00875]
Open	0.00186 [0.00600]	0.00274 [0.00706]	0.00514 [0.0103]
Birth order number	0.00522*** [0.000819]	0.00470*** [0.000903]	0.00588*** [0.00140]
Mother's age	-0.0123*** [0.00136]	-0.0124*** [0.00150]	-0.0155*** [0.00233]
Mother's age square	0.000163*** [0.0000212]	0.000164*** [0.0000234]	0.000207*** [0.0000362]
Years edu.	-0.00173*** [0.000325]	-0.00170*** [0.000359]	-0.00148** [0.000648]
Urban	-0.0121*** [0.00312]	-0.0123*** [0.00351]	-0.0139** [0.00558]
Constant	0.287*** [0.0211]	0.292*** [0.0234]	0.350*** [0.0361]
Birthmonth FE	Yes	Yes	Yes
Country-Bthyear FE	Yes	Yes	Yes
SB FE	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes
N	75512	62277	29835
r2	0.0355	0.0371	0.0414
Outcome Mean	0.0832	0.0843	0.0967

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. Column (1) focuses on all mine types, Colum (2) the heavy metal list without coal, and Column (3) with only Gold.

6.1.4 Migrant Selection

It is also detrimental to the analysis to understand the role played by in and out migration. All the above-mentioned empirical strategies struggle with residential sorting, namely the migratory movements correlated to industrial mining activity. For treatments based on geographical proximity, the results are biased if the opening of a mining site generates selected migration from treatment to control areas (or the reverse). For instance, it might trigger selection in the in-migration: the workforce might migrate from control (far from the mine) to treatment areas, and the workforce might be a selected population. Using the DHS, both [Benshaul-Tolonen, 2018] and [Von der Goltz and Barnwal, 2019] control for this phenomenon by excluding in-migrants from the analysis (either restricting the sample to individuals that have *always lived here*, or by excluding only new migrants that came two years before the opening of a mining site). However, the selection bias due to out-migration has been only lightly dealt with and there seems to be a way to improve. Indeed, we cannot omit the fact that if a mining site pollutes or creates effects that are highly detrimental to local populations, only those with the highest means might well be able to move away to a less exposed area, leaving the poorest and most vulnerable ones behind and thus creating an upward bias to our estimates. To deal with this issue, [Benshaul-Tolonen, 2018] has (i) looked at the negative change in infant health in the control communities (same trend), and has found no deterioration in the distance bins further away from mines ; (ii) ran a Spatial lag model to capture non-linear effects with distance from the mine. [Mamo et al., 2019] have implemented a Spatial Durbin model to control for spatial spillovers (at the grid cell level).

We delve deeper into the selection of our surveyed sample by looking at the migration status of respondents, keeping in mind that we can only trace in-migration and not out-migration. DHS surveys the number of years the respondent has lived in the village, town or city where she was interviewed, and we consider a person to be a migrant if she answered a specific number of years to this question. We find no significant effect of our treatment variable on the migration outcome (column 1 of table 9), when controlling for migration, restricting to migrants or stayers sample (columns 2 and 3). This absence of significant results provides a first evidence of a migrant selection bias. Yet, still further work needs to be undertaken to build stronger evidence.

Table 9: Effects of industrial mining activity on under 24 mortality - Topographic Treatment - Migrant Selection

	Migrant	Death < 24m	
		Non-migrant sample	Migrant sample
	(1)	(2)	(3)
Downstream × Open	0.0325	0.0284	0.0253
	[0.0342]	[0.0228]	[0.0174]
Downstream	-0.0293	-0.0192	-0.0182
	[0.0269]	[0.0160]	[0.0123]
Open	-0.0161	-0.00201	-0.0108
	[0.0216]	[0.0180]	[0.0142]
Birth order number	0.000695	0.00329*	0.00258
	[0.00203]	[0.00197]	[0.00163]
Mother's age	0.0148***	-0.00962***	-0.0146***
	[0.00319]	[0.00306]	[0.00281]
Mother's age square	-0.000247***	0.000132***	0.000195***
	[0.0000498]	[0.0000479]	[0.0000440]
Years edu.	0.00227**	-0.00153**	-0.00295***
	[0.00105]	[0.000756]	[0.000649]
Urban	0.0274**	-0.0258***	-0.0104*
	[0.0120]	[0.00820]	[0.00628]
Constant	0.382***	0.250***	0.359***
	[0.0496]	[0.0468]	[0.0440]
N	36377	14518	21807
R2	0.166	0.0743	0.0599
Outcome Mean	0.600	0.0870	0.0962

Notes: Standard errors clustered at the village level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines.

6.1.5 Other tests

We find stability in our results when dropping fixed effects one by one: birth month, commodity, and sub-basins birth year trend (Table 10) until keeping the two-way fixed effects in Column (4) (i.e keeping the mine sub-basin fixed effect and the country-birthyear fixed effect).

Table 10: Effects of industrial mining activity on under 12, 24 mortality - Topographic Treatment - Fixed Effects and controls dropping one by one

	Death < 24m			
	(1)	(2)	(3)	(4)
Downstream × Open	0.0229**	0.0227**	0.0201**	0.0195**
	[0.00985]	[0.00983]	[0.00958]	[0.00953]
Downstream	-0.0174***	-0.0174***	-0.0182***	-0.0178***
	[0.00673]	[0.00671]	[0.00669]	[0.00670]
Open	0.00213	0.00219	0.00345	0.00262
	[0.00715]	[0.00716]	[0.00715]	[0.00708]
Birth order number	0.00488***	0.00482***	0.00480***	0.00479***
	[0.000918]	[0.000917]	[0.000916]	[0.000916]
Mother's age	-0.0126***	-0.0125***	-0.0125***	-0.0125***
	[0.00152]	[0.00152]	[0.00152]	[0.00152]
Mother's age square	0.000167***	0.000166***	0.000166***	0.000165***
	[0.0000237]	[0.0000237]	[0.0000237]	[0.0000237]
Years edu.	-0.00174***	-0.00176***	-0.00176***	-0.00177***
	[0.000365]	[0.000364]	[0.000365]	[0.000364]
Urban	-0.0121***	-0.0147***		
	[0.00356]	[0.00514]		
Constant	-0.0121***	-0.0122***	-0.0120***	-0.0120***
	[0.00356]	[0.00356]	[0.00356]	[0.00356]
Country-Bthyear FE	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes
SB Bthyear trend	Yes	Yes	Yes	No
Commodity FE	Yes	Yes	No	No
Birthmonth FE	Yes	No	No	No
N	60814	60814	60814	60814
R2	0.0365	0.0362	0.0361	0.0355
Mean	0.0851	0.0851	0.0851	0.0851

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Column (1) includes all the Fixed Effects from the main specification, Column (2) excludes birth month FE, Column (3) main primary commodity FE, and Column (4) excludes the sub-basin trend, to display exactly a two-way FE model

6.2 Sensitivity analysis

We make sure that our results are robust when controlling for the fact that 81 percent of our final sample was hand-checked (column 2 of table 11). We find no significant effect when we restrict to the non hand-checked sample (column 3) but notice that our coefficient of interest remain from the same sign and order of magnitude. Our results are therefore mainly found thanks to our hand work (column 4), that is extensively described in the appendix.

Table 11: Effects of industrial mining activity on under 12, 24 mortality - Topographic Treatment - Sensitivity Controls

	Death < 24m			
	(1)	(2)	(3)	(4)
Downstream × Open	0.0229**	0.0228**	0.0252	0.0249**
	[0.00985]	[0.00984]	[0.0309]	[0.0106]
Downstream	-0.0174***	-0.0176***	-0.0175	-0.0193***
	[0.00673]	[0.00673]	[0.0236]	[0.00703]
Open	0.00213	0.00212	-0.0400	0.00148
	[0.00715]	[0.00715]	[0.0329]	[0.00763]
Birth order number	0.00488***	0.00488***	0.00402*	0.00500***
	[0.000918]	[0.000918]	[0.00233]	[0.00100]
Mother's age	-0.0126***	-0.0126***	-0.0154***	-0.0122***
	[0.00152]	[0.00152]	[0.00421]	[0.00164]
Mother's age square	0.000167***	0.000167***	0.000223***	0.000158***
	[0.0000237]	[0.0000237]	[0.0000682]	[0.0000253]
Years edu.	-0.00174***	-0.00174***	-0.00199***	-0.00161***
	[0.000365]	[0.000365]	[0.000718]	[0.000424]
Urban	-0.0121***	-0.0121***	-0.0126	-0.0120***
	[0.00356]	[0.00356]	[0.0108]	[0.00381]
Hand Checked		0.0463		
		[0.0330]		
Constant	0.296***	0.257***	0.340***	0.292***
	[0.0238]	[0.0372]	[0.0651]	[0.0255]
N	60814	60814	8695	52112
r2	0.0365	0.0366	0.0571	0.0380
Outcome Mean	0.0851	0.0851	0.0781	0.0863

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. Sample is on heavy metal mines.

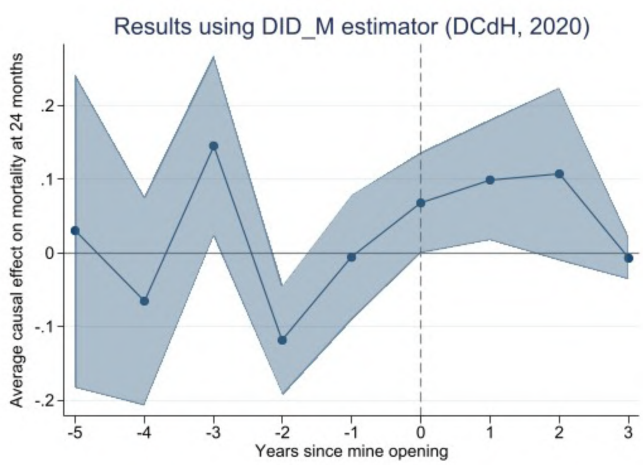
7 Discussion

7.1 Heterogeneous treatment effects with two-way fixed effects: [de Chaisemartin and D'Haultfœuille, 2020]

In the previous section, we estimated the effects by using standard difference-in-differences designs. However, recent developments in the estimation of difference-in-differences in staggered adoption designs ([Borusyak et al., 2021], [Goodman-Bacon, 2018], [Callaway and Sant'Anna, 2019], [de Chaisemartin and D'Haultfœuille, 2020]) show that the estimated ATT is a weighted sum of different ATTs with weights that may be negative (cf. Table [WORK IN PROGRESS]). The negative weights are an issue when the treatment effect is heterogeneous between groups over time, as one could have the treatment coefficient in those regressions is negative while the treatment effect is positive in every group and time period. We will show that in our context, standard estimation could indeed be exposed to these negative weighting issues and it could lead to substantial estimation errors. Therefore, we use [de Chaisemartin and D'Haultfœuille, 2020] estimation procedure to estimate the treatment effects in groups switching from no treatment to treatment compared to those remaining untreated. Another possibility is to take the "not-yet-treated" as a comparison group, [Callaway and Sant'Anna, 2019].

The estimator proposed by [de Chaisemartin and D'Haultfœuille, 2020] ensures that the treatment effects are estimated using only comparisons of units switching from no treatment to treatment compared to those remaining untreated. We also use their placebo estimate of [de Chaisemartin and D'Haultfœuille, 2020] that compares the evolution of child mortality from $t-10$ to $t-1$ in villages that are treated and those not treated (t being the date of opening of the site). The results are plotted in figure 11 where we see that there's a positive and significant effect of the mine opening three years before the official start date, which corresponds to the exploration and investment phase. We find coefficients higher in magnitude than in the event from our main results, but still need to take them with caution as we are still working on mastering the extent, hypothesis, and conditions of these new estimators.

Figure 11: [de Chaisemartin and D'Haultfœuille, 2020] Dynamic estimator



Notes : The Figure plots the results from the dynamic estimator built by [de Chaisemartin and D'Haultfœuille, 2020], using the $\text{did}_multiplegtcommandfromSTATA, forthe24-monthmortalityrates.$
Sources :$authors'elaborationonDHSandSNLdata.$

[WORK IN PROGRESS]

7.2 Additional robustness check: cell fixed effect panel

We will implement a panel analysis at the grid cell level, with information on its upstream-downstream status. This has the advantage of dealing with the DHS 2-10 km random reshuffling issue, as we will take the average infant mortality by cell and no longer rely on individual-level analysis.

[WORK IN PROGRESS]

8 Conclusion

This paper identifies a negative externality of industrial mining on local population living standards, as we show that industrial mining sites increase infant mortality in surrounding villages, indirectly through the contamination of water resources. We propose a staggered Difference-in-Difference strategy, where the treatment relies on an upstream-downstream topographic treatment. We isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison. We compare the effect on the health of villages located upstream and downstream of a mine deposit, before and after its opening. We find that being downstream when a mine opens increases by 2.3 percentage points (p.p) the 24-month mortality rate, which corresponds to an increase of 27% of the mortality rates. This result is an important contribution to the literature, as it enters into the debate on the positive and negative effects of industrial mining activity on health which finds a reduction in 12 months mortality rates using a treatment based on proximity to the mine [Benshaul-Tolonen, 2018]. First, the replication of a DiD analysis using geographical distance to the mine as a proxy for exposition to mining activity does not hold using our extended sample, which suggests the limited external validity of such a result. Second, our main result and analysis are robust to several robustness checks and underline the necessity to use a topographic treatment to understand the ambiguous effects of mining on living standards in developing countries and to be able to identify the negative externalities such as environmental pollution. Our main contribution is to be the first, to the extent of our knowledge, to identify the channel of water pollution, even if we identify it indirectly.

Further work for this paper will be to reproduce the main analysis for other health outcomes available within the DHS, such as looking at the effect on fertility, anthropometric measures, cough, diarrhea, anemia, and so on. Our main results find an increase in the 24-months mortality rate, but not for the 12-month mortality rate, suggesting that the absorption of contaminated water has delayed effects on biophysical development. Looking at other health outcomes would help understanding the process of pollution affecting child health in more details. A second on going work is use an instrumental variable strategy, using the international price of metal commodity as an IV for mine opening, as discussed in Section 3.2. We will also use production quantity outcomes available in the SNL database (but for a limited number of mines) to look at the heterogeneity of the effect according to the intensity of the pollution (that will be done also using the IV strategy). A large discussion on the broader magnitude effects and external validity of our analysis is also to be conducted. Further work can be done also to refine the definition of upstream and downstream positions using river networks and digital elevation maps, inspired by [Duflo and Pande, 2007] and [Garg et al., 2018], which could differientate the impacts on ground-water and on surface river water resources. First, we will control for the presence of river networks according to the magnitude of the river flow and importance, using the HydroSHEDS data on river basins. In a wokring paper,[Taylor, 2021] uses geological structure form gridded dataset of soil, intact regolith and sedimentary deposit thicknesses from [Pelletier et al., 2016] to build a global indicator of ground water potential (shallow bedrocks being correlated with the presence of aquifers), strategy that could be replicated as well in order to differientate between contamination of surface and ground water resources.

References

[Del, 2020] (2020). 2020 state of the artisanal and smallscale mining sector. washington, d.c.: World bank. Technical report, World Bank.

[Africa Bank, 2022] Africa Bank (2022). Mining industry prospects in africa. `https://blogs.afdb.org/fr/afdb-championing-inclusive-growth-across-africa/post`. Accessed: 2022-04-28.

[Ahlerup et al., 2020] Ahlerup, P., Baskaran, T., and Bigsten, A. (2020). Gold mining and education: A long-run resource curse in africa? *The Journal of Development Studies*, 56(9):1745–1762.

[Aragón and Rud, 2016] Aragón, F. M. and Rud, J. P. (2016). Polluting industries and agricultural productivity: Evidence from mining in ghana. *The Economic Journal*, 126(597):1980–2011.

[Atkin, 2016] Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

[Balietti et al., 2018] Balietti, A., Page, L., Pande, R., Rowe, K., and Sudarshan, A. (2018). Lease splitting and dirty entrants: The unintended deforestation consequences of india's environmental clearance process reform. *PEDL Research Papers.*

[Bazillier and Girard, 2020] Bazillier, R. and Girard, V. (2020). The gold digger and the machine. evidence on the distributive effect of the artisanal and industrial gold rushes in burkina faso. *Journal of Development Economics*, 143:102411.

[Benshaul-Tolonen, 2018] Benshaul-Tolonen, A. (2018). Local Industrial Shocks and Infant Mortality. *The Economic Journal*, 129(620):1561–1592.

[Berman et al., 2017] Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review*, 107(6):1564–1610.

[Borusyak et al., 2021] Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation.

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

[Briffa et al., 2020] Briffa, J., Sinagra, E., and R, B. (2020). Heavy metal pollution in the environment and their toxicological effects on humans. *Heliyon*, (6(9)):e04691.

[Callaway and Sant'Anna, 2019] Callaway, B. and Sant'Anna, P. H. C. (2019). Difference-in-differences with multiple time periods.

[Chen et al., 2013] 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.

[Chuhan-Pole et al., 2017] Chuhan-Pole, P., Dabalen, A. L., and Land, B. C. (2017). Mining in africa : Are local communities better off? *Africa Development Forum*, (Washington, DC: World Bank and Agence Francaise de developpement).

[Cobbina et al., 2013] Cobbina, Kumi, and Myilla (2013). Small scale gold mining and heavy metal pollution: Assessment of drinking water sources in datuku in the talensi-nabdam district. *International Journal of Scientific and Technology Research*, 2:96–100.

[Coelho and Texeira, 2011] Coelho, P. and Texeira, J. (2011). Mining activities: Health impacts. *Elsevier.*

[Corno and de Walque, 2012] Corno, L. and de Walque, D. (2012). Mines, migration and hiv/aids in southern africa. *Journal of African Economies*, 21.

[Cossa et al., 2022] Cossa, H., Dietler, D., Macete, E., Munguambe, K., Winkler, M., and Fink, G. (2022). Assessing the effects of mining projects on child health in sub-saharan africa: a multi-country analysis. *Globalization and Health*.

[Cust and Poelhekke, 2015] Cust, J. and Poelhekke, S. (2015). The local economic impacts of natural resource extraction. OxCarre Working Papers 156, Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

[de Chaisemartin and D'Haultfœuille, 2020] de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

[Dike et al., 2020] Dike, I., Onwurah, C., Uzodinma, U., and Onwurah, I. (2020). Evaluation of pb concentrations in selected vegetables and portable drinking water, and intelligent quotients of school children in ishiagu-a pb mining community: health risk assessment using predictive modelling. *Environ Monit Assess*, (192(2):126).

[Do et al., 2018] Do, Q.-T., Joshi, S., and Stolper, S. (2018). Can environmental policy reduce infant mortality? evidence from the ganga pollution cases. *Journal of Development Economics*, 133:306–325.

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

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

[Edwards et al., 2013] Edwards, D., Sloan, S., Weng, L., Dirks, P., Sayer, J., and Laurance, W. (2013). Mining and the african environment. *Conservation Letters*, 7.

[El-Kady and Abdel-Wahhab, 2018] El-Kady, A. A. and Abdel-Wahhab, M. A. (2018). Occurrence of trace metals in foodstuffs and their health impact. *Trends in Food Science and Technology*, 75:36–45.

[Garg et al., 2018] Garg, T., Hamilton, S. E., Hochard, J. P., Kresch, E. P., and Talbot, J. (2018). (not so) gently down the stream: River pollution and health in indonesia. *Journal of Environmental Economics and Management*, 92:35–53.

[Global Energy Monitor Wiki, 2021] Global Energy Monitor Wiki (2021). Heavy metals and coal. `https://www.gem.wiki/Heavy_metals_and_coal#cite_note-Toppin-3`. Accessed: 2022-04-28.

[Glöser et al., 2017] Glöser, S., Hartwig, J., Wheat, D., and Faulstich, M. (2017). The cobweb theorem and delays in adjusting supply in metals' markets. *System Dynamics Review*, 32.

[Goodman-Bacon, 2018] Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Working Paper 25018, National Bureau of Economic Research.

[Greenstone and Hanna, 2014] Greenstone, M. and Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, 104(10):3038–72.

[He et al., 2020] He, A., Li, X., Ai, Y., Li, X., Li, X., Zhang, Y., Gao, Y., Liu, B., Zhang, X., Zhang, M., Peng, L., Zhou, M., and Yu, H. (2020). Potentially toxic metals and the risk to children's health in a coal mining city: An investigation of soil and dust levels, bioaccessibility and blood lead levels. *Environ Int*, (141:105788).

[He and Perloff, 2016] He, G. and Perloff, J. M. (2016). Surface water quality and infant mortality in china. *Economic Development and Cultural Change*, 65(1):119–139.

[Jayachandran, 2009] Jayachandran, S. (2009). Air quality and early-life mortality : evidence from indonesia's wildfires. *Journal of Human Resources*, 44(4):pp. 3038–72.

[Khan et al., 2016] Khan, T., Nguyen, T., and Ohnsorge, Franziska an Schodde, R. (2016). From commodity discovery to production. *Policy Research Working Paper*, World Bank, Washington, DC(7823).

[Kotsadam and Tolonen, 2016] Kotsadam, A. and Tolonen, A. (2016). African mining, gender, and local employment. *World Development*, 83:325 – 339.

[Madilonga et al., 2021] Madilonga, R., Edokpayi, J., Volenzo, E., Durowoju, O., and Odiyo, J. (2021). Water quality assessment and evaluation of human health risk in mutangwi river, limpopo province, south africa. *Int J Environ Res Public Health*, (18(13):6765).

[Mamo et al., 2019] Mamo, N., Bhattacharyya, S., and Moradi, A. (2019). Intensive and extensive margins of mining and development: Evidence from sub-saharan africa. *Journal of Development Economics*, 139:28 – 49.

[Obasi et al., 2020] Obasi, N., Obasi, S., Nweze, E., Amadi, S., Aloke, C., and Aloh, G. (2020). Metal pollution and human health risk assessment of soils and edible plants in farmlands around enyigba lead-zinc mining site, ebonyi state, nigeria. *Environ Monit Assess*, (192(5):292).

[Oje et al., 2010] Oje, O., Uzoegwu, P., Onwurah, I., and Nwodo, U. (2010). Environmental pollution levels of lead and zinc in ishiagu and uburu communities of ebonyi state, nigeria. *Bull Environ Contam Toxicol*, (85(3):313-7).

[Pelletier et al., 2016] Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P., Niu, G., Willimas, Z., Brunke, M., and Gochis, D. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal Of Advances In Modeling Earth Systems*, pages 41–65.

[Programme, 2019] Programme, U. W. W. A. (2019). *The United Nations world water development report 2019: leaving no one behind*.

[Romero and Saavedra,] Romero, M. and Saavedra, S. The effects of gold mining on newborns' health (2017).

[Strobl and Strobl, 2011] Strobl, E. and Strobl, R. O. (2011). The distributional impact of large dams: Evidence from cropland productivity in africa. *Journal of Development Economics*, 96(2):432–450.

[Taylor, 2021] Taylor, C. (2021). Irrigation and climate change: Long-run adaptation and its externalities. *Working Paper*.

[Taylor et al., 2009] Taylor, C. D., Schulz, K. J., Doebrich, J. L., Orris, G., Denning, P., and Kirschbaum, M. J. (2009). Geology and nonfuel mineral deposits of africa and the middle east. *U.S. Geological Survey*.

[UN Environment Program , 2022] UN Environment Program (2022). Our work in africa. `https://www.unep.org/regions/africa`. Accessed: 2022-04-28.

[USAID, 2013] USAID (2013). Geographic displacement procedure and georefferenced data release policy for the demographic and health surveys. Dhs spatial analysis reports 7, United States Agency International Development.

[van der Ploeg, 2011] van der Ploeg, F. (2011). Natural resources: Curse or blessing? *Journal of Economic Literature*, 49(2):366–420.

[Venables, 2016] Venables, A. J. (2016). Using natural resources for development: Why has it proven so difficult? *Journal of Economic Perspectives*, 30(1):161–84.

[Von der Goltz and Barnwal, 2019] Von der Goltz, J. and Barnwal, P. (2019). Mines: The local wealth and health effects of mineral mining in developing countries. *Journal of Development Economics*, 139:1 – 16.

A Appendix

A.1 Descriptive Statistics

A.1.1 Data

Table 12 displays for each country the number and years of DHS waves, and the total number of DHS clusters and children under 5 years old, that we use for our empirical strategy. We see that the overall DHS sample gather 36 countries overall Africa, from 1986 to 2018. In our main empirical analysis, we decided to only keep DHS countries that had at least two survey rounds, in order to have comparable temporal variation across countries. Finally, our final sample accounts for countries (cf. Table 13): Tanzania, Burkina-Faso, Ghana, Zimbabwe, Mali, Democratic Republic of Congo, Guinea, Namibia, Madagascar, Cote d'Ivoire, Sierra Leone, Liberia, Nigeria, Senegal, Ethiopia, Uganda, Botswana, Malawi, Cameroon, Morocco, Niger, Kenya, Mauritania, Rwanda, Burundi, Lesotho, Togo, Eswatini, Algeria, Benin, Eritrea, Republic of the Congo, Guinea-Bissau, Somalia, Sudan, Tunisia, Djibouti, Equatorial Guinea (by order of importance in terms of mining activity according to Figure 15).

Table 12: DHS surveys in sample

Country	Survey Years	#Clusters	#Children<5Y
AO	2015	625	14177
BF	1993, 1999, 2003, 2010	1413	36744
BJ	1996, 2001, 2012, 2017	1752	31884
BU	2010, 2016	930	20824
CD	2007, 2013	836	27307
CF	1994	230	2639
CI	1994, 1998, 2012	674	12227
CM	1991, 2004, 2011, 2018	1619	31279
EG	1992, 1995, 2000, 2003, 2005, 2008, 2014	7741	75394
ET	2000, 2005, 2010, 2016	2313	42173
GA	2012	334	5911
GH	1993, 1998, 2003, 2008, 2014	2037	17931
GN	1999, 2005, 2012, 2018	1289	26588
KE	2003, 2008, 2014	2391	32235
KM	2012	252	3134
LB	1986, 2007, 2013	776	16224
LS	2004, 2009, 2014	1199	10269
MA	2003	480	6030
MD	1997, 2008	860	15932
ML	1996, 2001, 2006, 2012, 2018	1867	52996
MW	2000, 2004, 2010, 2015	2655	56688
MZ	2011	610	10950
NG	1990, 2003, 2008, 2013, 2018	3830	106848
NI	1992, 1998	503	11332
NM	2000, 2006, 2013	1290	13630
RW	2005, 2008, 2010, 2014	1176	21927
SL	2008, 2013	787	17483
SN	1993, 1997, 2005, 2010, 2012, 2014, 2015, 2016, 2017	2572	73084
SZ	2006	274	2706
TD	2014	624	18441
TG	1988, 1998, 2013	768	13869
TZ	1999, 2010, 2015	1259	20520
UG	2000, 2006, 2011, 2016	1765	37603
ZA	2017	671	3397
ZM	2007, 2013, 2018	1585	29105
ZW	1999, 2005, 2010, 2015	1431	19847

Table 13: DHS surveys in regression sample

Country	Survey Years	#Clusters	#Children<5Y
BF	1993, 1999, 2003, 2010	694	23,846
BJ	2001, 2012, 2017	62	1,911
BU	2010, 2016	317	8,280
CD	2007, 2013	82	5,092
CI	1994, 1998, 2012	196	4,838
CM	1991, 2004, 2011, 2018	90	2,513
ET	2000, 2005, 2010, 2016	100	2,956
GH	1993, 1998, 2003, 2008, 2014	1,217	12,074
GN	1999, 2005, 2012, 2018	360	11,775
KE	2003, 2008, 2014	233	4,130
LB	1986, 2007, 2013	190	7,537
LS	2004, 2009, 2014	336	2,810
MD	1997, 2008	131	3,301
ML	1996, 2001, 2006, 2012, 2018	570	19,147
MW	2000, 2004, 2010, 2015	207	6,651
NG	1990, 2003, 2008, 2013, 2018	105	3,993
NI	1992, 1998	40	1,105
NM	2000, 2006, 2013	138	2,175
RW	2005, 2008, 2010, 2014	713	14,615
SL	2008, 2013	377	13,717
SN	1993, 1997, 2005, 2010, 2012, 2014, 2015, 2016, 2017	363	10,111
TG	1988, 1998, 2013	104	2,187
TZ	1999, 2010, 2015	325	6,866
UG	2000, 2006, 2011, 2016	305	9,031
ZM	2007, 2013, 2018	364	10,966
ZW	1999, 2005, 2010, 2015	468	8,307

Tables 14, 15, 16, 17 and 18 display the descriptive statistics of all our outcome and control variables across six different samples: (1) all children under five year old living in the vicinity of an industrial mine, which correspond to our topographic analysis (2) same but restricted to children having reached the 12 months in order to calculate the 12 months mortality rate (3) same with 24 months. We then exclude in the samples (4), (5), (6) the children living in the mine's same subbasin. These descriptive figures are important in order to show that our analysis does not suffer from selection biases across the samples we use for our different regressions.

Table 14: Descriptive Statistics of mortality rates

	Mean (1)	SD (2)	Med (3)	Min (4)	Max (5)	N (6)
Living in the matched mine's subbasin						
All children	0.356	0.479	0	0	1	163,056
Children having reached 12months	0.356	0.479	0	0	1	128,317
Children having reached 24 month)	0.357	0.479	0	0	1	94,565
Mortality Rates						
Death <12m	0.064	0.245	0	0	1	128,317
Death <12m (outside mine's subbasin)	0.065	0.247	0	0	1	82, 609
Death <24m	0.084	0.278	0	0	1	95,565
Death <24m (outside mine's subbasin)	0.085	0.279	0	0	1	60,849
Death < 1m	0.032	0.176	0	0	1	161,374
Death < 1m (outside mine's subbasin)	0.032	0.176	0	0	1	103,920

Notes: We present the mortality rates at n months, conditionnally on having reached n months, for the whole sample and the sample living outside of the mine's same subbasin.

Table 15: Descriptive Statistics of timing and position relative to mines

	Mean (1)	SD (2)	Med (3)	Min (4)	Max (5)	N (6)
Open mine at birth						
All children	0.394	0.489	0	0	1	163,056
Children having reached 12 months	0.382	0.486	0	0	1	128,317
Children having reached 24 months	0.370	0.482	0	0	1	94,565
All children, outside mine's subbasin	0.372	0.483	0	0	1	104,996
Children having reached 12 months, outside mine's subbasin	0.361	0.480	0	0	1	82,609
Children having reached 24 months, outside mine's subbasin	0.346	0.476	0	0	1	60,849
Living downstream of a mine						
All children	0.458	0.498	0	0	1	163,056
Children having reached 12 months	0.459	0.498	0	0	1	128,317
Children having reached 24 months	0.459	0.459	0	0	1	94,565
All children, outside mine's subbasin	0.159	0.366	0	0	1	104,996
Children having reached 12 months, outside mine's subbasin	0.160	0.366	0	0	1	82,609
Children having reached 24 months, outside mine's subbasin	0.159	0.366	0	0	1	60,849
Distance from the matched mine						
All children	30.316	27.801	20.199	0.170	99.992	163,056
Children having reached 12 months	30.322	27.822	20.137	0.170	99.992	128,317
Children having reached 24 months	30.342	27.850	20.179	0.170	99.992	94,565
All children, outside mine's subbasin	43.478	26.516	36.126	1.076	99.992	104,996
Children having reached 12 months, outside mine's subbasin	43.491	26.547	36.146	1.076	99.992	82,609
Children having reached 24 months, outside mine's subbasin	43.543	26.567	36.239	1.076	99.992	60,849

Notes: We present the descriptive statistics within six different samples: (1) all children (2) children having reached 12 months (3) children having reached 24 months (4) all children not living in the mine's subbasin (5) children having reached 12 months and not living in the mine's subbasin (6) children having reached 24 months and not living in the mine's subbasin.

Table 16: Descriptive Statistics of children's characteristics (A)

	Mean (1)	SD (2)	Med (3)	Min (4)	Max (5)	N (6)
Birth order number						
All children	3.704	2.443	3	1	17	163,056
Children having reached 12 months	3.702	2.440	3	1	17	128,317
Children having reached 24 months	3.710	2.439	3	1	17	94,565
All children, outside mine's subbasin	3.723	2.436	3	1	17	104,996
Children having reached 12 months, outside mine's subbasin	3.723	2.434	3	1	17	82,609
Children having reached 24 months, outside mine's subbasin	3.725	2.428	3	1	17	60,849
Male						
All children	0.506	0.499	1	0	1	163,056
Children having reached 12 months	0.506	0.499	1	0	1	128,317
Children having reached 24 months	0.505	0.499	1	0	1	94,565
All children, outside mine's subbasin	0.508	0.499	1	0	1	104,996
Children having reached 12 months, outside mine's subbasin	0.508	0.499	1	0	1	82,609
Children having reached 24 months, outside mine's subbasin	0.509	0.499	1	0	1	60,849
Small size at birth						
All children	0.159	0.366	0	0	1	153,130
Children having reached 12 months	0.156	0.362	0	0	1	120,549
Children having reached 24 months	0.153	0.360	0	0	1	88,915
All children, outside mine's subbasin	0.160	0.366	0	0	1	98,410
Children having reached 12 months, outside mine's subbasin	0.156	0.363	0	0	1	77,334
Children having reached 24 months, outside mine's subbasin	0.155	0.361	0	0	1	59,957
Average size at birth						
All children	0.454	0.498	0	0	1	153,130
Children having reached 12 months	0.455	0.498	0	0	1	120,549
Children having reached 24 months	0.456	0.498	0	0	1	88,915
All children, outside mine's subbasin	0.456	0.498	0	0	1	98,410
Children having reached 12 months, outside mine's subbasin	0.456	0.498	0	0	1	77,334
Children having reached 24 months, outside mine's subbasin	0.455	0.498	0	0	1	59,957
Large size at birth						
All children	0.387	0.487	0	0	1	153,130
Children having reached 12 months	0.389	0.488	0	0	1	120,549
Children having reached 24 months	0.391	0.488	0	0	1	88,915
All children, outside mine's subbasin	0.384	0.486	0	0	1	98,410
Children having reached 12 months, outside mine's subbasin	0.388	0.498	0	0	1	77,334
Children having reached 24 months, outside mine's subbasin	0.390	0.488	0	0	1	59,957

Notes: We present the descriptive statistics within six different samples: (1) all children (2) children having reached 12 months (3) children having reached 24 months (4) all children not living in the mine's subbasin (5) children having reached 12 months and not living in the mine's subbasin (6) children having reached 24 months and not living in the mine's subbasin.

Table 17: Descriptive Statistics of children's characteristics (B)

	Mean (1)	SD (2)	Med (3)	Min (4)	Max (5)	N (6)
Cough in last 2 weeks						
All children	0.239	0.426	0	0	1	145,095
Children having reached 12 months	0.230	0.442	0	0	1	112,458
Children having reached 24 months	0.212	0.409	0	0	1	81,606
All children, outside mine's subbasin	0.237	0.426	0	0	1	93,337
Children having reached 12 months, outside mine's subbasin	0.229	0.420	0	0	1	72,291
Children having reached 24 months, outside mine's subbasin	0.210	0.404	0	0	1	52,424
Fever in last 2 weeks						
All children	0.252	0.434	0	0	1	145,034
Children having reached 12 months	0.248	0.432	0	0	1	112,421
Children having reached 24 months	0.221	0.415	0	0	1	81,584
All children, outside mine's subbasin	0.250	0.433	0	0	1	93,333
Children having reached 12 months, outside mine's subbasin	0.247	0.431	0	0	1	72,294
Children having reached 24 months, outside mine's subbasin	0.220	0.414	0	0	1	52,426
Diarrhea in last 2 weeks						
All children	0.166	0.372	0	0	1	146,097
Children having reached 12 months	0.161	0.368	0	0	1	113,251
Children having reached 24 months	0.123	0.329	0	0	1	82,199
All children, outside mine's subbasin	0.168	0.374	0	0	1	94,073
Children having reached 12 months, outside mine's subbasin	0.163	0.370	0	0	1	72,867
Children having reached 24 months, outside mine's subbasin 0.126	0.331	0	0	1	52,841	

Notes: We present the descriptive statistics within six different samples: (1) all children (2) children having reached 12 months (3) children having reached 24 months (4) all children not living in the mine's subbasin (5) children having reached 12 months and not living in the mine's subbasin (6) children having reached 24 months and not living in the mine's subbasin.

Table 18: Descriptive Statistics of mothers' characteristics

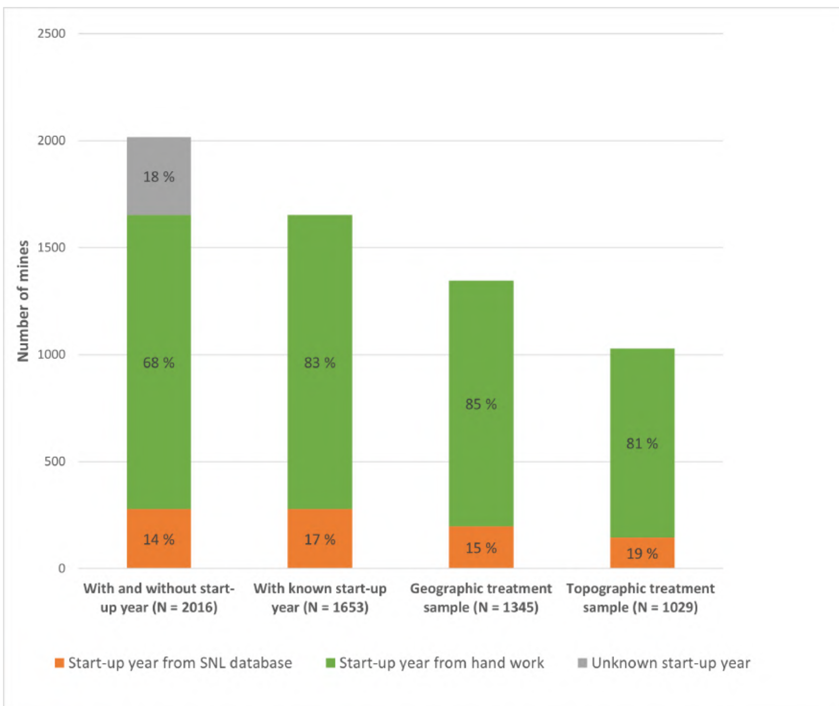
	Mean (1)	SD (2)	Med (3)	Min (4)	Max (5)	N (6)
Mother's age						
All children	29.088	7.027	28	15	49	163,056
Children having reached 12 months	29.546	6.982	29	15	49	128,317
Children having reached 24 months	30.033	6.940	29	15	49	94,565
All children, outside mine's subbasin	29.212	7.034	28	15	49	104,996
Children having reached 12 months, outside mine's subbasin	29.677	6.990	29	15	49	82,609
Children having reached 24 months, outside mine's subbasin	30.157	6.953	29	15	49	60,849
Years of education						
All children	3.419	4.088	1	0	22	162,992
Children having reached 12 months	3.402	4.077	1	0	22	128,265
Children having reached 24 months	3.364	4.066	1	0	22	94,524
All children, outside mine's subbasin	3.044	3.893	0	0	21	104,996
Children having reached 12 months, outside mine's subbasin	3.024	3.881	0	0	21	82,609
Children having reached 24 months, outside mine's subbasin	2.973	3.859	0	0	21	60,849
Migrant						
All children	0.600	0.490	1	0	1	101,039
Children having reached 12 months	0.602	0.490	1	0	1	79,472
Children having reached 24 months	0.604	0.489	1	0	1	58,339
All children, outside mine's subbasin	0.594	0.491	0	0	1	63,421
Children having reached 12 months, outside mine's subbasin	0.596	0.491	0	0	1	49,717
Children having reached 24 months, outside mine's subbasin	0.600	0.490	0	0	1	36,398
Urban						
All children	0.290	0.454	0	0	1	161,304
Children having reached 12 months	0.289	0.453	0	0	1	127,243
Children having reached 24 months	0.287	0.453	0	0	1	93,970
All children, outside mine's subbasin	0.261	0.439	0	0	1	104,489
Children having reached 12 months, outside mine's subbasin	0.260	0.439	0	0	1	82,282
Children having reached 24 months, outside mine's subbasin	0.257	0.437	0	0	1	60,659
Piped water as main drinking water source						
All children	0.256	0.436	0	0	1	163,056
Children having reached 12 months	0.257	0.438	0	0	1	128,317
Children having reached 24 months	0.256	0.437	0	0	1	94,565
All children, outside mine's subbasin	0.249	0.432	0	0	1	104,996
Children having reached 12 months, outside mine's subbasin	0.250	0.433	0	0	1	82,609
Children having reached 24 months, outside mine's subbasin	0.249	0.432	0	0	1	60,849
Visited health facility in last 12 months						
All children	0.607	0.489	1	0	1	145,018
Children having reached 12 months	0.591	0.492	1	0	1	114,557
Children having reached 24 months	0.585	0.493	1	0	1	89,956
All children, outside mine's subbasin	0.596	0.491	1	0	1	93,093
Children having reached 12 months, outside mine's subbasin	0.580	0.494	1	0	1	73,435
Children having reached 24 months, outside mine's subbasin	0.573	0.495	1	0	1	54,365
Ever had miscarriage						
All children	0.144	0.351	0	0	1	144,841
Children having reached 12 months	0.148	0.355	0	0	1	114,457
Children having reached 24 months)	0.152	0.359	0	0	1	85,001
All children, outside mine's subbasin	0.144	0.351	0	0	1	92,313
Children having reached 12 months, outside mine's subbasin	0.149	0.356	0	0	1	72,903
Children having reached 24 months, outside mine's subbasin	0.154	0.361	0	0	1	53,987

Notes: We present the descriptive statistics within six different samples: (1) all children (2) children having reached 12 months (3) children having reached 24 months (4) all children not living in the mine's subbasin (5) children having reached 12 months and not living in the mine's subbasin (6) children having reached 24 months and not living in the mine's subbasin.

A.1.2 Handwork

Out of the 3815 industrial mines recorded by the SNL database in Africa, 2016 were located within 100 km of a DHS cluster (with at least 2 waves of DHS). 278 had information on the opening and closing years within the database, and for the 1738 remaining mines, we searched for their years of opening and closure as well as their current activity status, i.e. whether the mining site looked active or inactive. The handwork consisted in reading the reports (comments and work history) available in the database, and browsing through the aerial images available on the SNL platform which provided the exact GPS coordinates and main location labels. This information was corroborated with online research (press releases, mining companies' websites, specialized websites on global mining activities, etc.) as well as Google maps and Google timelapse satellite imagery. Therefore, we were able to trace back to 1984 and check the start of the construction of a mining site and its expansion. A mine was noted as being still currently active if trucks could be seen around the site through Google maps satellite imagery. The exact startup year could not be determined for 18 percent of our sample (first bar of figure 12), and these mines are dropped in our regressions. In total, we, therefore, hand-checked 83% of the mines located within 100 km of a DHS cluster, and for which we know their year of opening (second bar of figure 12). We then associate each DHS cluster according to the strategy describe in section 4 and are left with the sample used for the geographic treatment (third bar of figure 12). Last but not least, we restrict our samples to mines that have DHS clusters within the three closest downstream subbasins, which corresponds to our topographic treatment sample.

Figure 12: Description of hand work and industrial mines samples



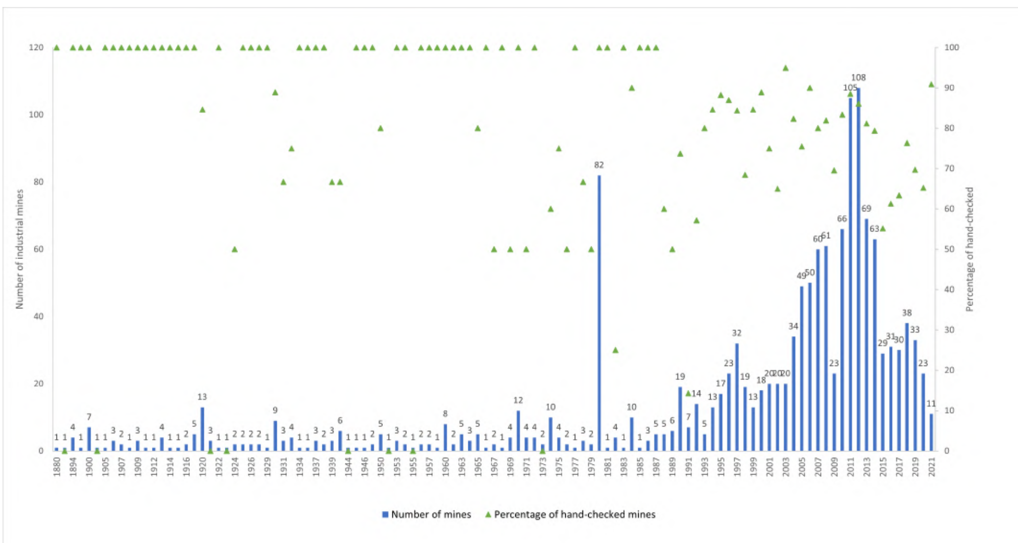
Notes : Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves (bar 1). Our regression analysis is conducted only on mines for which the startup year was available or could be retrieved by the handwork (bar 2). The geographic treatment consists of the mines which were matched according to the described methodology in section XX (bar 3). The topographic treatment sample consists of mines located within the three closest subbasins of DHS (bar 4).
Sources : authors' elaboration on DHS and SNL data.

The exact startup year could not be determined for 18 percent of our sample (first bar of figure 12), and these mines are dropped in our regressions. In total, we, therefore, hand-checked 83% of the mines

located within 100 km of a DHS cluster, and for which we know their year of opening (second bar of figure 12). We then associate each DHS cluster to its closest mine: and are left with the sample used for the geographic treatment (third bar of figure 12). Last but not least, we restrict our samples to mines that have DHS clusters within the three closest downstream subbasins, which corresponds to our topographic treatment sample. Among the sample of mines with startup year, 83.2 percent opened after 1981 which is the first year of birth within the DHS child surveys. For each of the following graphs, we study the whole sample of 2016 mines and plot the percentage of mines that were hand-checked and the percentage of mines that ends up having a startup year and are thus included in our study. We conduct this analysis on all the available mines within 100 km of a DHS cluster in order to be transparent on the creation of our sample compared to the original one.

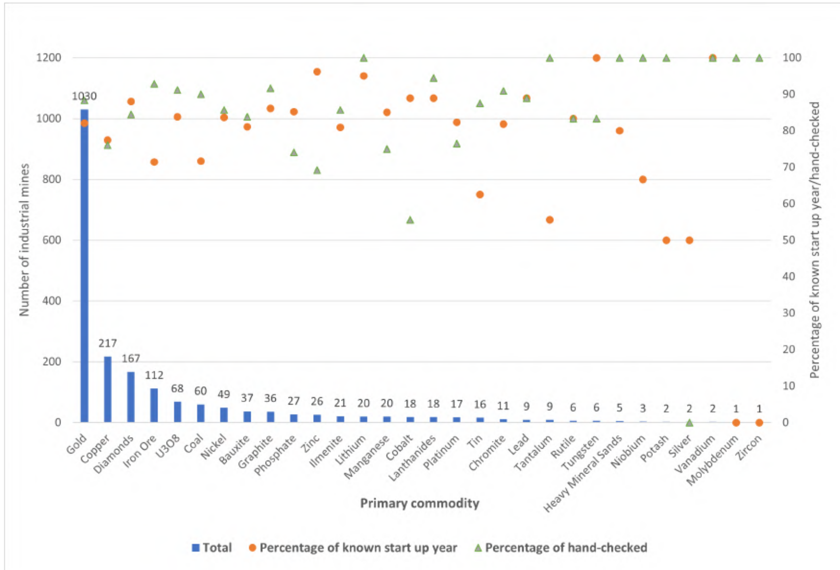
The distribution across each mining site's primary commodity of production can be found in Figure 14. Half of our sample consists in gold mining sites. Figure 15 and map XX [to be added] represent the distribution across country of location. Ownership information is available for 65 percent of our sample and the main owners are from the USA, UK, Canada, Australia, and China (Figure 17). 331 out of the 2016 mines have opening and closure dates. On average, industrial mining sites have opened during 18,3 years in this subsample, but the median is 8 years. Figure **??** represents the distribution across length of opening.

Figure 13: Mines across start-up year



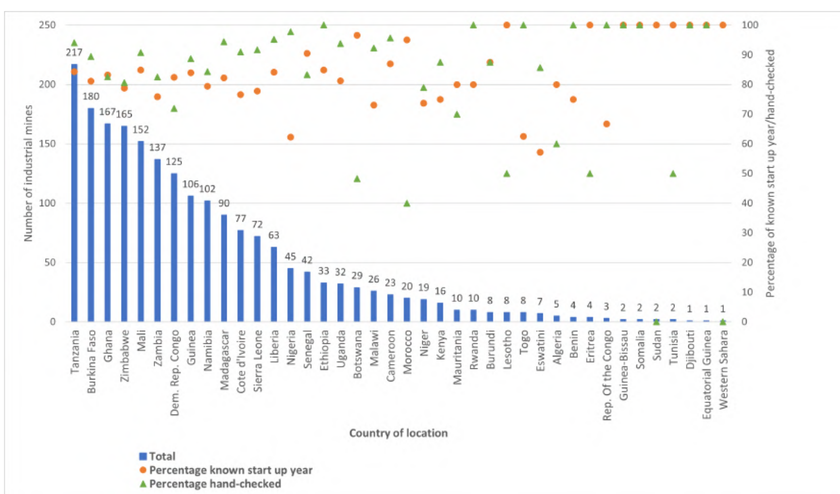
*Notes :*This figure displays the number of mines that opened during a specific year. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by hand work.

Figure 14: Mines across primary commodity



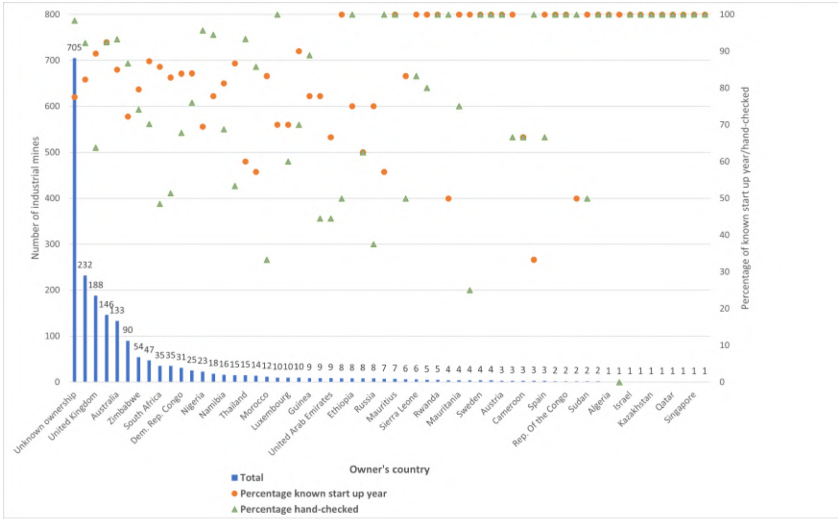
Notes :This graph displays the number of mines for each of the primary commodities available in the SNL database. Our main results are based on heavy metals and coal. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by handwork.

Figure 15: Mines across country of location



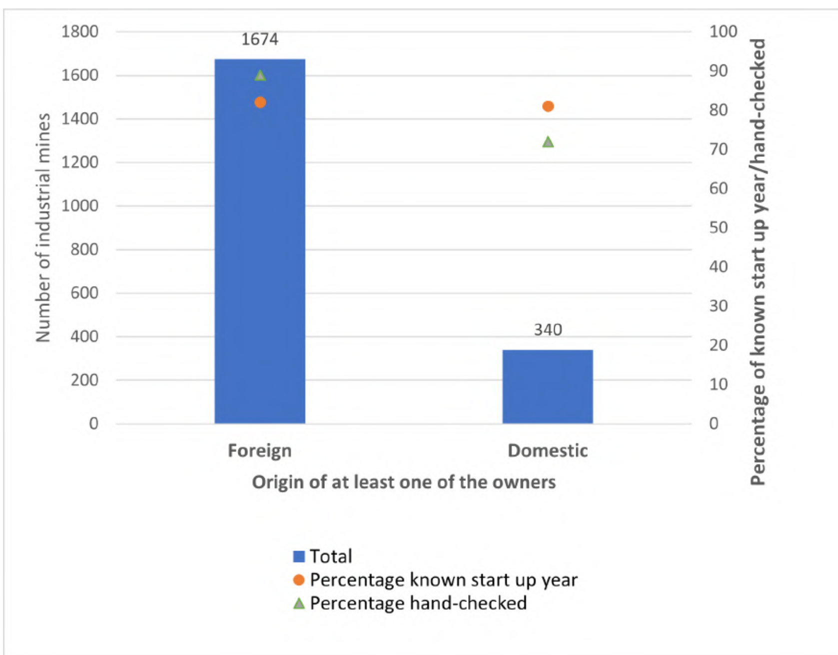
Notes :This graph displays the number of mines by country of location. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by handwork.

Figure 16: Mines across owner's country



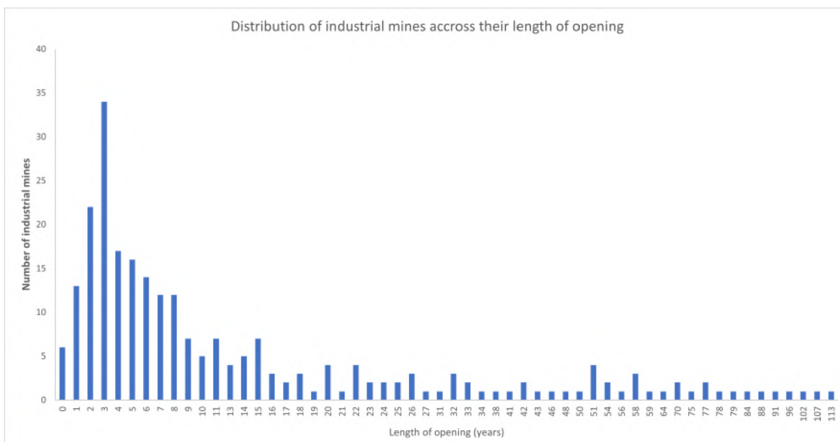
*Notes :*This graph displays the number of mines by owning company's registration country. Each mine can be owned by several companies from different countries, and we can thus attribute the same mine to several countries. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by handwork.

Figure 17: Mines across foreign and domestic ownership



*Notes :*This graph displays the number of mines across their domestic or foreign ownership. Each mine can be owned by several companies from different countries, and we consider domestic ownership if at least one of the owners is registered in the country where the mine is located. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by handwork.

Figure 18: Length of timing activity



Notes : This graph displays the average length of activity for the 331 mines for which we know both the start up and closure year. Our baseline sample is composed of the 2016 mines that are located within 100 km of a DHS cluster with at least two survey waves. We retrieve the information on the start up year for 1653 mines either from the SNL database or by handwork.

A.2 Context

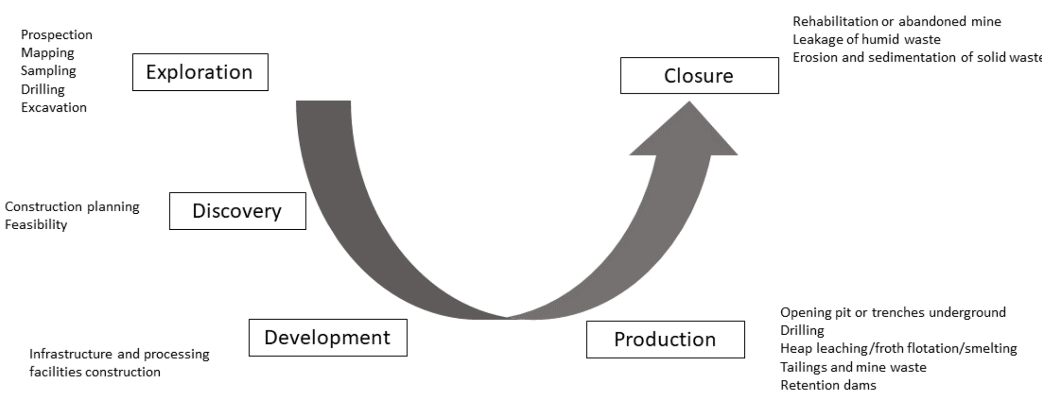
A.2.1 Mine Life Cycle

Figure 19: Expansion of the Sierra rutile plant, 1985-2020.

(a) Satellite image from 1985 (b) Satellite image from 2020

Notes : The two satellite images represent the expansion of the Sierra rutile plant, in Sierra Leone. Retention dams can be seen.
*Sources :*Google Earth engine Timelapse.

Figure 20: Industrial mine's life cycle



Notes : The figure schematizes the main stages of an industrial mining project .
Sources : Authors' elaboration, largely inspired by Coelho, Teixeira and Goncalves (2011)

A.2.2 Mine types

Table 19 gives the chemical properties of each metal, including their main chemical compounds (Column 1), their density (Column 2), and displays their share in the main estimation sample (in terms of

Table 19: Metals, chemical properties and sample distribution

Metals	Main chemical coumpounds (1)	density (gcm^{-3}) (2)	Nb. Mines (3)	Total Individual Sample (%) (4)
Heavy Metals				
Gold	Gold	19.3	581	41.88
Copper	Copper	8.96	89	5.03
Iron ore	Iron	7.87	54	8.72
U308	Uranium	8.39	36	1.60
Nickel	Nickel	8.9	25	5.06
Platinum	Platinum	21.45	21	0.43
Zinc	Zinc	7.14	19	2.46
Chromite	Iron Chromium	[4.5,5.09]	16	0.57
Ilmenite	titanium	4.6	14	3.67
Lanthanides	Lanthane(57) Lutecium(71)	[6.1,9.8]	13	1.95
Manganese	Manganese	7.21	12	0.62
Tin	Tin	[5.7;7.26]	10	4.87
Cobalt	Cobalt	8.9	7	0.56
Tungsten	Tugsten	19.25	6	1.06
Tantalum	Tantalum	16.69	5	0.15
Vanadium	Vanadium	6.12	4	0.04
Niobium	Niobium	8.57	3	0.39
Heavy Mineral Sands	Zirconium Titanium Tungsten Thorum	[4.5,17.6]	3	0.16
Silver	Silver	10.49	1	0.00
Lead	Lead	11.29	1	0.06
Non-Heavy Metals				
Diamonds	Carbon	3.5	115	11.73
Coal	Carbon Mercury? Arsenic?	1.35	55	2.19
Bauxite	Aluminium	2.79	23	1.94
Graphite	Carbon	2.26	21	0.82
Phosphate	Phosphate	1.83	14	2.78
Lithium	Lithium	0.53	14	0.80
Rutile	titanium	4.23	2	0.29
Potash(Salt)	Potassium	0.89	1	0.17

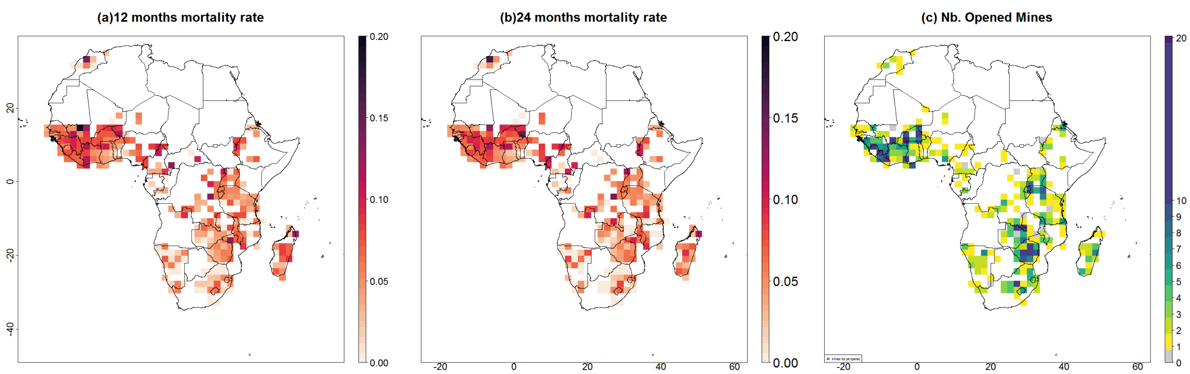
Notes:

the number of mines Column (3) and Total Individual Sample (Column (4)). Heavy metals are defined according to their density as being greater than $5gcm^{-3}$ [Briffa et al., 2020]. If small amounts of heavy metals can be mandatory, a high and abnormal concentration of heavy metals may cause health issues due to chronic toxicity. Heavy metals released in mining activity are toxic elements that degrade the environment and human biology. This is the case as well for heavy metals released during the mining and burning of coal, which is linked to toxic heavy metals such as lead, mercury, arsenic, nickel [Global Energy Monitor Wiki, 2021]. This is the reason why the main regression analysis includes heavy metals and coal mines, to capture the negative externalities linked to the most toxic mines. Table 29b from Section 29b shows the stability of the main result according to the sample selection of mines and including the whole sample.

A.3 Empirical Strategy

A.3.1 Identification Strategy

Figure 21: Outcomes spatial distribution



Notes : Figures (a) and (b) represent the means of 12/24 month mortality rates for each DHS waves available (listed in table 12), from 1986 to 2019. Means are computed at the grid level (100km mean size). The mortality rates are estimated without the children that did not reach 12/24 months at the time of the survey. Figure (c) displays the stock of mines that have opened before 2019 (including mines that opened before 1986). Means are computed at the grid level (100km mean size).
Sources : authors' elaboration on DHS and SNL data.

A.3.2 Descriptive Statistics and Parallel Trends

Table 20 gives the balance table for control variables, for the sample regressions both for 12 and 24 months mortality rates. Figures 22, 23 and 24 display the spatial variation of the main outcomes per period, for the restricted sample used in the main regression analysis.

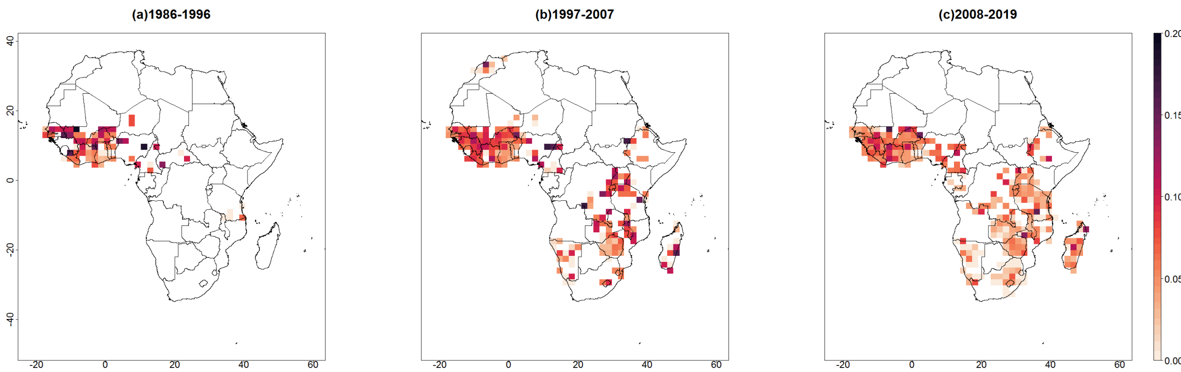
We plot in Figure 25 the distribution of mines opened within 100 km upstream or within the 3 closest subbasins downstream during a child's birthyear, so as to see which countries gather the highest number of industrial mining activity in the vicinity of surveyed households over 1986-2018. Ghana, Zimbabwe, Tanzania, Zambia, Guinea and Sierra Leone have the highest density of open mines nearby DHS clusters, while Benin, Burundi, Cameroon, Lesotho and Niger have the lowest number of open mining sites. This figure also represents the variation in the number of mines which opened between the first and last year of surveys for each country. We can thus grasp the context of change in industrial mining activity over our period of interest. Ghana, Tanzania, Guinea, Mali, and Burkina-Faso witness the highest number of mine openings between 1986 and 2018.

Table 20: Balance Table - Double Difference with Topographic Treatment - Descriptive Statistics

	Before Mine Opening					After Mine Opening					Within Up.	Within Dwn.	Within
	Upstream		Downstream		Diff	Upstream		Downstream.		Diff			
	N	Mean /(SD)	N	Mean /(SD)	(4-2) /(p.v)	N	Mean /(SD)	N	Mean /(SD)	(9-7) /(p.v)	(7-2) /(p.v)	(9-4) /(p.v)	(12-11) /(p.v)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Under 12 Children Sample													
Household Charateristics													
% Urban Household													
All	44725	0.274	8102	0.16	-0.113	24697	0.275	5085	0.26	-0.015	0.001	0.099	0.098
		(0.446)		(0.367)	(0)		(0.446)		(0.438)	(0.025)	(0.762)	(0)	(0)
Mines	279		235			204		186			104	63	10
Mother Charateristics													
Age													
All	44725	29.736	8102	29.585	-0.151	24697	29.678	5085	29.309	-0.37	-0.058	-0.276	-0.218
		(7.032)		(7.011)	(0.075)		(6.896)		(7.035)	(0.001)	(0.295)	(0.028)	(0.066)
Years of Education													
All	44725	2.383	8102	3.083	0.7	24697	3.82	5085	4.694	0.875	1.437	1.611	0.175
		(3.549)		(3.79)	(0)		(4.18)		(4.096)	(0)	(0)	(0)	(0)
% Migrant													
All	26125	0.598	5557	0.571	-0.026	14553	0.605	3482	0.586	-0.02	0.008	0.014	0.007
		(0.49)		(0.495)	(0)		(0.489)		(0.493)	(0.035)	(0.132)	(0.177)	(0.583)
Mines	249		187			157		138			68	35	7
Under 24 Children Sample													
Household Charateristics													
% Urban Household													
All	33688	0.272	6114	0.162	-0.11	17465	0.269	3582	0.269	0	-0.003	0.106	0.11
		(0.445)		(0.369)	(0)		(0.443)		(0.443)	(0.958)	(0.438)	(0)	(0)
Mines	279		234			192		172			92	49	9
Mother Charateristics													
Age													
All	33688	30.185	6114	29.924	-0.26	17465	30.233	3582	29.929	-0.304	0.048	0.005	-0.044
		(6.999)		(6.963)	(0.007)		(6.843)		(7.033)	(0.018)	(0.451)	(0.974)	(0.842)
Years of Education													
All	33688	2.347	6114	3.063	0.716	17465	3.795	3582	4.697	0.902	1.449	1.635	0.186
		(3.528)		(3.794)	(0)		(4.17)		(4.07)	(0)	(0)	(0)	(0.737)
% Migrant													
All	19223	0.602	4148	0.57	-0.032	10533	0.61	2494	0.591	-0.019	0.008	0.021	0.013
		(0.489)		(0.495)	(0)		(0.488)		(0.492)	(0.08)	(0.196)	(0.098)	(0.655)
Mines	248		186			150		133			61	30	6

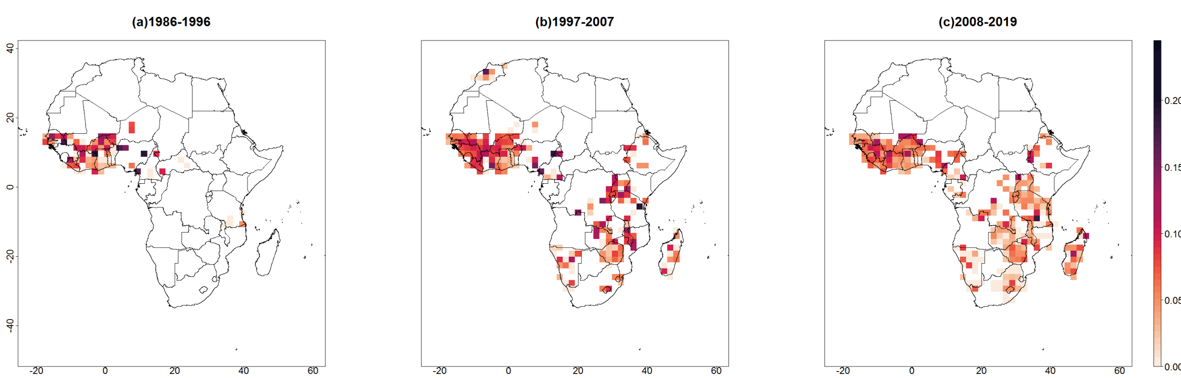
Notes: Standard errors and p-values in parentheses. Outcomes descriptive statistics of under 12 and 24 months mortality,for villages Upstream and Downstream mining sites, for individuals born before and after the opening of the mine.

Figure 22: Spatial variation of 12 month mortality rates per periods - Restricted Sample



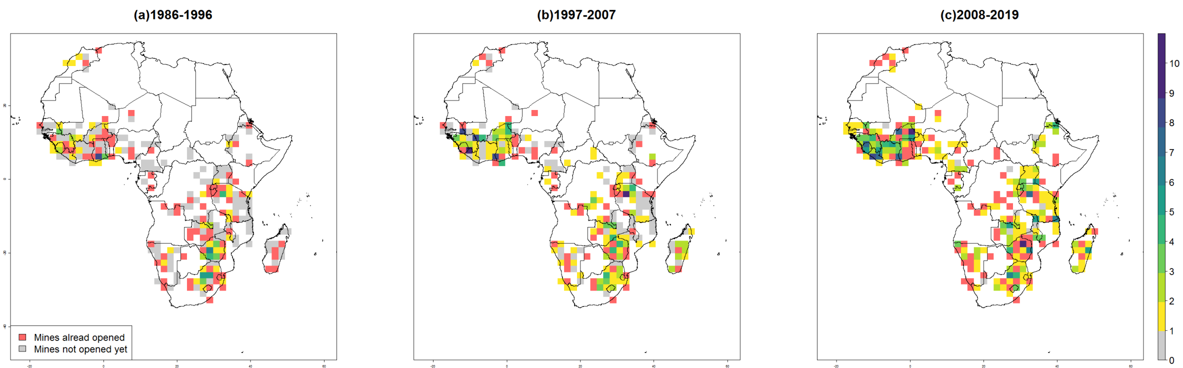
Notes : The figures represent the means of 12 month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008 and (c) 2008-2019, for the sample of the main analysis. The mortality rates are estimated without the children that did not reach 12 months at the time of the survey.
Sources : authors' elaboration on DHS data.

Figure 23: Spatial variation of 24 months mortality rates per periods - Restricted Sample



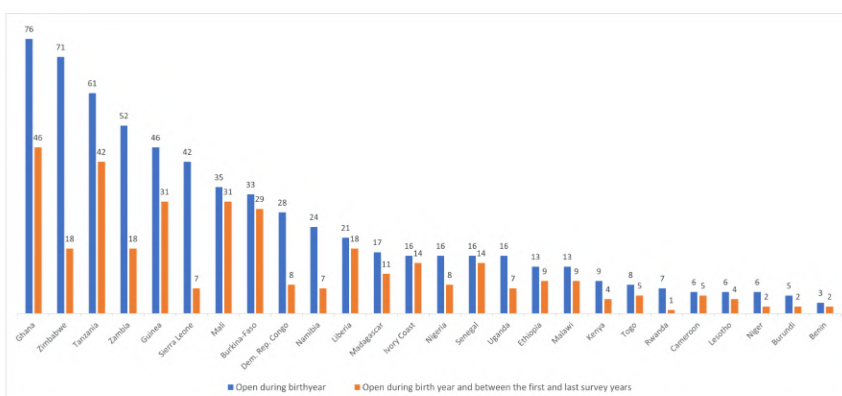
Notes : The figures represent the means of 24 month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008 and (c) 2008-2019, for the sample from the main analysis. The mortality rates are estimated without the children that did not reach 24 months at the time of the survey.
Sources : authors' elaboration on DHS data.

Figure 24: Spatial variation of mine opening per periods - Restricted Sample



Notes : The figures represent the number of mines that opened during the periods over the grid area (160km on average). A red grid represents an area where no mine opened over the period, but where at least one mine has opened before the period. A grey cell represents an area where no mine opened over the period, but where at least one mine will open in the future.
Sources : authors' elaboration on SNL data.

Figure 25: Number of open mines during birth year and between first and last wave



Notes : The figure represents the number of mines that were opened during the birth year of children located within our topographic treatment sample by country, and the number of mines that were opened during the birth year of children located within our topographic treatment sample and which opened between the first and last year of survey for each country.
Sources : authors' elaboration on SNL and DHS data.

A.4 Geographic Treatment

In this section, we propose to replicate the empirical strategy of [Benshaul-Tolonen, 2018], who finds that a mine opening is associated with a 5.5 percentage point decrease in 12 months mortality rates. The identification strategy relies on a treatment based on proximity, comparing individuals living nearby to those living further from an industrial mine. In this estimation, geographical proximity is used as a proxy for positive as well as negative externalities of industrial mining, such as exposure to mining pollution.
The identification strategy relies on a difference-in-difference strategy (DiD), comparing infant mortality in areas within 10 km of a mine deposit (treatment group) to infant mortality in DHS clusters further away from a mine deposit (10-100km, control group), before and after the opening of the mine deposit and within the district. As the strategy is a two-way fixed-effects, including a district-fixed effect, the comparison is made within this level of administrative delimitation. The identification can be formally written as:

$$\begin{aligned} Death_{i,v,c,m,SB} = &\alpha_0 + \alpha_1 Opened_{birthyear,i,v} + \alpha_2 MineDeposit_{[0;10km]v} \\ &+ \alpha_3 Opened_{birthyear,i,v} \times MineDeposit_{[0;10km]v} + \alpha_4 X_i \\ &\gamma_d + \gamma_{d-bthtrend} + \gamma_{c,birthyear} + \epsilon_v \end{aligned} \tag{3}$$

With $Death_{i,v,c,district}$ a dummy equals to one if child i from DHS village v (within district d) of country c, has reached the n^{th} month and has died (n being 12 for the 12 month old mortality, 24 and so on). $Opened_{birthyear,i,v}$ is a dummy equal to 1 if at least one mine located within 10 km for the treatment group, or within 100 km for the control group, has opened before child i's year of birth (this cohort comparison can be considered here as a source of tripe difference).$MineDeposit_{[0;10km]v}$ is a dummy of proximity (1 if village DHS v is within 10 km (or 100 km) of a mine deposit), X_i a vector of child/mother level controls (mother's age and age square, years of education, urban status). Finally, γ_d is a district fixed effect, $\gamma_{d-bthtrend}$ a district birthyear linear trend, and $\gamma_{c,birthyear}$ a country-birthyear fixed effect. Please note that the matching of DHS clusters to mines relies on the same strategy as in [Benshaul-Tolonen, 2018], and assigns a DHS cluster to the closest mine (without consideration of its opening status). Thanks to this matching, if a DHS cluster is both in the treatment and control groups of two different mines (i.e within 10km of mine A and within 10-100km from Mine B), we assign it mechanically to the treatment group (so linked to mine A). This creates bias explained in Section 4.1, which explains the choice for a district fixed effects and reduces the noise linked to DHS random displacements.

Firstly, we give our estimators from the exact replication of [Benshaul-Tolonen, 2018] results, using our own calculation, and find similar impacts (Tables 21 and 22). Second, we propose the replication of the results using our extended sample, including more countries, DHS waves, types of mines, and mines hand checked, and show the results from [Benshaul-Tolonen, 2018] is mainly determined by the choice of countries.

A.4.1 Exact replication of [Benshaul-Tolonen, 2018]

The geographic treatment proxies exposure to mining activity using the distance to the site and follows partly the analysis from [Benshaul-Tolonen, 2018], and finds contradictory impacts on infantile mortality. In order to understand better how our results can be compared to the literature, we propose in this section a replication exercise of the main result from [Benshaul-Tolonen, 2018]. The first difference between the two analysis is the sample, as [Benshaul-Tolonen, 2018] uses 43 gold mines that match with 31 DHS surveys from nine countries (Burkina Faso, Cote D'Ivoire, Ethiopia, Ghana, Guinea, Mali, Senegal, Tanzania, and DRC [18]. However, when matching the DHS cluster to the same industrial mining sites from [Benshaul-Tolonen, 2018], no DHS from DRC remained. At the end, the analysis is only on the 8 first countries, in accordance with Figure A6 from Appendix of [Benshaul-Tolonen, 2018]), for a whole sample of 1-year-old children of 48,151. In the main analysis of this paper, we pull all the available DHS in Sub-Saharan Africa that matches a total sample of more than 700 industrial mining sites. For this replication analysis, we used the same mines and DHS survey rounds as [Benshaul-Tolonen, 2018]. Please

[18] Please note that we used the same DHS survey that we found from the online replication code of [Benshaul-Tolonen, 2018]

note that we have few differences in terms of the whole sample, as [Benshaul-Tolonen, 2018] counts 37,365 children *vs* 41,902 for us, that might be explained by the way we calculated the 100km buffer distance [19]. The second main difference between both analyses is the independent variable, as we use as a shock the opening of the industrial mine whereas [Benshaul-Tolonen, 2018] uses the activity status based on production data given by the SNL product. This accounts for interim years, between the opening and final closing of the mine, where the production has been on hold. We replicated this variable for this exercise.

Table 21: Replication [Benshaul-Tolonen, 2018] Main Results

Dependent variable	Infant mortality first 12 months			
Sample :	Children (1)	Children drop spillover (2)	Boys (3)	Girls (4)
Industrial × mine deposit (at birth)	-0.0472**	-0.0474*	-0.0289	-0.0781***
	[0.0230]	[0.0260]	[0.0320]	[0.0301]
Mine deposit [0;10km]	0.0392**	0.0546***	0.0517**	0.0561**
	[0.0169]	[0.0195]	[0.0229]	[0.0231]
Mother's age	-0.0145***	-0.0154***	-0.0155***	-0.0152***
	[0.00190]	[0.00210]	[0.00274]	[0.00297]
Mothers's age × Mother's age	0.000222***	0.000236***	0.000223***	0.000245***
	[0.0000302]	[0.0000335]	[0.0000435]	[0.0000475]
Years edu.	-0.00214***	-0.00230***	-0.00272***	-0.00184**
	[0.000489]	[0.000547]	[0.000827]	[0.000760]
Urban$_h h$	-0.0125***	-0.0120**	-0.00710	-0.0183***
	[0.00428]	[0.00480]	[0.00687]	[0.00659]
Birth-month FE	Yes	Yes	Yes	Yes
Country birth year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
District BirthYear trend	Yes	Yes	Yes	Yes
Drop10-30 km away	No	Yes	Yes	Yes
Drop investment phase	No	Yes	Yes	Yes
Mean of outcome	0.102	0.104	0.110	0.099
Mean(treatment, pre-treatment)	0.154	0.163	0.173	0.153
Observations	41902	34228	17534	16694

Notes: $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from [Benshaul-Tolonen, 2018] and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth. Columns 2-5 drop the two years preceding th opening year, defined as investment phase in [Benshaul-Tolonen, 2018] and the individuals living within 10-30km of the closest industrial mine. Mean (treatment, pre-treatment) is the sample for the treatment group before the mine were active. dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

Table 21 displays the replication of the main results from [Benshaul-Tolonen, 2018] Table 2. We find that a mine opening within 10 kilometers is associated with a 4.7 percentage point decrease in infant mortality rates, while [Benshaul-Tolonen, 2018] found 5.5 p.p. Our results is slightly less significant than from [Benshaul-Tolonen, 2018], and we identify a different impact according to gender, with a significant reduction of girl mortality rates of 7 p.p *vs* a non-significant reduction for boys, which differs from the previous study. To follow [Benshaul-Tolonen, 2018] example, we excluded in Columns 2-5 from Table 21 individuals born within 10-30 kilometers of the closest industrial mining site and those born the two years before the opening of a mine, which is a proxy for the investment phase according to [Benshaul-Tolonen, 2018].
Please note that in accordance with the descriptive statistics from [Benshaul-Tolonen, 2018] we have in the sample a very high mean of under 12-month mortality rates (from 10 to 17 % according to the groups). These are relatively high numbers, that do not match with world bank data. This is because

[19]In the replication codes of [Benshaul-Tolonen, 2018], one can observe that the distance has been determined using the STATA command nearstat [...] dband(0,25) which relies on different projections (not specified) as ours from R libraries, explaining the small sample differences

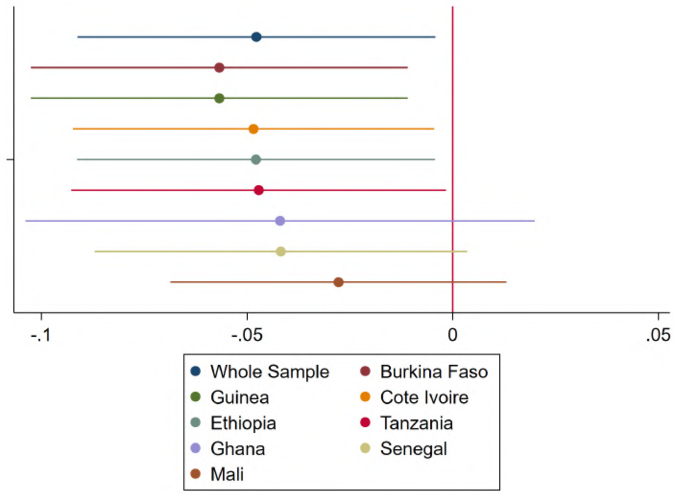
[Benshaul-Tolonen, 2018] drops all the individuals that are still alive but did not reach the age of 12 months yet to measure the mortality, in order to avoid growing mechanically the mortality rates of these cohorts. Unfortunately, we can read in the codes that if the living individuals were dropped, the children that died before their 12 months from these specific cohorts were not dropped: mechanically, the mortality rates for all the years preceding the survey rounds are 100 %, which explain the high mean of outcomes. For replication purposes, we propose to keep this variable and correct this in Table 22, where we observe average mortality rates around 7%. Figure 26 replicates the Figure A6 from [Benshaul-Tolonen, 2018], which shows the coefficient estimates of the main regression for *industrial* × *mine deposit* on infant mortality, each regression excluding the sample from one country as indicated by the country name. The Figure 26 shows that results are highly sensitive to the presence of Mali, Senegal, and Ghana in the sample (whereas they do not consist for the majority of the sample (5847, 1098 and 5595 respectively).

Table 22: Replication [Benshaul-Tolonen, 2018] Main Results

Dependent variable	Infant mortality first 12 months corrected			
Sample :	Children (1)	Children drop spillover (2)	Boys (3)	Girls (4)
Industrial × mine deposit (at birth)	-0.0494**	-0.0471*	-0.0439	-0.0631**
	[0.0229]	[0.0244]	[0.0317]	[0.0298]
Mine deposit [0;10km]	0.0394**	0.0587***	0.0682***	0.0513**
	[0.0179]	[0.0198]	[0.0255]	[0.0235]
Mother's age	-0.0118***	-0.0123***	-0.0120***	-0.0124***
	[0.00175]	[0.00196]	[0.00256]	[0.00283]
Mothers's age × Mother's age	0.000182***	0.000189***	0.000172***	0.000203***
	[0.0000279]	[0.0000312]	[0.0000405]	[0.0000452]
Years edu.	-0.00143***	-0.00152***	-0.00204***	-0.000803
	[0.000455]	[0.000510]	[0.000772]	[0.000715]
Urban$_h h$	-0.0106***	-0.0113***	-0.00501	-0.0196***
	[0.00384]	[0.00436]	[0.00661]	[0.00600]
Birth-month FE	Yes	Yes	Yes	Yes
Country birth year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
District BirthYear trend	Yes	Yes	Yes	Yes
Drop10-30 km away	No	Yes	Yes	Yes
Drop investment phase	No	Yes	Yes	Yes
Mean of outcome	0.079	0.080	0.083	0.077
Mean(treatment, pre-treatment)	0.109	0.118	0.120	0.115
Observations	40386	32873	16823	16050

Notes: $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from [Benshaul-Tolonen, 2018] and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth. Columns 2-5 drop the two years preceding th opening year, defined as investment phase in [Benshaul-Tolonen, 2018] and the individuals living within 10-30km of the closest industrial mine. Mean (treatment, pre-treatment) is the sample for the treatment group before the mine were active. dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

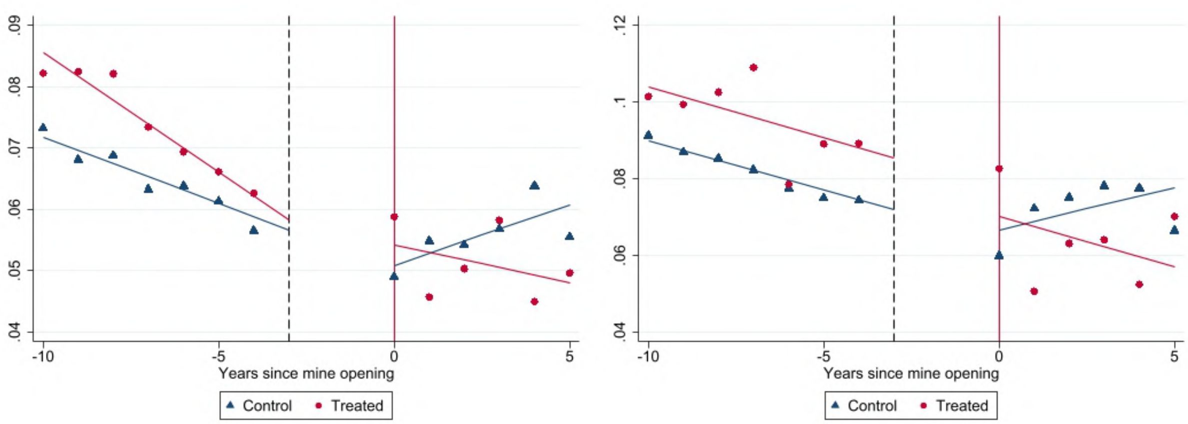
Figure 26: Regression results when dropping one country at the time



A.4.2 Full sample analysis

Figure 27 plots the linear trends of the 12 and 24 months mortality rates for the geographic treatment, including our overall mine and DHS sample. We see that the linear trends assumption seems to be validated for the 24-month mortality, but not for the 12-month mortality rates.

Figure 27: Linear Trends dropping investment phase - Geographic Treatment



(a) Infant mortality Rate 12 months

(b) Infant mortality Rate 24 months

Table 23 and Table 24 display the results, replicating [Benshaul-Tolonen, 2018] estimation strategy, with our overall sample of mines and DHS surveys. Table 23 focuses on the 12-month mortality rates and shows that we find a significant reduction of infantile mortality by 0.8 p.p only when controlling for migrants (column (2)). Columns (1) and (2) display the results for the whole sample, while columns (3) and (4) while dropping the spillovers effects (areas between [10-30]km and the tho years before the mine opening, which represents the investment phase in [Benshaul-Tolonen, 2018]). Columns (5) and (6) replicate the analysis for the male sample while columns (7) and (8) for the girls.

Table 24 displays result for the 12-month mortality rates (Columns (1)-(4)) and 24 Months mortality rates (Columns (5)-(8)), and compares the estimators when not including the migrant control variable (Columns (1), (3) (5) and (7)), and when including it (Columns ((2),(4),(6) and (8)). We also display the estimators for the restricted sample of rural areas (Columns (3),(4), (7), and (8)). We observe a significant reduction of 12 months mortality rates in Column (2), i.e for the overall sample while controlling for migrants, and find no results otherwise. This absence of results suggests that using proximity as a proxy for exposure to mining activity average contradictory effects, including both positive and negative externalities, and shows the importance of our main estimation strategy which relies on topographic position.

Table 23: Geographic Treatment

	Infant mortality first 12 months							
	All		Drop spillover		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Indus. × deposit	-0.00259	-0.00823**	-0.00189	-0.00575	0.00250	-0.00302	-0.00513	-0.00807
	[0.00329]	[0.00418]	[0.00407]	[0.00537]	[0.00570]	[0.00764]	[0.00522]	[0.00674]
Deposit	0.00130	0.00374	0.00103	-0.000128	0.00632	0.0113	-0.00366	-0.0109*
	[0.00252]	[0.00317]	[0.00392]	[0.00500]	[0.00546]	[0.00708]	[0.00513]	[0.00628]
Birth order	0.00389***	0.00315***	0.00360***	0.00320***	0.00349***	0.00304***	0.00382***	0.00356***
	[0.000345]	[0.000428]	[0.000423]	[0.000518]	[0.000606]	[0.000742]	[0.000549]	[0.000671]
Mother's age	-0.0105***	-0.0107***	-0.0102***	-0.0110***	-0.0116***	-0.0128***	-0.00884***	-0.00924***
	[0.000541]	[0.000668]	[0.000669]	[0.000824]	[0.000953]	[0.00119]	[0.000903]	[0.00111]
agesquare	0.000147***	0.000151***	0.000142***	0.000156***	0.000163***	0.000183***	0.000121***	0.000127***
	[0.00000853]	[0.0000106]	[0.0000106]	[0.0000131]	[0.0000150]	[0.0000187]	[0.0000142]	[0.0000175]
Years edu.	-0.000877***	-0.00103***	-0.000874***	-0.00101***	-0.000881***	-0.00103***	-0.000873***	-0.000968***
	[0.000135]	[0.000167]	[0.000164]	[0.000200]	[0.000238]	[0.000290]	[0.000216]	[0.000265]
Urban	-0.00610***	-0.00725***	-0.00708***	-0.00906***	-0.00825***	-0.0111***	-0.00563**	-0.00622**
	[0.00135]	[0.00172]	[0.00169]	[0.00214]	[0.00235]	[0.00297]	[0.00227]	[0.00289]
migrant		0.00543***		0.00509***		0.00255		0.00754***
		[0.00120]		[0.00145]		[0.00208]		[0.00196]
Constant	0.229***	0.232***	0.226***	0.240***	0.251***	0.273***	0.201***	0.206***
	[0.00826]	[0.0101]	[0.0103]	[0.0126]	[0.0146]	[0.0181]	[0.0138]	[0.0169]
Birth-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ctry-bthyr FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist-bthyr trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drop10-30 km	No	No	Yes	Yes	No	No	No	No
Drop t-2	No	No	Yes	Yes	No	No	No	No
N	359219	243645	236573	165202	119860	83570	116696	81601

Notes: $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from [Benshaul-Tolonen, 2018] and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth. dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

Table 24: Effects of industrial mining activity on under 12, 24 mortality - Geographic Treatment - All Households

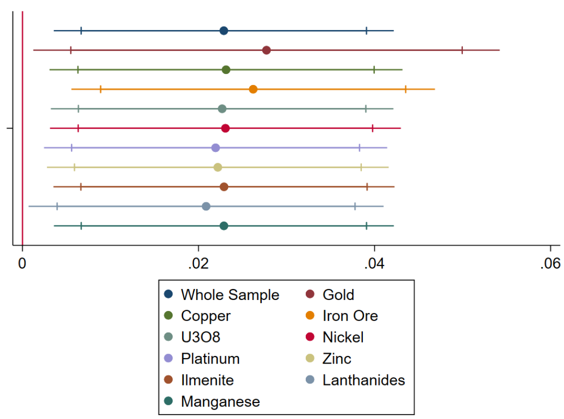
	Death <12m				Death < 24m			
	All	All	Rural	Rural	All	All	Rural	Rural
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
indus.×deposit	-0.00259	-0.00823**	-0.00259	-0.00627	0.000248	-0.00264	0.000248	-0.00248
	[0.00329]	[0.00418]	[0.00329]	[0.00509]	[0.00431]	[0.00535]	[0.00431]	[0.00657]
Deposit	0.00130	0.00374	0.00130	0.00313	0.000627	0.00121	0.000627	0.000859
	[0.00252]	[0.00317]	[0.00252]	[0.00368]	[0.00321]	[0.00411]	[0.00321]	[0.00477]
Indus.	0.00131	0.00222	0.00131	0.00340	0.00116	0.00122	0.00116	0.00190
	[0.00155]	[0.00200]	[0.00155]	[0.00230]	[0.00201]	[0.00259]	[0.00201]	[0.00297]
Birth order	0.00389***	0.00315***	0.00389***	0.00353***	0.00512***	0.00401***	0.00512***	0.00447***
	[0.000345]	[0.000428]	[0.000345]	[0.000500]	[0.000440]	[0.000549]	[0.000440]	[0.000642]
Mother's age	-0.0105***	-0.0107***	-0.0105***	-0.0116***	-0.0115***	-0.0124***	-0.0115***	-0.0140***
	[0.000541]	[0.000668]	[0.000541]	[0.000787]	[0.000704]	[0.000873]	[0.000704]	[0.00103]
Age square	0.000147***	0.000151***	0.000147***	0.000161***	0.000151***	0.000167***	0.000151***	0.000187***
	[0.00000853]	[0.0000106]	[0.00000853]	[0.0000122]	[0.0000110]	[0.0000136]	[0.0000110]	[0.0000159]
Years edu.	-0.000877***	-0.00103***	-0.000877***	-0.000792***	-0.00145***	-0.00157***	-0.00145***	-0.00132***
	[0.000135]	[0.000167]	[0.000135]	[0.000219]	[0.000173]	[0.000215]	[0.000173]	[0.000283]
Urban	-0.00610***	-0.00725***	-0.00610***		-0.00940***	-0.00995***	-0.00940***	
	[0.00135]	[0.00172]	[0.00135]		[0.00175]	[0.00222]	[0.00175]	
migrant		0.00543***		0.00514***		0.00727***		0.00630***
		[0.00120]		[0.00144]		[0.00155]		[0.00186]
Constant	0.229***	0.232***	0.229***	0.247***	0.273***	0.286***	0.273***	0.315***
	[0.00826]	[0.0101]	[0.00826]	[0.0120]	[0.0109]	[0.0134]	[0.0109]	[0.0159]
Birthmonth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cty-Bthyr FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SB FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SB/Bthyr trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Commodity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	359219	243645	359219	179155	265735	179729	265735	132398
R2	0.0195	0.0235	0.0195	0.0281	0.0289	0.0337	0.0289	0.0393
Mean	0.0630	0.0653	0.0630	0.0688	0.0816	0.0851	0.0816	0.0903

Notes: Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Proximity and Opened are dummies which indicate whether the individual lives in a DHS village within 10 km of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines.

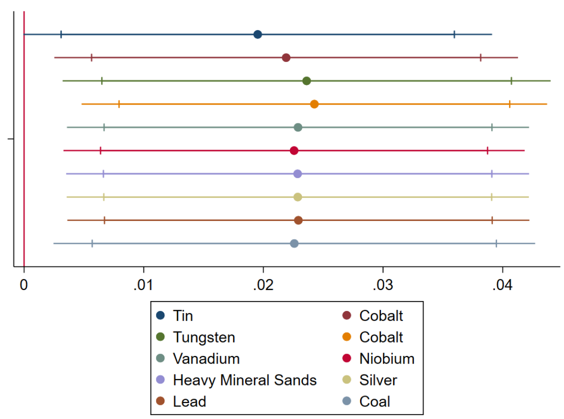
A.5 Robustness

Figure 28 displays the DiD estimators for different regression with restricted samples, meaning while dropping each metal one by one, using the sample for the 24 months mortality rates, and the heavy metals and coal mine sample. This suggests that our main results are not driven by a specific metal. Accordingly, Figure 29, plots the interaction estimators while dropping countries one by one and show that our analysis is not driven by a particular country.

Figure 28: Regression results when dropping one heavy metal one by one.

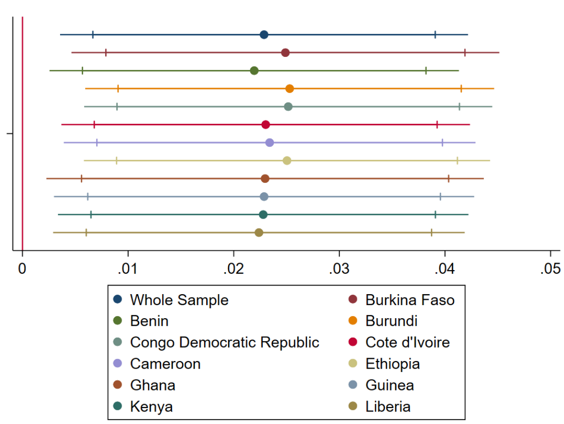


(a)

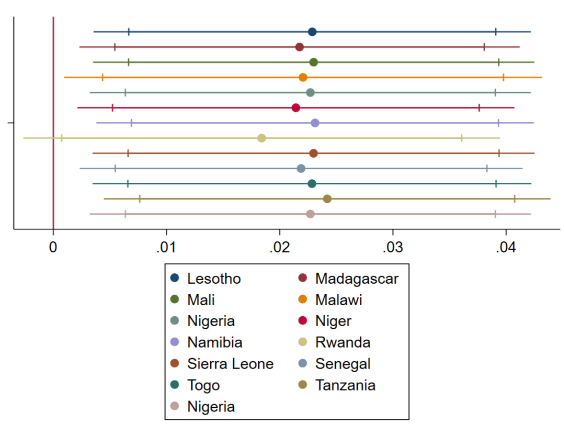


(b)

Figure 29: Regression results when dropping one country one by one.



(a)



(b)