Can White Elephants Kill? Unintended Consequences of Infrastructure Development in Peru

Antonella Bancalari [*]

November 2019

JOB MARKET PAPER

Abstract

Public infrastructure development is prone to inefficiencies that can result in poor-quality implementation, but the consequences are unclear. This paper studies the effect on infant and under-five mortality of a nation-wide expansion of sewerage infrastructure, conducted by the Government of Peru between 2005 and 2015. I use novel administrative panel data at the district level and exploit random geography-driven variation in project allocation to instrument for sewerage diffusion. I document an increase in under-five mortality in districts that experienced greater sewerage diffusion. The result is linked to hazards from the construction works and was exacerbated by delays and mid-construction abandonment. The potential health benefits of sewerage fail to manifest even after completion of projects due to lack of household connectivity.

JEL codes: C36, H51, I15, J18, N36, O18

Keywords: infrastructure, public expenditure, sewerage, mortality, Peru

---

[*] London School of Economics and Institute for Fiscal Studies (email:a.bancalari@lse.ac.uk). This project has benefited from data and discussions with individuals from numerous institutions in Peru, including the Ministry of Housing, Construction and Sanitation, Ministry of Health, Ministry of Economics and Finance and the World Bank's Sanitation Programme. I am grateful to Orazio Attanasio, Britta Augsburg, Alex Armand, Oriana Bandiera, Robin Burgess, Joan Costa-i-Font, Pramila Krishna, Berkay Ozcan and Simon Quinn for their useful comments. I also thank participants at CSAE Oxford, Royal Economic Society Junior Symposium, European Society for Population Economics and Population Association of America Conferences and at the LSE STICERD Economics and Social Policy, IFS-UCL Economics, Oxford University Economics, EUDN Development Economics, NOVAFRICA-NOVA SBE, Sussex Economics and the School of Economics at Kent. All work and opinions in this paper are my own and do not necessarily represent the views of the Government of Peru or other institutions. My work as a PhD student was supported by the London School of Economics Doctoral Scholarship.

1 Introduction

There is currently a global effort to achieve the Sustainable Development Goals (SDGs) by 2030. Crucial advances are being made in areas ranging from improving health and life expectancy and making human settlements safe and resilient to promoting inclusive and sustained economic growth. The assumption behind much of this policy agenda is that investing in infrastructure is crucial for attaining these development goals.

However, to date, key actors like low- and middle-income country (LMIC) governments and the World Bank have placed much more emphasis on the volume of infrastructure expenditure (World Bank, 2017), rather than the quality of that expenditure (Besley and Ghatak, 2006). Economic research has been very useful at identifying the effectiveness of infrastructure projects — e.g. dams, roads and electricity networks— once they are completed and in use (Duflo and Pande, 2007; Dinkelman, 2011; Lipscomb et al., 2013; Rud, 2012). What is less clear is the consequence of such projects while they are still underway.

This paper seeks to fill this gap in the literature. Specifically, I look at the development of sewerage projects. Sewerage lends itself to such analysis for two reasons. First, due to high fixed costs and economies of scale, the sewerage industry is a natural monopoly — i.e. one large provider, which is usually a government. Absent competition, monopolies have weak efficiency incentives that may result in overlooking implementation quality. Second, sewerage projects are highly disruptive because they entail extensive excavation, large building sites and traffic disruption. I look specifically at the effect of the diffusion of sewerage systems in Peru on the mortality of children under the age of five (under-five). Under-fives are the most vulnerable segment of the population because of their weak immune system and poor awareness of health risks.

The diffusion of sewerage in Peru is an excellent case to study because the scale of this public intervention was national, allowing for considerable spatial variation in implementation. The Government of Peru spent USD 3 billion to start more than 5,000 sewerage projects. I construct a district-level panel of 1,400 districts for every year between 2005 and 2015 by combining several sources of novel administrative data and grid-cell level spatial data. Specifically, I rely on detailed data on expense plans and timing of expenditures to identify the number of projects in construction and those completed in a given district.

Peru's natural geographic variation is ideal for the instrumental variable utilized in this study. I construct an instrument capturing how sewerage diffusion would have evolved over time had project placement been based solely on cost considerations. I rely on the fact that geographic characteristics — i.e. land slope, elevation, area and river density— affect a district's technical suitability for low-cost sewerage projects. Subject to a time-variant nation-wide budget constraint, I predict that the government of Peru would have allocated more projects to cheaper districts, and would have done so earlier in the period of study. The identification assumption is that no other factors affecting mortality rates (e.g. citizen's preference for preventive healthcare and other infrastructure and policies) changed over time along the same spatial lines as the predicted allocation of projects. A number of tests bolster the validity of my identification. I find that my instrument is not related to other types of infrastructure development, residential sorting, selective migration or geography-specific mortality trends.

I find that with every extra sewerage project that was launched, under-five mortality increased by 6 percent over the baseline. The primary determinants of this mortality increase are waterborne diseases and accidents. Notably, I find no effect of sewerage construction on the mortality caused by other diseases and complications unrelated to infections or hazards.

Furthermore, I find that delays and half-finished sewerage projects exacerbated the risks. Only half of the projects started were ever completed, and half of those took more than five years to be completed. With each extra year a district was exposed to sewerage works, the average annual change in under-five mortality increased by 14 percentage points.

One would expect the benefits of sewerage systems to manifest upon project completion. However, I find no effects on early-life mortality from an additional sewerage project being completed. In line with this finding, I document that providing access to public sewers does not increase the connectivity of households to sewerage systems. This find is evidence of the last mile problem — the inability of governments to connect costly infrastructure to the final user (Ashraf et al., 2016).

The paper makes three contributions. First, the paper broadens the literature on public goods by moving beyond assessing inefficiencies to encompass social costs. Influential papers have identified the determinants of waste in government spending (Bandiera et al., 2009; Rasul and Rogger, 2018) and institutional arrangements that prevent these inefficiencies (Besley and Burgess, 2002). However, there is a need to gain a better understanding of how inefficiencies in the provision of public goods jeopardise economic development and wellbeing. For example, Burgess et al. (2015) acknowledge this need in the context of a misallocation of public resources in Kenyan road building, by stating that “linking [our] findings to aggregate economic outcomes represents a key priority for future research”.

Second, this paper contributes to the literature on infrastructure effectiveness by extending the scope of analysis to the potential risks generated by projects that are still in progress. There is growing evidence in this literature on the effectiveness of electrification and large dams in improving labour and productivity (Dinkelman, 2011; Rud, 2012), and decreasing poverty (Dinkelman, 2011; Lipscomb et al., 2013; Duflo and Pande, 2007). More closely related papers find that environmental hazards from large infrastructure affect early-life mortality (Mettetal, 2019; Cesur et al., 2015; Gupta and Spears, 2017).

Finally, this study informs the literature on public health by exploring the effects of sewerage at scale in a contemporary setting (Watson, 2006; Alsan and Goldin, 2018). Recent studies in LMICs have mainly focused on the effectiveness of private sanitation infrastructure (Geruso and Spears, 2018), and come from experimental studies with a limited time-horizon and geographical setting (Duflo et al., 2015). My study, by contrast, focuses on a nation-wide setting and a longer temporal focus.

The rest of the paper proceeds as follows: section 2 provides background and section 3 explains the data and presents descriptive statistics. Section 4 provides details of the instrumental variable strategy. Section 5 and section 6 present the results of the effect of construction and completion of sewerage, respectively. Each of these sections describes the mechanisms driving the results. Section 7 concludes by discussing the significance of the study for a wider body of literature as well as potential extensions to other institutional contexts and other types of infrastructure.

2 Sewerage diffusion in Peru

Half of Peru's households lacked sewerage connectivity in 2005. To remedy this, the National Sanitation Plan for 2006-2015 set the target of increasing access to sewerage in urban areas, representing the first national goal of sewerage diffusion in Peru. In this period, the Government of Peru spent more than USD 3 billion to start 5,000 projects in more than 1,100 districts. The roll-out of sewerage projects across districts was not random. Starting sewerage projects depended on the demand of citizes and willigness of local municipalities; municipal resources; and willingness of the Central government to expand access in certain districts. Between 2005 and 2015, most projects were implemented by local municipalities: more than 56 were implemented by district municipalities and almost 30 percent by province municipalities (Appendix Figure 12). For projects implemented by district municipalities, unobservable characteristics of the district population (e.g. citizen's demand for public health) as well as the willingness and ability of the municipality to develop social infrastructure are correlated with both sewerage diffusion and early-life mortality. Furthermore, district municipalities can only implement sewerage projects if they are incorporated into the National System of Public Investment (SNIP), which requires having: (i) access to Internet; (ii) approval from the municipal council to receive technical assistance in formulation and implementation of investment projects from the Central government; and (iii) an annual budget above one million soles (approximately 200,000 sterling pounds). In line with these criteria, richer municipalities with a revenue above the median and with access to Internet by 2005 started a greater amount of sewerage projects (Appendix Figures 8 and 9). For the portfolio of projects formulated by the Ministry of Sanitation, the National Sanitation Plan 2006-2015 states that previously unattended and poor areas should be prioritized when expanding access to sewerage. This was not the case since more sewerage projects were started in districts with a lower percentage of the population with unmet basic needs and with a higher sewerage connectivity by 2005 (Appendix Figures 10 and 11). Interviews in the Ministry of Sanitation revealed that lobbying shifts the prioritization of projects into districts with greater political value. These confounding factors pose reverse causality and omitted variable bias concerns. Richer districts starting earlier and more sewerage projects also had lower early-life mortality by 2005 and experienced different trends in early-life mortality than poorer districts.

In order to address endogeneity in sewerage diffusion, I exploit exogenous variation linked to implementation costs and funds allocation. According to the SNIP guidelines, to implement a project, it requires achieving technical and economic viability. Both depend crucially on the design and direct project costs as cheaper projects are more likely to be implemented. The cheapest sewerage system is the conventional gravity system, connected to a treatment plant working through anaerobic digestion (Panamerican Center of Sanitation Engineering and Environmental Sciences, 2005). To be able to install this system, the area must have a steep gradient, be in a low altitude with enough oxygen and close to a body of water to discharge the effluent. More advanced and expensive technology is required in areas that do not have these characteristics. Projects in suitable districts fow low-cost sewerage systems are more likely to achieve technical and economic viability based solely on cost considerations. I exploit these geographical factors that affect sewerage diffusion to construct an instrumental variable, as explained in Section 4.

Sewerage diffusion also depends on funds allocation, as not all projects declared economic and technically viable have funds to be started and not all projects started have guaranteed resources to be completed. The largest source of funding was local: 40 percent of sewerage projects were financed by district royalties and 22 percent by local tax revenue. 30 percent of projects were funded by transfers from the national government (Appendix Figure 13). Confounding factors such as voters preferences and clientelism can affect the allocation of funds, so I rely on the nationwide expenditure as a component of my instrument. I explain the construction of the instrument in detail in Section 4.

3 Data and descriptive statistics

3.1 Data

I construct a district-level panel dataset of more than 1,400 districts in Peru from 2005 to 2015 by combining data from several novel sources. I construct infant and under five mortality using vital statistics registries and population forecasts. For the core dataset measuring sewerage diffusion, I compile and combine project-level data from viability studies and annual budget reports, which allows identifying when a project is under construction and completed. To construct the instrumental variable, I use grid-cell level spatial data, including elevation (from which I compute gradient), river flow and district boundaries. In addition, I draw on population forecasts to control for time-variant population density and district population size. The final dataset is an unbalanced panel of 1,408 districts spanning 2005-2015: 10,032 district-year observations for the infant-mortality sample and 10,494 district-year observations for the under-five mortality sample.

The outcome variables are constructed using vital records provided by the Ministry of Health and population forecasts built by the National Institute of Statistics and Informatics (INEI for its Spanish acronym) for every calendar year between 2005 and 2015 at the district level. The vital records provide the number of infants born alive and number of deaths of infants (under 1 year old) and children under 5 years old. The mortality data is disaggregated by cause of death following the International Classification of Diseases - ICD10. The population forecast provides data on the number of children under 5 years old. I construct infant mortality (IMR) and under-5 mortality (U5MR) rates for each district d and year t, using as the denominator the population at risk, as described by Preston et al. (2001):

$$IMR_{dt} = \frac{\text{Deaths infants aged 0-11 months}_{dt}}{\text{Live births}_{dt}} x1000$$

$$U5MR_{dt} = \frac{\text{Deaths children aged 0-59 months}_{dt}}{\text{Population aged 0-5 months}_{dt}} x1000$$

Infant-mortality rates in Peru have a right skewed distribution due to incomplete birth registration: the coverage was 93 percent by 2005 (UNICEF, 2005). To deal with outliers, I apply a winsorizing procedure to observations above the 90th percentile of the distribution of the infant mortality rate. To alleviate concerns linked to the quality of the vital registers in Peru, I compare the nation-wide trends of the computed infant mortality and under-five mortality rate with the trends of mortality rates drawn from several nationally representative surveys. Comparing the mortality rates computed from vital statistics with the rates from different surveys, I find that the former are slightly lower in level, but the trends do not differ greatly (See Appendix Figures 14 and 15).

To measure sewerage diffusion, I use raw data from viability studies registered in the National System of Public Investment (SNIP for its Spanish acronym) and budget reports from the Integrated System of Financial Administration (SIAF for its Spanish acronym) of the Ministry of Economy and Finance. These sources provide information on the number of sewerage projects declared viable between 2005 and 2015 in a given district and detailed project-level data on the budgeted investment and accrued investment by years. Using this information, I set as the starting year the one in which a given project receives the first disbursement. Because the Ministry of Sanitation does not keep a record of project completion, I follow their advice to set the year of completion as the one in which the budgeted investment is accrued by at least 90 percent. I set the years in which projects are under construction as the ones between start and completion. Projects without completion year but with start year are defined as in construction until the final year of the dataset.

I construct three alternative indicators of sewerage diffusion at the district level to identify effects not only once the infrastructure is completed, but also during its construction phase: (i) cumulative number of sewerage projects started; (ii) number of sewerage projects in construction; and (iii) cumulative number of sewerage projects completed. Indicators (i) and (iii) are constructed as cumulative given that sewerage infrastructure is a long-lasting investment whose access persists across years, entailing complementarities across systems. An important limitation is that sanitation projects are formulated in a sub-area of districts (the smallest jurisdictional level in Peru), but this is not easily identifiable (i.e. no address nor geo-codes) and there is no early-life mortality data at the same level. For projects formulated at a higher governmental level that lack of data on the number of projects per district, I assign one project to each district within the corresponding province or region. This approach is not capturing the intensity of sewerage diffusion within each of the districts, but it is done in only 3.7 percent of the districts ever intervened.

I use spatial data provided by the Ministry of Environment to compute geographic characteristics influencing the cost of sewerage development. I rely on this data to construct an instrumental variable. The spatial data includes information on surface elevation for multiple cells (1x1 km) which I match to district boundaries in 2015. I construct indicators for four main geographical characteristics: elevation, gradient, area and river density. First, I use the information on surface elevation at each cell to compute the fraction of district area in four different elevation categories considering quintiles of the elevation distribution: [0-250] meters above the level of the sea (henceforth mamls), {250-500] mamls, {500-1,000] mamls, and above 1,000 mamls. Second, I compute gradient using surface elevation at each cell and neighboring cells. I construct indicators capturing the fraction of district area falling into four gradient categories: (i) [0-0.8] percent, (ii) {0.8, 4.19] percent, (iii) {4.19-13] percent, and (v) above 13 percent. The first category captures flat areas below or equal 0.8 percent in which sewerage construction is costliest as determined by technical guidelines (Panamerican Center of Sanitation Engineering and Environmental Sciences, 2005). The remaining categories are created considering quintiles of the gradient distribution. I use quintiles because this ensures enough variation across categories, while allowing capturing differences in elevation and gradient within districts (compared to, say, using the mean per district). Appendix Figure 16 and Figure 17 shows districts in Peru vary largely in their ruggedness and altitude. Third, I compute the total area within the boundaries of each district. Finally, I compute river density as the fraction of the district area that falls in inland waters. Appendix Figure 18 shows that river density varies greatly across districts of Peru.

I draw on data from the National Register of Municipalities (RENAMU for its Spanish acronym) to measure municipal characteristics. As explained in Section 2 only districts that had access to Internet, high resources and approval to receive technical assistance were able to formulate and implement sewerage projects. I control for these characteristics as a robustness check. From RENAMU, I also get reports of whether water and faecal sludge is treated in the district. I use these variables to explore whether sewerage diffusion had any impact on the removal of bacteria and contaminants from the sources of drinking water and wastewater. Data on the treatment of water is available only between 2008 and 2014 and data on the treatment of sludge is available between 2006 and 2014.

Furthermore, to compute measures of sewerage connectivity, I compile household-level data from three Census rounds: 2005, 2010 and 2017. I use this data to evaluate if sewerage diffusion increased the percentage of households connected to the public sewers. I also use this data to compute the percentage of households with a head having attained above secondary school and connected to the electricity network in each district. These variables are alternative outcomes used to evaluate if sewerage diffusion affected early-life mortality rates through changes in the population composition (i.e. selective migration).

Finally, I compute measures of other infrastructure development that could have affected early-life mortality rates beyond sewerage diffusion. I use the SIAF budget reports from the Ministry of Economy and Finance to identify the level of expenditure on transportation, energy and health.

This data is available at the district level between 2007 and 2014 (2015 only available for transport expenditure).

3.2 Descriptive statistics

Figure 1 shows that sewerage diffusion happened at both the extensive and intensive margin. Between 2005 and 2015, more than 5,000 projects were started in almost 1,200 disricts: 80 percent of all districts were ever intervened. The majority of projects consisted in the installation of new systems (almost 80 percent), as opposed to the improvement of old pipe networks. On average, districts started two sewerage projects during the period of study and some districts started as many as 95 projects.

Table 1 provides descriptive statistics for the beginning and end period of analysis. The first and third columns provide the sum for the variables of interest and the mean for the geographical and control variables for 2005 and 2015, respectively. The second and fourth columns provide the standard deviation for the geographical and control variables for 2005 and 2015, respectively. The last column shows the data source used to compute the variables. Between 2005 and 2015, Peru started 6,090 sewerage projects, out of which 4,783 were construction and expansion of new systems and 1,307 were improvement of existing lines. In this decade, the national infant mortality rate decreased by almost a third from 20 to 7 per 1,000 births and the under-five mortality rate decreased from 3 to 2 per 1000 children.

According to the 2005 Peruvian Census, Peru had 1,830 districts belonging to 196 provinces and 25 regions. An average district had a population density of 642 people per km^2 in 2005: 23 thousand people living in an average territorial area of 636 km^2. Between 2005 and 2015 the population growth rate was 1.3 percent. The table also shows descriptive statistics of the key geographic factors influencing the cost of sewerage installation, revealing that there is great variation within and across districts in Peru along these lines. On average, the largest share of area of districts falls in the highest elevation category (74 percent), followed by the lowest category (15 percent) and all categories have a relatively high standard deviation (20 percent). Districts in the sample tend to have rugged terrains. On average, the lowest share of area falls in the flattest gradient category (only 10 percent) and the largest share in the steepest category (37 percent). River density is on average 53 kms per km^2 and there is great variation across districts (124 standard deviation).

In the decade of study, the average revenue of a district municipality quadrupled from 4 million to 15 million Nuevos Soles (approximately USD 4.5 million) and many municipalities gained access to Internet; the share of municipalities with access to Internet increased from 37 percent to 92 percent. In 2005, 66 percent of municipalities were registered as requiring technical assistance for the formulation of investment projects and 22 percent managed a health centre. The former decreased by 9 percentage points and the latter increased by 10 percentage points by 2015. In 2005, 11 percent of the district mayors were affiliated to the government's political party; this share remained similar by 2015.

Districts improved greatly their access to public services in the period of analysis. On average, the share of households with heads having completed secondary education in a district increased by 12 percentage points between 2005 and 2015. In 2005, on average, 56 percent of households were connected to electricity and 25 percent to sewerage systems and 23 percent of municipalities reported that sewage is treated in their district. The average sewerage and electricity connectivity increased by more than 20 percentage points and the percentage of municipalities reporting that sewage effluent is treated increased by 7 percentage points. Furthermore, public expenditure increased over the period of analysis in the transportation, energy and health sectors.

4 Empirical strategy

In order to understand the consequences of sewerage diffusion on early-life mortality, I rely on an instrumental variable approach. In Section 4.1 I explain how I construct the instrument, followed by Section 4.2 where I describe the estimation strategy.

4.1 Instrument: project allocation by technical suitability

The instrument I use is a prediction of how sewerage diffusion would have evolved over the decade of study had investments been based only on exogenous cost considerations. Exploting geographic characteristics, I rank all districts in Peru based on their technical suitability for low-cost sewerage projects and allocate projects following this ranking, subject to a nation-wide budget constraint and arbitrary maximum allocation threhold. The key identification assumption is that no other factors affecting mortality rates independently moved over time along the same spatial lines as the predicted allocation of projects. In other words, I assume that behavioural changes and the implementation of other health policies or social infrastructure that affects early-life mortality did not move from the most suitable districts for low-cost sewerage in early years to slightly less suitable districts in later years. My identification strategy ultimately relies on discontinuities created by a budget constraint and arbitrary threshold of maximum project allocation per district. This threshold leaves extra generation capacity that is subsequently reallocated to other districts further down the ranking. The intensity of the predicted sewerage diffusion varies across years and districts and this forms the basis of my instrumental variable strategy. Lipscomb, Mobarak and Barham (2013) demonstrate that isolating the variation in infrastructure linked to exogenous geographic cost and budget considerations is useful for studying the effects of large infrastructure projects.

Notably, relying on the technical suitability of a district makes the instrument comply with the monotonicity assumption. While the instrument may have no effect on sewerage diffusion in some districts, say very suitable district with low political will (never-takers) or unsuitable districts with high political will (always-takers), all districts affected by the instrument (compliers) are affected in the same way. In other words, all suitable districts predicted to receive more and earlier sewerage projects are more likely to implement more sewerage projects earlier. It is sensible to assume that no district decreased its likelihood of experiencing sewerage diffusion by being more technically suitable (defiers).

I follow three main steps to construct the instrument.

First: "Potential" nation-wide projects per year

For every year, I identify the number of sewerage projects that the government would have been able to start and complete. To do so, I divide the national expenditure on sewerage projects by the average cost of a project. The national expenditure on projects to construct new, expand and improve sewerage systems is identified based on the total disbursement made to all sewerage projects in a given year. The average cost of a sewerage project is calculated from the cost of all sewerage projects. The nation-wide budget for sewerage projects increased year to year and this generated variation over-time on the expenditure on sewerage projects. To get an idea of the over-time variation in projects "potentially" implemented, the budget spent in 2005 allows implementing 20 sewerage projects, in 2010 allows an additional 800 projects and in 2015 an additional 950 projects (see Figure 2).

Second: Ranking of districts based on technical suitability index

For each district, I compute an index that captures its technical suitability for implementing low-cost sewerage systems. Although sewerage diffusion is likely to respond mainly to demand-side factors, such as socio-economic characteristics and political will, it also responds to exogenous geographical factors.

The gradient of the terrain plays a major role in determining a district's suitability for low-cost projects. The cheapest sewerage system is the conventional gravity system, in which steepness allows faecal sludge to flow rapidly through pipes from houses to disposal areas (Romero Rojas, 2000). Fewer pipes and lower depths are required to install pipe networks in steeper districts, reducing even further the costs (Hammer, 1986). In very flat areas, it is necessary to install costly electric bombs to pump water and effluent (Panamerican Center of Sanitation Engineering and Environmental Sciences, 2005). Furthermore, elevation above the level of the sea is another topographic factor that affects districts' suitability for low-cost sewerage projects. The cheapest wastewater treatment plant works in low altitude areas because it requires oxygen to work through aerobic digestion (i.e. the biological decomposition of organic sludge (Romero Rojas, 2000). Sludge requires additional costly treatment (i.e. the injection of oxygen and chemicals) in areas with high altitude . The cost of sewerage projects also depends on the availability of water to discharge effluent. Factors linked to geographical dispersion also affect the district's technical suitability for sewerage and related costs. Considering that the span of settlements is greater in larger districts, developing sewerage systems in districts with large territorial areas requires installing longer networks of pipes. This increases both the complexity and cost of projects.

A regression of the total number of projects developed in a given district between 2005 and 2015 on the above-described geographic factors confirms the hypotheses raised by the engineering literature. I estimate the following ordinary least square (OLS) regression:

$$S_d = \sum_{k=2}^{4} \beta_{1k} Gr_{dk} + \sum_{k=2}^{4} \beta_{2k} E_{dk} + \beta_3 A_d + \beta_4 R_d + \epsilon_d \quad (1)$$

where S_d is total number of sewerage projects started in district d between 2005 and 2015, Gr_d is the fraction of area of district d falling in each of the three steep categories k (flat gradient is the reference category), E_d is the fraction of area of district d falling in each of the three elevated categories k (low altitude is the reference category), A_d is the total territorial area within district boundaries and R_d is the district's river density (river length in km per area in km^2).

Table 2 column (1) presents the OLS coefficients and standard errors in brackets and column (2) presents the standardized beta coefficients. Table 2 shows that, as predicted by the engineering literature, steep gradient categories and river density favour sewerage diffusion, while elevation and district area affects is negatively associated with project placement. The omitted gradient category is the fraction of district area in the flat category (below 0.8 percent) and the omitted elevation category is the fraction of district area in the low altitude category (below 250 mamsl). We can see that steep gradient and elevation predicts the allocation of sewerage projects non-monotonically: the largest coefficient is the lower-middle ({0.8, 4.19] percent) gradient category and the highest elevation category (above 1,000 mamls).

I compute a technical suitability index for all districts in Peru using principal component analysis, including all the above-described geographic factors. The computed index is the first component with an eigenvalue larger than 1. I rely on the index to rank all districts in Peru. The highest-ranking districts are forecasted to receive sewerage projects earlier and more across years.

Third: Allocation of projects based on ranking and budget constraint

The final phase consists on constructing a time-variant instrument. To do so, I allocate "potential" projects across districts and years following the technical suitability rank. I start by placing one project per district in the highest-ranking districts until the number of "potential" projects is exhausted. For instance, for 2005, I place one project for each of the 20 highest ranking districts because the budget spent that year amounts to the average cost of 20 projects. I follow the same procedure for the following years until a district receives a maximum of 5 projects, which is the median of the distribution of projects allocated in intervened districts between 2005 and 2015. Projects that would have been allocated to higher-ranked districts that already hit the maximum are placed in lower-ranked districts. Therefore, by 2015, the highest-ranked districts would have received up to 5 sewerage projects, while the lowest-ranked districts would have received none. This creates a predicted allocation roll-out that provides variation across districts and years.

Description of the instrumental variable

Figure 3 depicts actual district-wise sewerage diffusion (measured as started projects) between 2005 and 2015. There is great variation across districts and greater intensity in the affluent and populous north coast as well as in the poorer north-centre region of the Andes. The regions that experienced relatively lower sewerage diffusion are the northeast region of the Amazon and the south of Peru. Figure 4 plots the number of sewerage projects predicted in each district between 2005 and 2015. In 2005, the allocation of projects starts in the northeast Amazon region of Peru. Sewerage then diffuses into the northwest coast and from 2009 into the southwest coast and Andean region. By 2015, all 25 regions of Peru would have had at least one district intervened. Ignoring the demand-side drivers of sewerage diffusion forces the prediction to over-allocate sewerage projects to unattended places like the northeast Amazon area and the south coast. This weakens the relevance of the instrument, but allows extracting the exogenous variation linked to geographical characteristics. The spatial correlation between actual and predicted sewerage diffusion seems to be low when comparing Figure 3 and Figure 4. In fact, the correlation coefficient is 0.34. The strength of this correlation in a model with district fixed effects determines the predictive power of the instrumental variable estimator. I test formally the relevance of the instrument in the first-stage estimation explained in the next section. The identification assumption is that other factors affecting early-life mortality rates did not independently move over time along the same spatial lines as the predicted allocation of projects.

4.2 Estimation strategy

I estimate the effect of sewerage diffusion on infant mortality and under-five mortality rates between 2005 and 2015 relying on variation in the intensity of sewerage projects started and in construction across districts and years and using predicted sewerage diffusion as an instrument. The instrumental variable strategy corrects for the bias introduced by the endogenous placement of sewerage systems. To formally evaluate the relationship between actual and predicted sewerage diffusion, I estimate the following first-stage regression:

$$S_{dt} = \alpha Z_{dt} + \gamma_d + \delta_t + \tau P_{dt} + \nu_{dt} \tag{2}$$

where S_{dt} denotes the cumulative number of sewerage projects started (or number of projects in construction or completed) and Z_{dt} is the number of projects predicted to be started and completed in district d and year t. This first stage estimation attempts to isolate the portion of the variation in sewerage diffusion that is attributable to exogenous cost considerations.

I estimate the effect of sewerage diffusion on infant mortality and under-five mortality rates using the following two-stage least square (2SLS) model:

$$MR_{dt} = \alpha_2 \hat{S}_{dt} + \gamma_2 d + \delta_2 t + \tau_2 P_{dt} + \nu_2 dt \tag{3}$$

where MR_{dt} denotes infant ($1q_o$) or under-5 mortality ($5q_o$) rates and $\hat{S_{dt}}$ is the instrumented cumulative number of sewerage projects started (or number of projects in construction or completed) in district d and year t. Because my endogenous variable captures treatment intensity, there is more than one causal effect for a given district: the effect of going from 0 to 1 project, from 1 to 2 projects, and so on. The following underlying functional relation generates the counterfactuals:

$$MR_{dt} = f_{dt}(S) \tag{4}$$

Equation 4 indicates what the mortality rate of district d in year t would be for any number of sewerage projects S, and not just for the realized value S_{dt}. Because S_{dt} takes on values in the set $0, 1, 2, 3, S_{max}$, there are S_{max} causal effects. In this case, the 2SLS estimates are a weighted average of the unit causal response along the length of the potential nonlinear causal relation described by $f_{dt}(S)$. The unit causal response is the average difference in potential mortality rates for compliers at point S, that is, districts driven by the instrument to implement a number of sewerage projects less than S to at least S.

The estimation strategy includes both district γ_d and year δ_t fixed effects. The former controls for time-invariant characteristics in districts and the latter for annual shocks common to all districts. In addition, all models control for time-varying population density and total district population (P_{dt}). In some specifications, I add as covariates municipal characteristics that were correlated with actual sewerage diffusion (as discussed in Section 2), including indicators of whether the district municipality has access to Internet and needs technical assistance to formulate investment projects and municipal revenue to control for public investment capabilities. I also add an indicator of whether the municipality manages at least one health centre to control for political will on health policy. If the instrumental variable strategy is as good as random when allocating sewerage projects, I expect controlling for these factors to affect only slightly the point estimates. Standard errors are clustered at the district level to deal with serial correlation due to the panel characteristic of the data.

Table 3 that the predicted sewerage diffusion is a relevant instrument for actual sewerage diffusion. This table presents the first-stage and reduced-form estimates for the cumulative number of started sewerage projects (Panel A) and number of sewerage projects in construction (Panel B). Column (1) in Panel A examines the cross-sectional relationship between the technical suitability index and the total number of projects started by 2015. Column (2) and column (3) examines the over-time relationship between predicted and actual sewerage diffusion and restrict the analysis to the IMR and U5MR sample, respectively. The dependent variable in columns (1) to (3) is sewerage diffusion. The dependent variable in column (4) is IMR and column (5) is U5MR. The constructed instrument is a good predictor for actual sewerage diffusion. Using cross-sectional variation, a percentage point increase in the suitability index increases by 10 the total number of started projects and this is statistically significant at the 1 percent level (Panel A). Using variation across years and controlling for district and year fixed-effects lowers the magnitude of the first-stage coefficients, but they remain highly significant. On average, an extra project predicted to be allocated in a district increases by 0.45 the cumulative number of started projects (Panel A) and 0.3 the number of projects in construction (Panel B). The Sanderson-Windmeijer F test of excluded instruments are high and above the rule of thumb (10) for all specifications, which confirms the relevance of the instrument. Columns 4 and 5 show a positive effect of predicted sewerage diffusion on both infant and under-five mortality rates, although only the latter is precisely estimated. Importantly, when restricting the analysis to years before the start of the first project in construction in a given district, I find that the instrument has no effect on early-life mortality.

5 Effect of sewerage construction on early-life mortality

The main result of this paper is that mortality increased in districts that were exposed to more sewerage projects under construction. Table 4 presents the estimated effect of the cumulative number of started sewerage projects and the number of sewerage projects in construction on a district's infant mortality rate (henceforth IMR) and under-five mortality rate (henceforth U5MR). Columns (1) to (4) show OLS estimates and column (5) to (8) show 2SLS estimates. All specifications include district and year fixed-effects. While the OLS estimates suggest that sewerage diffusion is associated with a reduction in IMR (evidence of project placement bias), the 2SLS estimates are positive (though less precisely estimated). Both the OLS and 2SLS estimates show that sewerage diffusion increased U5MR. The 2SLS results remain robust when including municipal characteristics (columns 6 and 8). On average, an extra sewerage project started increased the IMR by 1 death per 1,000 births and the U5MR by 0.2 deaths by 1,000 children. These results are translated into a 1.7 percent and 4.1 percent increase, respectively, from initial average mortality rates. Only the effect on U5MR is precisely estimate (at a 1 percent significance level) likely because the quality of the data is better for this outcome. These unintended mortality consequences are linked to the construction works required to install sewerage lines. The magnitude of the effect of construction is significantly larger than the effect of start. On average, the IMR increased by 1.4 deaths per 1,000 births and the U5MR by 0.3 deaths per 1,000 children with each additional sewerage project in construction. These results are translated into a 2.4 percent and 6.2 percent increase from initial average mortality rates.

The effect of sewerage diffusion on under-five mortality is larger than on infant mortality because of the different population at risk in each mortality rate. While infants are mostly inside dwellings, children often roam freely outside and hence are more exposed to outdoor pollutants and hazards from construction works. Moreover, infants are more likely to be exclusively breastfed, and thus, not exposed to contaminants in drinking water.

Comparing across the OLS (columns 1-4) and 2SLS specifications (columns 5-8) we see that the 2SLS estimates are larger. There are three possible reasons for this downward bias in OLS estimates. First, the compliers in the IV strategy (based on district's technical suitability for sewerage systems) may be different from the average district whose placement of sewerage systems could have been affected by socio-economic and political considerations or other demand-side factors. Areas that experienced sewerage diffusion because of endogenous factors instead of low cost considerations may be more able to mitigate hazards linked to the installation of sewerage lines. Richer districts, better politically connected and with greater willingness to improve living standards may be able to mitigate better the side-effects of sewerage construction. Second, OLS estimates reveal the expected project placement bias since richer municipalities with lower mortality experienced greater diffusion. Finally, the sewerage diffusion variable constructed by combining administrative records likely suffers from classical measurement error, while the geographical variables used to predict the placement of sewerage projects are measured quite precisely (based on 1x1 km satellite maps). The 2SLS estimates may be correcting the measurement error in the independent variable and addressing the associated attenuation bias.

Ignoring heterogeneity driven by treatment intensity, I also find a positive effect of a district ever being intervened on early-life mortality rates (see Appendix section 8.1). A variety of sensitivy checks bolster the robustness of the main results. I estimate the 2SLS model with district and year fixed-effects (without municipal controls) with a series of modifications. First, I restrict the sample of analysis to districts that started at least one sewerage project to make the sample of study more comparable. Furthermore, I exclude the capital and main province of Peru, Lima, and add an indicator of whether the district is located in the Amazon region given that peculiar factors of these areas could be driving the results. Finally, I replace the independent variable with a version top-coded at the 90th percentile of the distribution of sewerage projects to ensure that the results are not driven by outliers. The magnitude and precision of the estimated effect of an extra started sewerage project as well as an extra sewerage project in construction on under-five mortality rates remain robust and highly significant (see Appendix section 8.2).

5.1 Validity of the instrument

To interpret the results as the causal effect of sewerage diffusion on early-life mortality, the exclusion restriction must hold. In other words, the predicted sewerage diffusion across districts and years must affect IMR and U5MR only through actual sewerage diffusion. In this section, I provide evidence that supports the internal validity of the results.

The main threat to my identification strategy is the delivery of other infrastructure that could affect early-life mortality. On the one hand, infrastructure is frequently developed as a bundle. The estimated results could be driven by other types of infrastructure that are developed following the same spatial and temporal pattern as my instrument if these also pose health hazards, such as pollution from roads and energy plants (Marcus, 2017; Gupta and Spears, 2017). On the other hand, the allocation of funds to develop sewerage may move away funds from infrastructure beneficial for mortality. My results could be explained by other types of infrastructure that are beneficial for early-life health, but developed following the opposite pattern to my instrument. To alleviate these concerns, I control for district expenditure on transportation, energy and health. Table 5 presents the estimates on the impact of sewerage diffusion on U5MR when progressively including the above-mentioned controls to check if the sewerage diffusion channel holds. This exercise confirms the main results: on average, an extra sewerage project started increased the U5MR. The magnitude of the estimate remains similar, even greater than the original estimate when controlling for transport and health expenditure. Controlling for energy expenditure decreases the precision of the estimated effect of sewerage diffusion on U5MR, but this could be attributable to a decrease in sample size.

I further explore if the alternative infrastructure investments could explain the direct effect of the instrument on early-life mortality. In other words, I test whether my instrument is a strong predictor of variation in other infrastructure expenditure, and if so, if the predicted variation can explain the increase in mortality rates. Table 6 presents 2SLS estimates of transport, energy and health expenditure on early-life mortality rates using the predicted sewerage diffusion as an instrument. Columns (1) to (3) show the estimated effect on IMR and (4) to (5) on U5MR. None of the three alternative infrastructure developments explains the estimated effects in mortality. No estimate is statistically significant and in all cases the first-stage is weak.

Another concern would be if the instrument is capturing variation driven by a specific region with greater suitability for low-cost sewerage projects. In Table 7, I test the robustness of the estimated effect of sewerage diffusion on U5MR when controlling for geography-specific trends. I include as controls the following components interacted with year: in column (1) the flat gradient category, column (2) the low elevation category, column (3) the district area in km^2, column (4) a categorical variable capturing the three main geographical regions of Peru (coast, highlands and jungle), and column (5) an indicator for the Amazon region. The estimated effect of sewerage diffusion on U5MR remains robust: controlling for geographic-specific trends has little effect on the first-stage power, the 2SLS point estimates and the statistical significance. When controlling for elevation-specific trends, the magnitude remains similar, but the precision and F-test of excluded instrument are lower. This finding reveals that elevation is an important driver of the variation used in the instrument.

Another threat to my identification strategy is if my instrument is correlated with the distribution of rural population across districts. Because the instrument is computed using geographic factors such as gradient and elevation that are likely to affect residential sorting, the results could be driven by channels other than sewerage diffusion. Flat and steep districts with greater river density may be beneficial for agriculture and attract households with farming as their main occupation. This sorting could explain the main results since rurality has long being associated with higher mortality rates (Hathi et al., 2017). Figure 5 shows that the computed instrument does not correlate with rural population: districts with a percentage of rural population above the median by 2005 where have an identical distribution of predicted sewerage projects as those with a percentage of rural population below the median.

5.2 Mechanisms

There are several explanations for the observed rise in infant and under-five mortality and I perform tests to shed lights on possible mechanisms.

I first investigate whether sewerage diffusion affected early-life mortality rates through migration. The observed increase in mortality rates could be a result of a decrease in the denominator, namely the number of live births (IMR denominator) and the number of children under 5 years old (U5MR denominator). This decrease in births and population could be due to families moving away from disruptive infrastructure works. Columns 1 and 2 in Table 8 show that this is not the case: the estimated effects on live births and under-5 population go in the opposite direction, meaning that coefficients of the effect of sewerage diffusion on early-life mortality are underestimated. The main results could also reflect selective emigration of the most well-off households and immigration of poorer households. Disruptive sewerage works may create incentives for well-off household to move away, reducing houseing prices and rent and hence attracting poorer households. Columns 3 and 4 show that there is no evidence of sorting across districts. The effect of sewerage diffusion on the number of people with completed secondary education is not statistically significant. Although there is a negative and statistically significant effect on electricity connectivity, the results are restricted to a small sub-sample because the data is only available for 50 percent of the districts of analysis and for 2 years.

5.2.1 Effects by cause of death

Under-five mortality may increase due to poor-quality implementation of sewerage infrastruture works. Open ditches from the excavation works required to install sewerage pipes pose a number of hazards to children. Environmental dangers documented in Peru are linked to dust particles, stagnated ground water that created sources of vector-borne diseases and the use of ditches as landfill sites (El Comercio, 2018). Shockingly, there is evidence of children falling and drowning in ditches from sewerage works that were as deep as two meters, got filled with water from nearby sources and had no security fence (Correo, 2018). Another important risk linked to open ditches is traffic diversion into previously quite residential areas. An interview with an engineer expert on the implementation of sewerage projects disclosed that contractors frequently divert traffic in an unorganized matter —not putting in place effective signalling systems and this results in greater traffic accidents. An additional example of precarious technical implementation is how contractors handle old pipe networks when expanding sewerage networks from systems already in place. Safely handling wastewater from old pipes is costly and thus contractors lack incentives to prevent faecal sludge from running into residential areas. An interview with the Leader of the World Banks Water and Sanitation Programme in Peru revealed that in areas where drinking piped-water supply is intermittent and in the presence of cracks in water pipes, sewage leaks can be absorbed by water pipes once water provision is resumed. Ingesting faecal matter through drinking water has fatal consequences for young children.

Table 9 investigates the effect of sewerage in construction on different measures of mortality depending on the diseases and related-health problems that caused the death. The mortality data is disaggregated for general pathology groups following the WHO International Classification of Diseases (ICD 10). The outcome in column (1) is all deaths caused by water-borne diseases, including: infectious diseases (ICD-10 category I), peri-natal complications (ICD-10 category XVI), diseases of the digestive system (ICD-10 category XI) and malnutrition and other nutritional deficiencies (ICD-10 category IV). The outcome in column (2) is the mortality rate linked to external causes (ICD-10 category XX), which mostly include deaths caused by falls, drowning and traffic-related accidents. The following columns estimates the effect of sewerage works on deaths unrelated to sanitation and construction works. The outcome in column (3) is the mortality rate resulting from diseases of the respiratory system (category X) and in column (4) is the mortality rate due to congenital malformations mortality rate (ICD-10 category XVII). The outcome in column (5) is the mortality rate linked to all unrelated factors, including diseases of the nervous system (ICD-10 category VI), circulatory system (ICD-10 category IX) and neoplasms (ICD-10 category II).

If sewerage diffusion affects mortality through failures during the construction period and my estimates are well identified, we would only observe an increase in the U5MR of deaths caused by infections and accidents. I line with this prediction, I find that an extra sewerage project in construction increased the U5MR caused by water-borne diseases by 0.2 deaths per 1,000 children (9.4 percent increase from the initial rate) and the U5MR caused by accidents by 0.12 deaths per 1,000 children (9.6 percent increase from the initial rate). Notably, I find no effect of sewerage diffusion on unrelated causes.

5.2.2 Delays and project non-completion

Delays and mid-construction abandonment of sewerage projects may exacerbate the hazards that the implementation phase pose to infants and children. Strikingly, between 2005 and 2014, only half of the projects that ever received funds were completed (see Figure 6). This reveals that a large share of started sewerage projects was left unfinished. Interviews with local engineers indicate that an average sewerage project takes one year to be completed, but delays are frequent. For projects started and completed between 2005 and 2015, the mode and median completion time is two years. From the pool of projects started after 2005, only 50 percent of the projects took less than 5 years to be completed (Figure 7). All together, this is evidence of frequent delays and high mid-construction abandonment of sewerage projects in Peru. To shed lights on the determinants of project non-completion and duration, I formally estimate a discrete-time hazard model of the probability of completing sewerage projects and the years that take to complete a project. Appendix 8.4 shows that political dynamics as well as project and municipal characteristics affect project completion and duration.

I formally evaluate the effect of the time exposed to the construction of sewerage systems on early-life mortality. To do so, I compute the total years that each district is exposed to construction works between 2005 and 2015. I use the maximum number of sewerage projects predicted to be developed in a given district as an instrument for years exposed to construction. The first-stage reveals that the technical suitability for low-cost projects also predicts districts being exposed to construction works for longer, even conditional on the number of started projects. This positive association could be explained by low-cost projects relying on less advanced technology (Panamerican Center of Sanitation Engineering and Environmental Sciences, 2005). Table 10 shows the 2SLS effect of years exposed to construction on the average annual change of IMR (columns 1 and 2) and U5MR (columns 3 and 4). On average, an extra year that a district was exposed to sewerage works increased the average annual growth of IMR by 14 percentage points and the average annual growth of U5MR by 16 percentage points. To disentangle the effects of the number of sewerage projects from the length of sewerage works, I restrict the sample to districts with only 2 sewerage projects developed (i.e. higher number of districts with this number of projects started and greater variation in works duration).

6 Effect of completed sewerage on early-life mortality

If mortality increases due to the construction works needed to install sewerage systems, we expect no effects after projects are completed. Early-life mortality rates can even decline as a result of having access to sewerage systems given the potential of these systems to prevent infectious diseases. To estimate the effect of sewerage completion on early-life mortality, I rely on the lagged values of the cumulative number of completed sewerage projects and I instrument it using the lagged values of the predicted placement of sewerage projects based on geographic factors. I use lagged values of sewerage completion because dwellings may take a couple of years to connect to the public sewers following the installation of the main sewerage line. In addition, using lagged values helps isolating the effect of functional sewerage systems since project completion is likely to happen while the construction of other sewerage projects is taking place. Table 11 presents the effect of the cumulative number of completed sewerage projects on a district's IMR and U5MR. The 2SLS estimates of the effects of an extra sewerage project completed on early-life mortality are not statistically significant. In line with my main hypothesis, early-life mortality increases during the construction phase and these unintended consequences dissipate once projects are completed (i.e. when closing open ditches).

6.1 Mechanisms

Even when projects are completed, the health benefits associated with sewerage systems may not materialize in the short-run given two main reasons. First, achieving less than universal connectivity means that neighbours are still contaminating the environment; the negative externalities from using rudimentary sanitation prone to leakages are at play (Augsburg and Rodríguez-Lesmes, 2018). Expanding access to sewerage systems may not ensure universal connectivity. Governments often do not guarantee the connection of expensive infrastructure to its final user, known as the last mile problem (Ashraf et al., 2016). Although the Peruvian norm establishes that it is compulsory for landlords to connect to public sewers when available, the enforcement of the norm is highly selective. This results in a skewed composition of connectivity rates towards those able to afford connecting their dwellings to public sewers or that have the willingness to do so (e.g.

tenants willing to pay higher rents, caring landlords and families better informed about sanitation risks and with preferences for health outcomes). Due to the negative health externalities linked to poor sanitation, achieving less than universal sewerage connectivity may not ensure improvements in the disease environment.

Second, even if universal connectivity is achieved, the sustainability of sewerage systems depends on the effectiveness of government agencies to operate and maintain the systems. A diagnose of the institutional quality of the public firms in charge of the operation and maintenance of sewerage systems in Peru revealed that more than 80 percent perform poorly, measured by its transparency, customer support, institutional management, financial and operational sustainability and work environment (Von Hesse, 2016). There is evidence that the bad performance of a public firm led to an inoperative treatment plant to contaminate local sources of water and agricultural fields and to deteriorate the disease environment (La Republica, 2015). A study revealed that in Latin American, particularly in Peru, only about 30 percent of wastewater is treated, with the remaining sludge being discharged in bodies of water used for drinking or irrigation purposes (Fay et al., 2017).

I then estimate the effects of sewerage diffusion on sewerage connectivity and the likelihood of treating water and sludge. Table 12 shows that using a two-year lag of the cumulative number of completed sewerage projects, connectivity decreased by 5 percentage points per project completed. A limitation of this approach is the completion of projects is highly correlated with the construction of other projects in the same district. To alleviate this issue, I estimate the effect of the number of total projects completed only after the last project is completed and I do not find statistically significant effects on household connectivity.

To investigate if sewerage diffusion improved the extent to which the population drinks safe water, I use municipal reports on whether piped water and sludge are treated. Table 12 shows that the probability of municipalities reporting that water is treated decreased by 0.07 percentage points with each additional completed project. Perhaps the installation of sewerage systems created a sense of protection that led municipalities to reallocated resources away from water treatment. There is no statistically significant effect on the probability of treating sludge, which is evidence of the bad quality of installed sewerage systems. Even when sewerage projects are completed, potential health benefits do not manifest if there is a sub-optimal operation and maintenance of these centrally-managed systems.

7 Conclusions

Large public infrastructure can be a driver of development and may be an effective way to set LMICs well on track to achieve the Sustainable Development Goals (SGDs). However, large infrastructure development can be highly disruptive and may result in negative unintended consequences. In this paper I examine the logic of this trade-off by asking the following empirical question: what are the consequences of infrastructure development? To answer this question, I focus on the diffusion of sewerage infrastructure, a natural monopoly with highly disruptive construction.

I provide evidence that under-five mortality increased with every additional project that was launched. The construction works exposed the population to hazards that increased early-life deaths caused by the incidence of waterborne diseases and accidents. Furthermore, I find that delays and mid-construction abandonment are highly prevalent, and exacerbated the negative consequences. Early-life mortality rates increased with each extra year that the district was exposed to unfinished sewerage projects. Finally, I find no effect from completed sewerage infrastructure projects on early-life mortality. Providing access to public sewers did not ensure that households would connect to this public good —evidence of the last mile problem.

My result that infrastructure development conducted by a monopolistic government can create considerable social costs requires wider investigation. Given that private participation has increased in the provision of public goods in Latin America and around the world, there is a need to understand if alternative arrangements can improve the quality of the implementation of infrastructure projects. Galiani et al. (2005) find large gains in connectivity and performance from the privatization of sewerage services in Argentina, which decreased child mortality by 5.3 percent over baseline. These estimates are exactly opposite to my estimated effect of developing sewerage projects on mortality. Such alternative institutional arrangement could therefore be a viable solution to solve the last-mile problem that the government of Peru is currently facing, and that many advanced economies faced in the previous century (Ashraf et al., 2016). Considering that each district in Peru constructed 2 sewerage projects on average, if sewerage services are privatized and the same gains as in Argentina are achieved, it would take 20 years to prevent the same number of deaths as the ones caused during the construction phase. Whether greater private participation in the implementation of sewerage projects can prevent child deaths or not is an empirical question yet to be answered. Besley and Ghatak (2006) predict that privatized solutions are only viable in the presence of strong legal systems and effective regulation. This prediction highlights the need to develop invisible infrastructure —i.e. institutions— when investing in visible infrastructure (Acemoglu and Dell, 2010).

My findings on the extent of mid-construction abandonment are in line with recent papers that explore the degree and determinants of this inefficiency (Rasul and Rogger, 2018; Williams, 2017). Overall, 13 percent of the expenditure in sewerage systems between 2005 and 2015 was allocated to projects that were never finished. A back-of-the-envelope calculation suggests that this waste equals one fifth of the expenditure in Education in Peru in 2015 (World Bank, 2016), which reflects the high social opportunity cost of project non-completion. While previous studies reveal that mid-construction abandonment of any infrastructure is a common waste in LMICs, I document that it is more than a waste: it can kill children. A key priority of future research is to identify which types of negative consequences can emerge from poor-quality implementation of other types of infrastructure. We must gain a better understanding of how dangerous ”white elephants” can be.

References

Acemoglu, D. and M. Dell (2010). Productivity Differences Between and Within Countries. *American Economic Journal: Macroeconomics 2*(1), 169–188.

Alsan, M. and C. Goldin (2018). Watersheds in child mortality: the role of effective water and sewerage infrastructure 1880 to 1915.

Ashraf, N., E. L. Glaeser, and G. A. M. Ponzetto (2016, may). Infrastructure, Incentives, and Institutions. *American Economic Review 106*(5), 77–82.

Augsburg, B. and P. A. Rodríguez-Lesmes (2018, jul). Sanitation and child health in India. *World Development 107*, 22–39.

Bandiera, O., A. Prat, and T. Valletti (2009, sep). Active and passive waste in government spending: Evidence from a policy experiment. *American Economic Review 99*(4), 1278–1308.

Besley, T. and R. Burgess (2002). The Political Economy of Government Responsiveness : Theory and Evidence from India Author (s): Timothy Besley and Robin Burgess Source : The Quarterly Journal of Economics , Vol . 117 , No . 4 (Nov ., 2002), pp . 1415-1451 Published by : The MIT Pres. *The Quarterly Jounral of Economics 117*(4), 1415–1451.

Besley, T. and M. Ghatak (2006). Public Goods and Economic Development. In A. V. Banerjee, R. Bénabou, and Dilip Mookherjee (Eds.), *Understanding Poverty*. Oxford Scholarship.

Cesur, R., E. Tekin, and A. Ulker (2015). Air Pollution and Infant Mortality: Evidence frmo the expansion of natural gas infrastructure.

Correo (2018, feb). Children drown in an open ditch from sewerage project.

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review 101*, 3078–3108.

Duflo, E., M. Greenstone, R. Guiteras, and T. Clasen (2015, sep). Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation.

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

El Comercio (2018, feb). Criadero de zancudos en Asia por obra paralizada.

Fay, M., A. Luis Alberto, C. Fox, U. Narloch, S. Straub, and M. Slawson (2017). Rethinking Infrastructure in Latin America and the Caribbean: Spending Better to Achieve More. Technical report, World Bank, Washington DC.

Galiani, S., P. Gertler, and E. Schargrodsky (2005). Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy 113*(1), 83–120.

Geruso, M. and D. Spears (2018). Neighborhood Sanitation and Infant Mortality. *American Economic Journal: Applied Economics 10*(2), 125–162.

Granados, C. and F. Sánchez (2014). Water Reforms, Decentralization and Child Mortality in Colombia, 19902005. *World Development 53*, 68–79.

Gupta, A. and D. Spears (2017, nov). Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households. *Journal of Environmental Economics and Management 86*, 262–276.

Hammer, M. (1986, jan). *Water and wastewater technology* (2nd ed.). John Wiley and Sons Inc.,New York, NY.

Hathi, P., S. Haque, L. Pant, D. Coffey, D. Spears, and C. Edu (2017). Place and Child Health: The Interaction of Population Density and Sanitation in Developing Countries. *Demography 54*, 337–360.

Kesztenbaum, L. and J.-L. Rosenthal (2016). Sewers' diffusion and the decline of mortality: The case of Paris, 18801914. *Journal of Urban Economics*, 1–13.

La Republica (2015, jun). Planta de Tratamiento de Epsel es un foco contaminante.

Lipscomb, M., A. M. Mobarak, and T. Barham (2013). Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics 5*(2), 200–231.

Marcus, M. (2017). On the road to recovery: Gasoline content regulations and child health. *Journal of Health Economics* (54), 98–123 Contents.

Mettetal, E. (2019, may). Irrigation dams, water and infant mortality: Evidence from South Africa. *Journal of Development Economics 138*, 17–40.

Panamerican Center of Sanitation Engineering and Environmental Sciences (2005). Guia para el Diseno de Tecnologias de Alcantarillado. Technical report, Organizacion Panamericana de la Salud (OPS/WHO) y Cooperacion Suiza para el Desarrollo (COSUDE), Lima, Peru.

Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography: measuring and modeling population processes*. Malden, MA: Blackwell Publishers.

Rasul, I. and D. Rogger (2018, feb). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *The Economic Journal 128*(608), 413–446.

Romero Rojas, J. A. (2000, aug). *Treatment of waste water: theory and principles of design*. Bogota.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from India . *Journal of Development Economics 97*, 352–367.

UNICEF (2005). State of the World's Children. Technical report, UNICEF, Geneva.

Von Hesse, M. (2016). Propuesta de Bases para una Política Nacional de Saneamiento. Technical report, Peruvian Ministry of Housing, Construction and Sanitation, Lima.

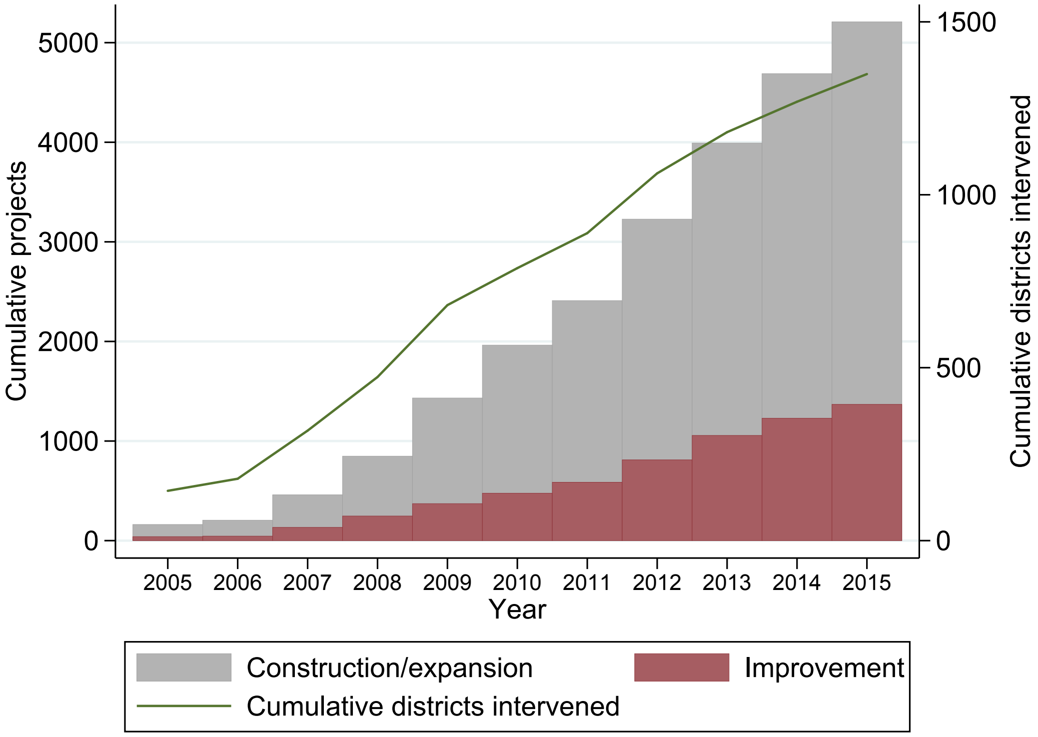
Watson, T. (2006, sep). Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations. *Journal of Public Economics 90*(8), 1537–1560.

Williams, M. J. (2017). The Political Economy of Unfinished Development Projects: Corruption, Clientelism, or Collective Choice? *American Political Science Review*.

World Bank (2016). World Development Indicators.

World Bank (2017). World Bank Lending (Fiscal 2017). Technical report, Washington D.C.

Figure 1: Number of started sewerage projects and districts intervened, 2005-2015



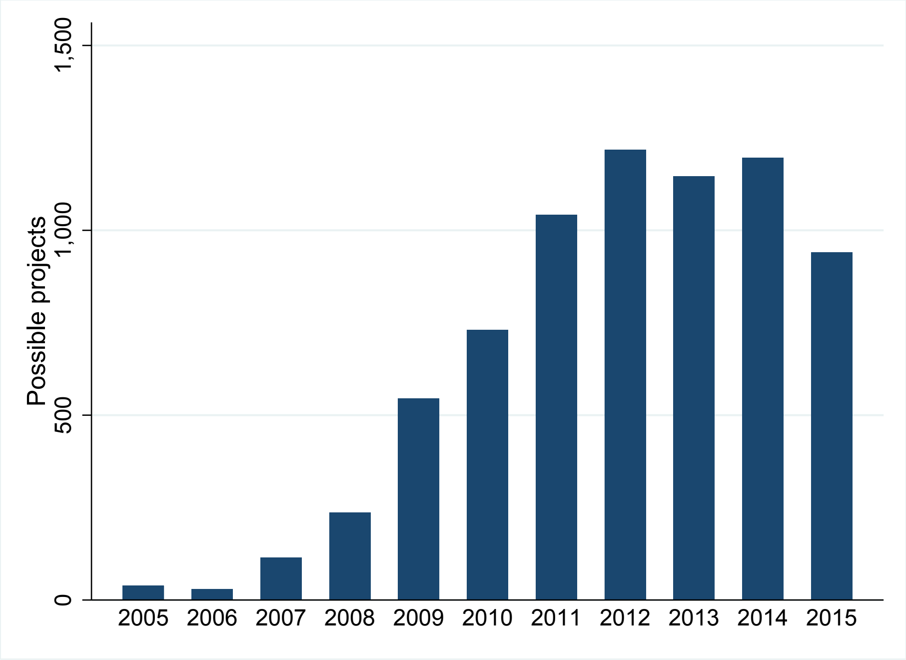
Note: This figure shows the cumulative number of started sewerage projects and districts intervened between 2005 and 2015. The y-axis on the left side indicates the cumulative number of started projects. The gray bars indicate the cumulative number of started projects for the construction of new and the expansion of sewerage systems. The lower red bars indicate the cumulative number of started projects for the improvement of existing sewerage systems. The y-axis on the right side indicates the cummulative number of districts intervened. A district is classified as intervened when at least one sewerage project was started. The green line indicates the cumulative number of districts intervened. Author's calculation using data from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Table 1: Summary statistics and data sources

	Beginning period		End period		Source
	(1)	(2)	(3)	(4)	(5)
	Sum		Sum		
1. Outcomes					
Deaths under 1y	6404		3820		Vital records
Births	325211		527088		Vital records
Deaths under 5y	8256		4987		Vital records
Population under 5y	2884748		2772323		INEI Pop forecast
Infant Mortality Rate	19.69		7.24		
Under-five Mortality Rate	2.86		1.79		
2. Sewerage diffusion					
Started sewerage	194		6090		SNIP and SIAF reports
Started sewerage (cons/exp)	156		4783		SNIP and SIAF reports
Started sewerage (imp)	38		1307		SNIP and SIAF reports
	Mean	SD	Mean	SD	
3. Geography					
Fraction district gradient $\leq$ 0.8%	0.10	0.23			GIS
Fraction district gradient {0.8-4.19]%	0.19	0.22			GIS
Fraction district gradient {4.19-13]%	0.34	0.20			GIS
Fraction district gradient above 13%	0.37	0.29			GIS
Fraction district elevation $\leq$ 250 mamls.	0.15	0.33			GIS
Fraction district elevation {250-500] mamls.	0.05	0.14			GIS
Fraction district elevation {500-1000] mamls.	0.06	0.15			GIS
Fraction district elevation above 1000 mamls.	0.74	0.41			GIS
District area (sq. km)	635.93	1655.50			GIS
River density (km/sq km)	53.32	124.30			GIS
4. Other					
Population density (pop/sq km)	642.91	2837.77	847.34	3188.96	Census and GIS
Population	23403.32	57020.49	32947.11	75973.03	Census
Municipal revenue (millions)	4.83	21.79	15.34	55.19	Municipal Registry
Internet access	0.37	0.48	0.92	0.27	Municipal Registry
TA in formulation of investment projects	0.66	0.47	0.57	0.49	Municipal Registry
Manages health centers	0.22	0.41	0.32	0.46	Municipal Registry
Major affiliated to the government party	0.11	0.31	0.13	0.33	Electoral data
Pctg. HH head secondary	0.22	0.15	0.34	0.16	Census
Sewerage connectivity	0.25	0.27	0.46	0.29	Census
Electricity connectivity	0.56	0.26	0.79	0.16	Census
Drinking water is treated			0.99	0.10	Municipal Registry
Sewage is treated	0.23	0.42	0.57	0.50	Municipal Registry
Transport expenditure (millions)	1.26	7.00	1.51	7.09	SIAF Reports
Energy expenditure (millions)	0.03	0.19	0.13	0.94	SIAF Reports
Health expenditure (millions)	0.60	2.33	0.28	1.33	SIAF Reports

Note: The beginning period is 2005 and the end period is 2015. Columns (1) and (3) provide the sum for the variables of interest and the mean for the geographical and control variables for 2005 and 2015, respectively. Columns (2) and (4) provide the standard deviation for control variables for 2005 and 2015, respectively, and column (2) for the cross-sectional geographical variables. Column (5) shows the data source used to compute each of the variables.

Figure 2: Number of sewerage projects potentially developed nation-wide



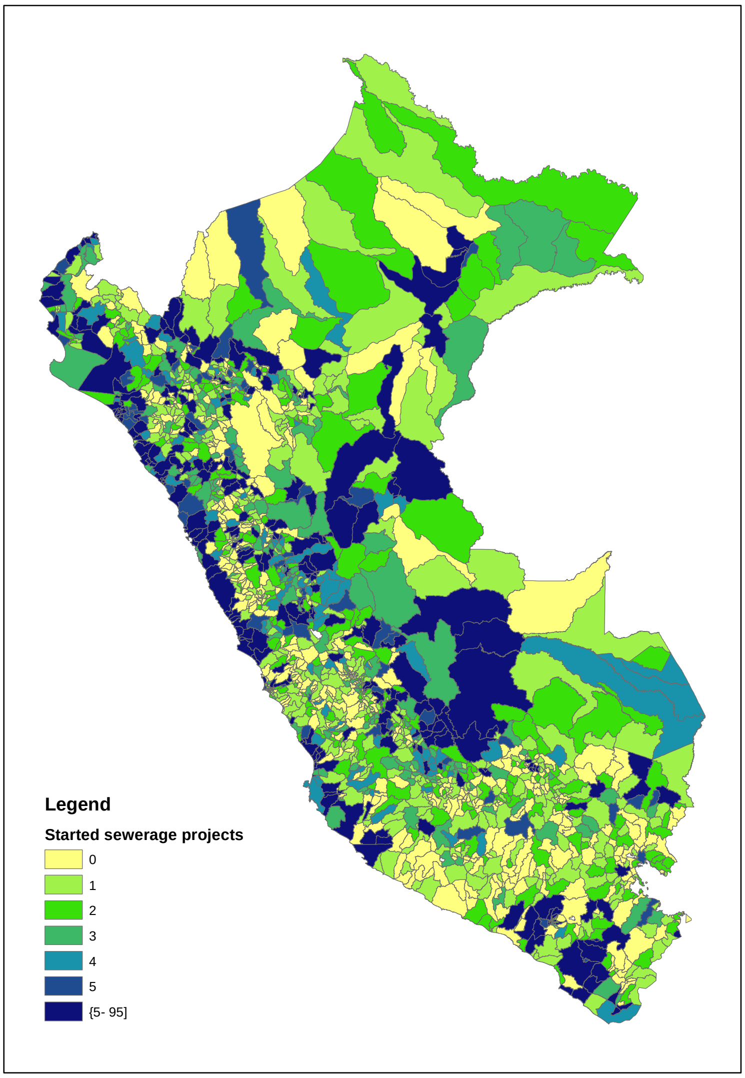
Note: Potential projects calculated by dividing the national expenditure on sewerage projects by the average cost of a project.

Table 2: Geographic cost parameters for sewerage diffusion

Dependent variable	Sewerage projects 2005-2015	
	OLS coeff. (1)	Beta coeff. (2)
Fraction district gradient $\{0.8\text{-}4.19]\%$	0.833 [2.047]	0.022
Fraction district gradient $\{4.19\text{-}13]\%$	2.315 [1.785]	0.064
Fraction district gradient above 13%	0.903 [1.542]	0.038
Fraction district elevation $\{250\text{-}500]$ mamls	-5.015^{***} [1.475]	-0.103
Fraction district elevation $\{500\text{-}1000]$ mamls	-1.425 [1.818]	-0.029
Fraction district elevation above 1000 mamls	-6.710^{***} [1.233]	-0.369
District area (sq. km)	-0.001^{**} [0.000]	-0.134
River density (km/sq km)	0.005^{*} [0.003]	0.096
Observations	1832	

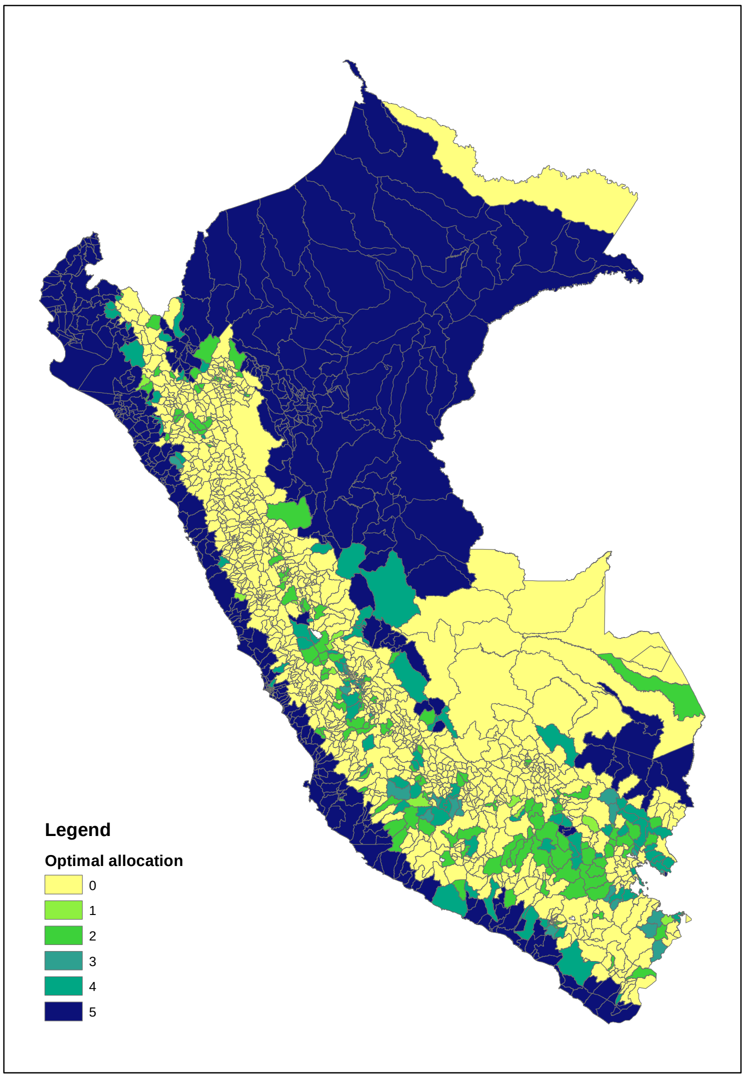
Note: The dependent variable is the number of started sewerage projects between 2005 and 2015. Column (1) shows the coefficients of an ordinary least squares (OLS) regression and column (2) shows the standardizes beta coefficients. The reference category for district gradient is flat gradient (below or equal 0.8) and for district elevation low elevation (below of equal 250 mamls.). Robust standard errors in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 3: Sewerage diffusion across districts in Peru, 2005-2015



Note: This map shows the district boundaries of Peru and the distribution across districts of the total number of sewerage projects started between 2005 and 2015. Light-shaded districts are those that started no or few sewerage projects and dark-shaded districts are those that started several (up to 95) sewerage projects. Author's calculation using data on the number of sewerage projects started between 2005 and 2015 from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Figure 4: Predicted sewerage diffusion across districts in Peru, 2005-2015



Note: This map shows the district boundaries of Peru and the distribution across districts of the predicted number of sewerage projects to be started and completed between 2005 and 2015. Light-shaded districts are those in which no or few sewerage projects were allocated and dark-shaded districts are those in which several sewerage projects were allocated. Author's calculation using data on the number of sewerage projects started between 2005 and 2015 from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Table 3: First-stage and reduced-form

	First-stage			Reduced-form	
	Cross-section	Over-time			
Dependent variable	Sewerage diffusion			IMR	U5MR
	(1)	(2)	(3)	(4)	(5)
Panel A: Started sewerage					
Suitability index	10.26^{***}				
	[1.623]				
Predicted cum. projects		0.455^{***}	0.447^{***}	0.457	0.0982^{**}
		[0.097]	[0.092]	[0.798]	[0.031]
Fstat	39.94	22.04	23.96		
Panel B: Sewerage in construction					
Predicted cum. projects		0.326^{***}	0.320^{***}		
		[0.069]	[0.065]		
Fstat		22.42	23.97		
Sample	All	IMR	U5MR	IMR	U5MR
District and year FE	No	Yes	Yes	Yes	Yes
District-year	776	10032	10494	10032	10494

Note: Column (1) in Panel A examines the cross-sectional relationship between the technical suitability index and the total number of projects started by 2015. Column (2) and column (3) examines the over-time relationship between predicted and actual sewerage diffusion. Columns (2) and (3) restrict the analysis to the IMR and U5MR sample, respectively. The dependent variable in column (1) is the total number of projects started by 2015 and in columns (2) and (3) it is the cumulative number of started sewerage projects (Panel A) and number of sewerage projects in construction (Panel B). The dependent variable in column (4) is IMR and in column (5) is U5MR. All regressions control for population density and total population. Standard errors clustered by province in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 5: Controlling for expenditure in other sectors

Dependent variable	Under-five mortality rate (U5MR)			
	(1)	(2)	(3)	(4)
Started sewerage	0.259^{**}	0.218	0.256^{**}	0.216
	[0.124]	[0.148]	[0.125]	[0.150]
Transport expenditure (ln)	0.006			0.006
	[0.009]			[0.010]
Energy expenditure (ln)		0.011		0.011
		[0.007]		[0.007]
Health expenditure (ln)			0.006	-0.002
			[0.008]	[0.010]
F-stat	13.09	8.096	12.85	7.775
District-year	8663	7547	8674	7547
Districts	1217	1069	1218	1069

Note: The dependent variable is the under-five mortality rate. Sewerage diffusion is instrumented by predicted sewerage diffusion. Each column controls progressively for expenditure in different sectors. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 4: Effect of sewerage diffusion on early-life mortality

	OLS				2SLS			
Dependent variable	IMR		U5MR		IMR		U5MR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Started sewerage	-0.484***	-0.441***	0.020***	0.011***	1.004	1.218	0.219***	0.230***
	[0.168]	[0.167]	[0.005]	[0.005]	[1.784]	[1.903]	[0.080]	[0.085]
Fstat					22.04	20.57	23.96	22.50
Sewerage in construction	-0.606**	-0.550**	0.027***	0.027***	1.402	1.702	0.307***	0.324***
	[0.247]	[0.246]	[0.007]	[0.007]	[2.488]	[2.654]	[0.112]	[0.120]
Fstat					22.42	20.89	23.97	22.40
Muni controls	No	Yes	No	Yes	No	Yes	No	Yes
Initial MR	56.44	56.44	4.82	4.82	56.44	56.44	4.82	4.82
District-year	10032	10032	10494	10494	10032	10032	10494	10494
Districts	1408	1408	1408	1408	1408	1408	1408	1408

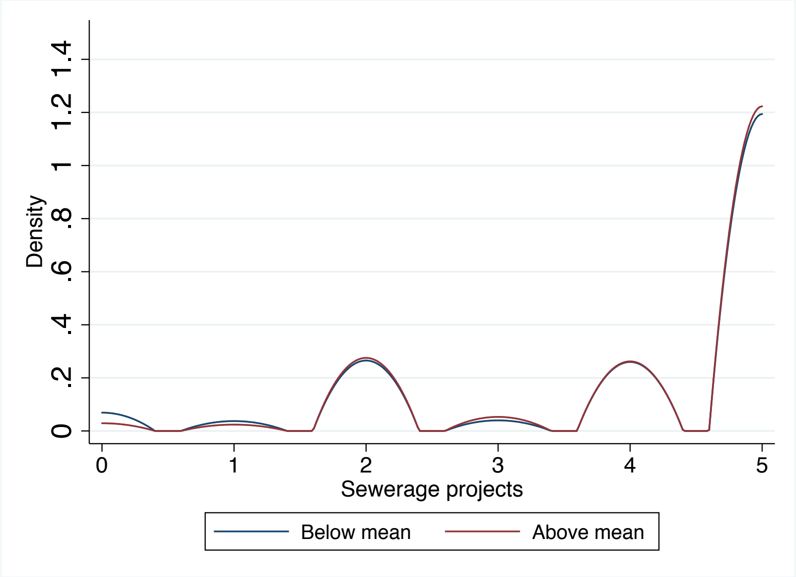
Note: This table presents the main results of the effect of the cumulative number of started sewerage project and the number of sewerage projects in construction on early-life mortality rates. Columns (1) to (4) show ordinary least-square (OLS) estimates and column (5) to (8) show 2SLS estimates. The dependent variable in columns (1), (2), (5) and (6) is the infant mortality rate (IMR) and in columns (3), (4), (7) and (8) is the under-5 mortality rate (U5MR). All regressions control for population density and total population and include district and year fixed effects. Muni controls include time-varying dummy variables indicating whether the district municipality has access to Internet, needs technical assistance to formulate investment projects, manages at least one health center and municipal income (ln). Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 6: IV not operating through alternative channels

Dependent variable	IMR			U5MR		
	(1)	(2)	(3)	(4)	(5)	(6)
Transport expenditure (ln)	22.38			1.444		
	[54.79]			[1.389]		
Energy expenditure (ln)		2.772			0.390	
		[7.241]			[0.305]	
Health expenditure (ln)			-5.579			-0.650
			[8.826]			[0.409]
F-stat	0.279	2.728	3.123	1.383	3.730	4.522
District-year	8222	7137	8233	8663	7547	8674
Districts	1217	1069	1218	1217	1069	1218

Note: The dependent variable is the infant mortality rate for columns (1) and (2) and the under-five mortality rate for columns (3) and (4). Each alternative channel is instrumented by predicted sewerage. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 5: Distribution of predicted projects by % rural population



Notes: This figure shows the distribution of predicted sewerage projects by the district's percentage of rural population. The blue distribution corresponds to districts with rural population below the median and the red distribution corresponds to districts above the median of the distribution of the percentage of rural population by 2005.

Table 7: Controlling for geography-specific trends

Dependent variable	Under-five mortality rate (U5MR)				
	Gradient	Elevation	Area	Region	Amazon
	(1)	(2)	(3)	(4)	(5)
Sewerage in construction	0.313^{**}	0.242	0.307^{**}	0.220^{*}	0.309^{**}
	[0.149]	[0.188]	[0.119]	[0.126]	[0.122]
F-stat	15.89	11.00	22.67	16.98	22.40
Initial MR	10494	10494	10494	10494	10494
District-year	1408	1408	1408	1408	1408

Note: The dependent variable is the under-five mortality rate. Sewerage diffusion is instrumented by the constructed prediction of sewerage diffusion. Each column includes the interaction of year with each geographic component either in isolation (columns 1-3) or jointly (column 4). The interacted component in column (1) is the flat gradient category, column (2) the low elevation category and column (3) the district area in km2. Column 4 controls for the interaction with all the geographic characteristics. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 8: Fertility and selective migration do not explain the results

Dependent variable	Births	Pop u5	Educ sec	Electricity
	(1)	(2)	(3)	(4)
Started sewerage	0.132^{***}	0.007^{*}	-0.001	-0.023^{***}
	[0.043]	[0.004]	[0.001]	[0.007]
Fstat	19.37	21.82	16.85	21.05
Sewerage in construction	0.180^{***}	0.009^{*}	-0.001	-0.037^{***}
	[0.059]	[0.005]	[0.002]	[0.012]
Fstat	20.51	22.62	16.08	18.13
Initial mean	304.7	2472.1	0.219	0.557
District-year	10032	10494	2630	1406
Districts	1408	1408	1014	703

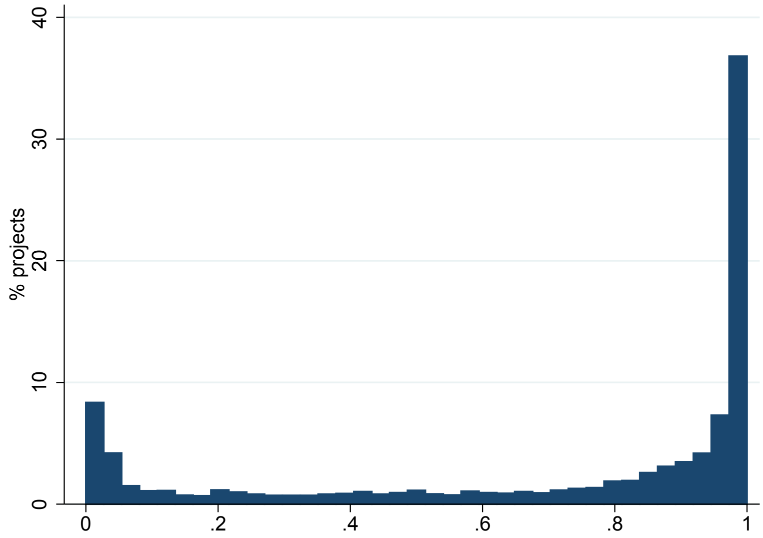
Note: The dependent variable in column (1) is number of live births (ln), column (2) is under-five population (ln), column (3) is percentage of household heads with secondary education completed and column (4) is percentage of households connected to the electricity network. Sewerage diffusion is instrumented by the constructed prediction of sewerage diffusion. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 9: Effect of sewerage diffusion on MR, by cause of death

	Waterborne (1)	Accidents (2)	Respiratory (3)	Malformation (4)	Other (5)
Started sewerage	0.169***	0.075*	-0.006	0.024	-0.016
	[0.066]	[0.044]	[0.037]	[0.029]	[0.047]
Fstat	21.82	21.82	21.82	21.82	21.82
Sewerage in construction	0.233***	0.103*	-0.009	0.033	-0.022
	[0.090]	[0.060]	[0.051]	[0.040]	[0.065]
Fstat	22.62	22.62	22.62	22.62	22.62
Initial mean	2.265	1.248	0.736	0.388	1.302
District-year	10494	10494	10494	10494	10494
Districts	1408	1408	1408	1408	1408

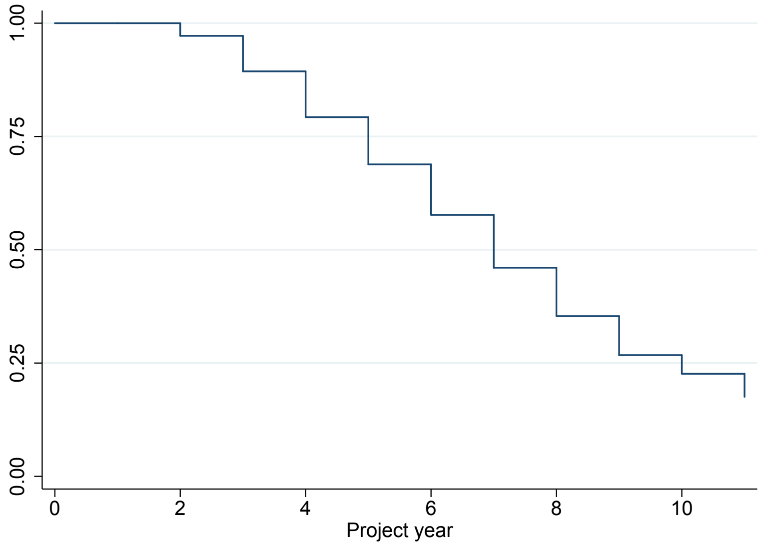
Note: The dependent variable is the under-five mortality rate, split by cause of death related to waterborne diseases (column 1), accidents (column 2), respiratory diseases (column 3); malformations (column 4); other unrelated causes (column 5).The number of projects in construction is instrumented by the prediction of sewerage diffusion. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 6: Completion rate, projects between 2005-2015



Notes: Completion rates computed as total accrued investment until 2015 divided by budgeted investment. Sample restricted to projects started before 2015 because of right-censoring of the data (finished in 2015) and the fact that sewerage projects optimally take one year to be completed

Figure 7: Project completion hazard rate



Notes: Completion time computed as the number of years it takes for a project to be completed (ever accrued more than 90 percent of the budgeted investment).

Table 10: Effect of time exposed to construction works on early-life mortality

Dependent variable	Δ Infant Mortality		Δ Under-5 Mortality	
	(1)	(2)	(3)	(4)
Years exposed to construction	0.322*	0.355	0.142**	0.163*
	[0.180]	[0.229]	[0.0688]	[0.0925]
Fstat	9.943	5.519	9.990	5.679
1 year mean	0.787	0.787	0.247	0.247
Muni/project controls	No	Yes	No	Yes
Districts	186	186	200	200

Note: The dependent variable is the average annual change in infant mortality rates (columns 1 and 2) and under-five mortality rates (column 3 and 4). "Years exposed to construction" is instrumented by the maximum predicted number of projects (Panel A) and by the suitability index (Panel B). All regressions control for initial population density and total population. Muni controls include initial characteristics capturing whether the district municipality had access to Internet, needed technical assistance to formulate investment projects, managed at least one health center and municipal income (ln). Project controls include the average budgeted investment and the total number of projects started between 2005 and 2015. Sample restricted to districts in which construction works ever took place (i.e. at least 1 sewerage project started). Robust standard errors in brackets. Statistical significance denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Effect of sewerage diffusion on connectivity and treatment

Dependent variable	Connectivity (1)	Δ Connectivity (2)	Treat Water (3)	Treat Sludge (4)
Completed sewerage (L2)	-0.0556***	-0.0236	-0.105**	-0.0316
	[0.0198]	[0.0156]	[0.0417]	[0.0642]
Fstat	18.01	21.77	15.04	15.32
District-year	2814	1405	9712	11040
Projects completed	-0.0179	-0.0137	1.369	1.562
	[0.0135]	[0.0121]	[2.592]	[4.118]
Fstat	20.85	27.13	0.304	0.175
District-year	4220	1405	9712	12307
Initial mean	0.227	0.227	0.838	0.234
Districts	1408	1408	1408	1408

Note: The dependent variable in column (1) is the percentage of households connected to sewerage, column (2) the change in sewerage conenctivity between 2005-2015, column (3) is an indicator of whether the municipality reports that water is treated in the district and column (4) whether sludge is treated in the district. The cumulative number of started and completed projects are instrumented using the predicted number of sewerage projects. All regressions control for population density and total population. Standard errors clustered by district in brackets, except for column (2) that presents robust standard errors in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

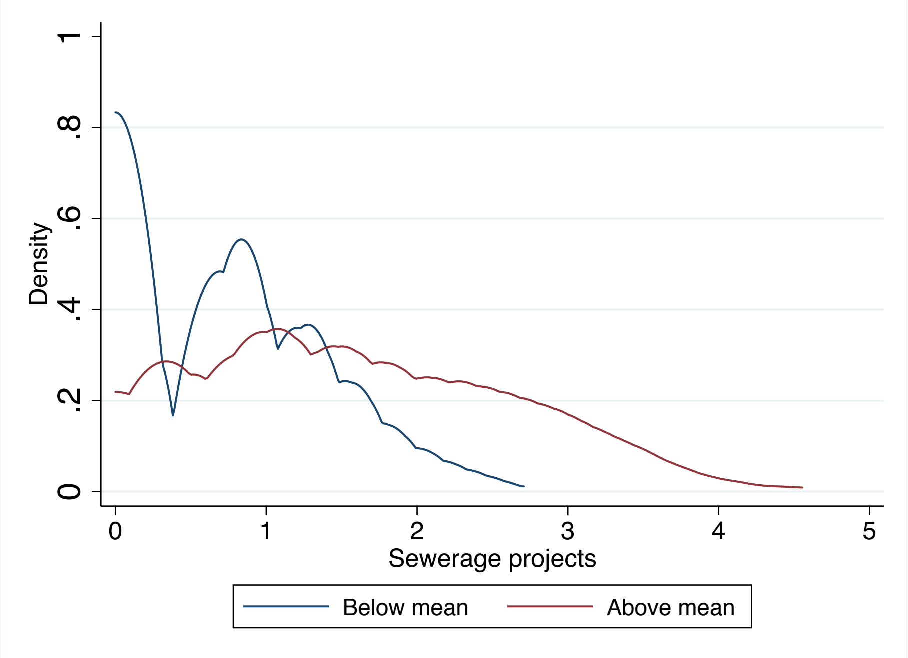
Table 11: Effect of sewerage completion on early-life mortality

	OLS				2SLS			
Dependent variable	IMR		U5MR		IMR		U5MR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Completed sewerage (L2)	-1.459***	-1.414***	0.008	0.008	2.047	3.784	0.160	0.148
	[0.527]	[0.529]	[0.016]	[0.016]	[6.146]	[6.443]	[0.211]	[0.217]
Muni	No	Yes	No	Yes	No	Yes	No	Yes
Fstat	.	.	.	.	10.54	9.991	11.33	10.68
District-year	7973	7973	8336	8336	7973	7973	8336	8336
Districts	1324	1324	1354	1354	1324	1324	1354	1354
Completed sewerage (L3)	-2.024***	-2.027***	-0.017	-0.017	-0.167	2.472	-0.065	-0.063
	[0.693]	[0.700]	[0.019]	[0.019]	[6.762]	[6.843]	[0.231]	[0.230]
Muni	No	Yes	No	Yes	No	Yes	No	Yes
Fstat	.	.	.	.	11.32	11.69	12.24	12.49
District-year	6960	6960	7316	7316	6960	6960	7316	7316
Districts	1276	1276	1320	1320	1276	1276	1320	1320
Initial MR	56.44	56.44	4.82	4.82	56.44	56.44	4.82	4.82
Muni controls	No	Yes	No	Yes	No	Yes	No	Yes

Note: This table presents the results of the effect of the lagged (L2 is a 2-year lag and L3 is a 3 year lag) cumulative number of completed sewerage project on early-life mortality rates. Columns (1) to (4) show ordinary least-square (OLS) estimates and column (5) to (8) show 2SLS estimates. The dependent variable in columns (1), (2), (5) and (6) is IMR and in columns (3), (4), (7) and (8) is U5MR. All regressions control for population density and total population and include district and year fixed effects. Muni controls include time-varying dummy variables indicating whether the district municipality has access to Internet, needs technical assistance to formulate investment projects, manages at least one health center and municipal income (ln). Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

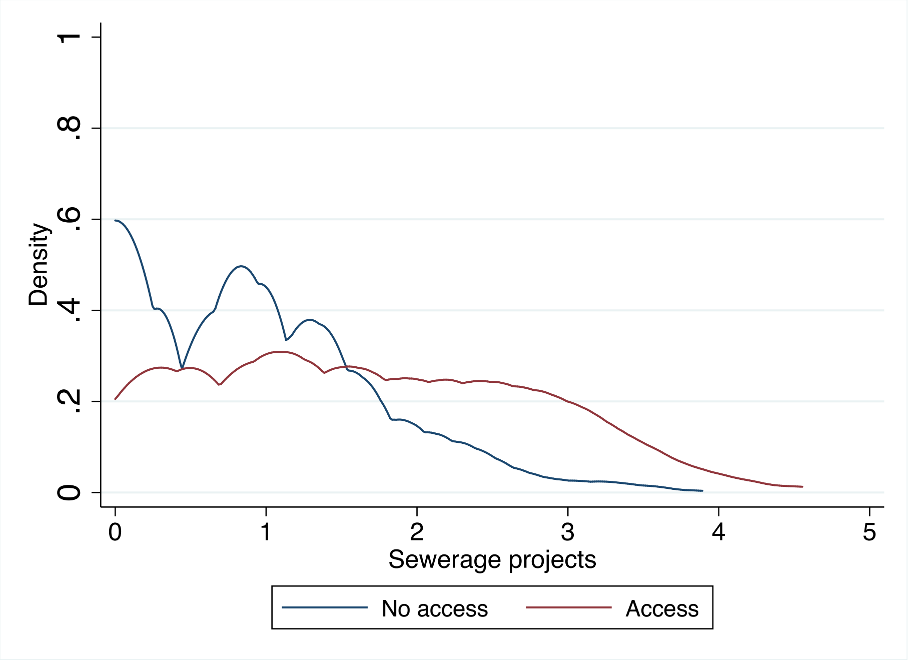
8 Appendix

Figure 8: More sewerage projects allocated to richer municipalities



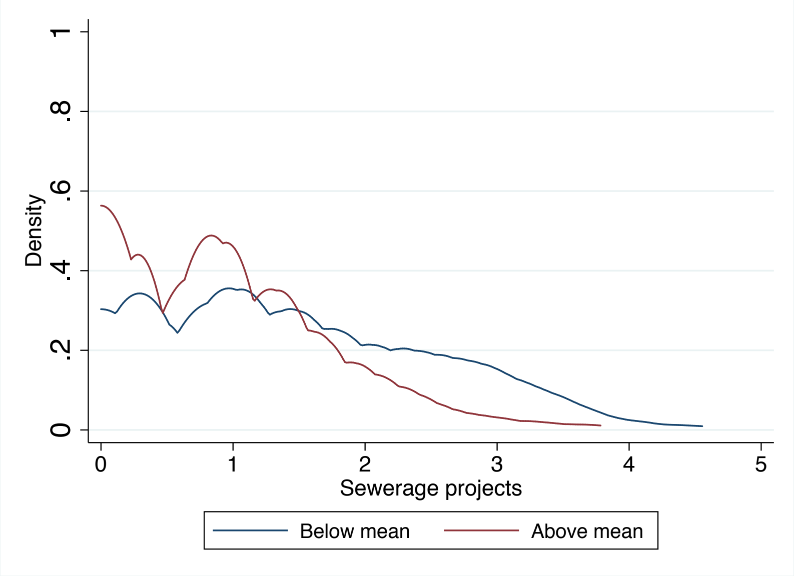
Note: This figure shows the distribution of started sewerage projects by initial municipal revenue. The blue distribution corresponds to municipalities with budget below the median and the red distribution corresponds to municipalities with budget above the median of the distribution of municipal budget by 2005.

Figure 9: More sewerage projects allocated to municipalities with Internet access



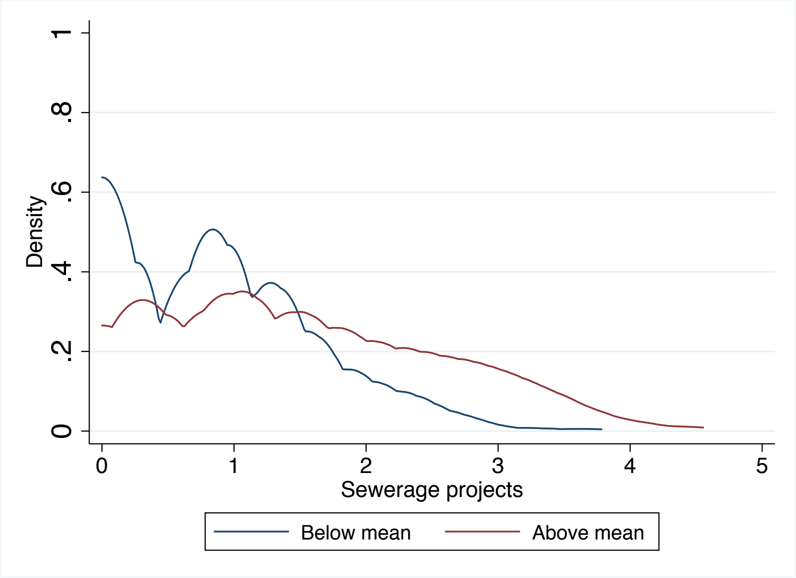
Note: This figure shows the distribution of started sewerage projects by initial Internet access. The blue distribution corresponds to municipalities without access and the red distribution corresponds to municipalities with Internet access by 2005.

Figure 10: Distribution of projects by % population with unmet basic needs



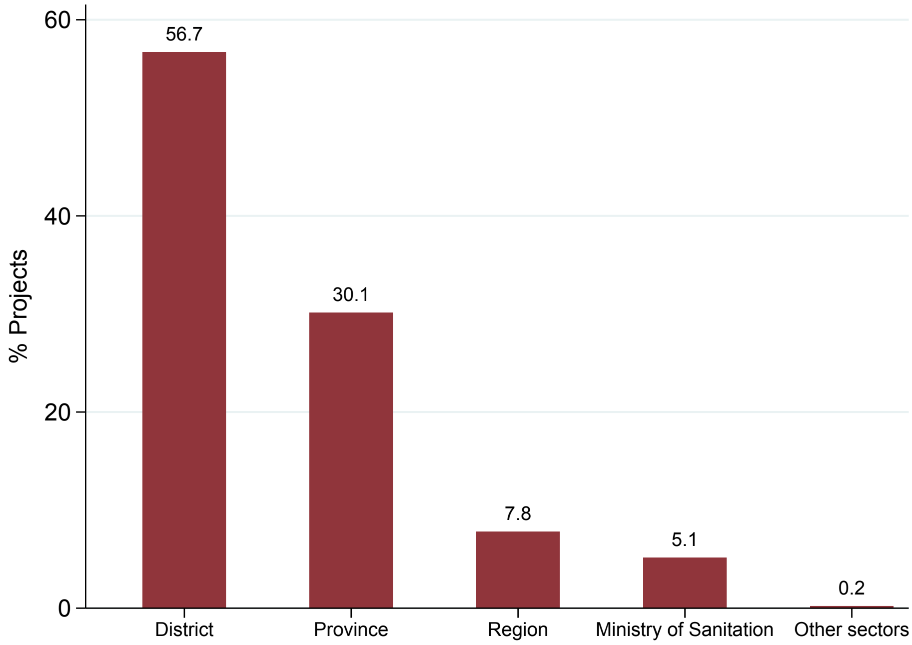
Notes: This figure shows the distribution of started sewerage projects by the district's percentage of households with unmet basic needs. The blue distribution corresponds to districts with a percentage of household with unmet basic needs below the median and the red distribution corresponds to disricts above the median of the distribution by 2005.

Figure 11: Distribution of predicted projects by % sewerage connectivity



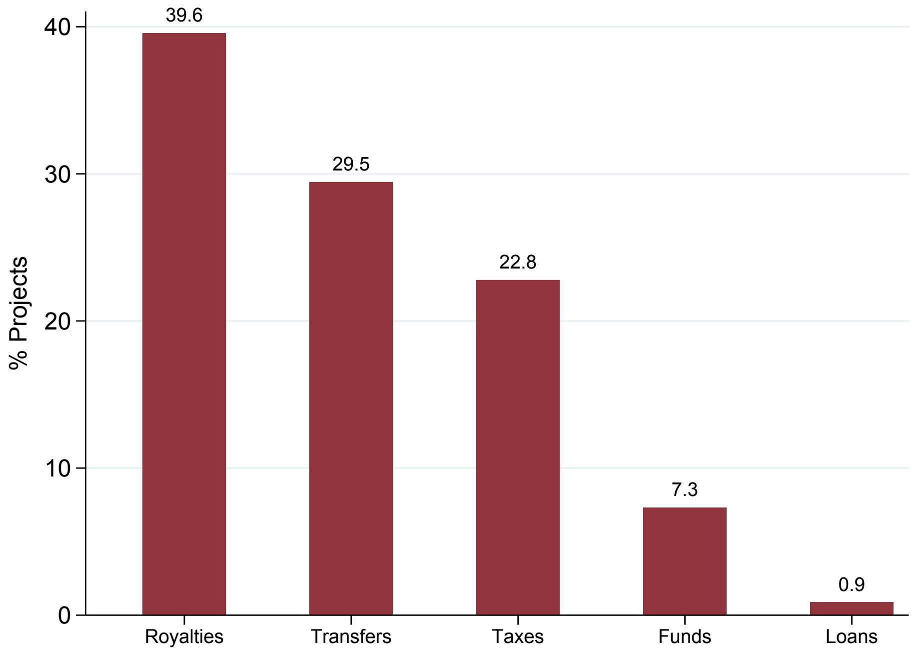
Notes: This figure shows the distribution of started sewerage projects by the district's percentage of households already connected to sewerage. The blue distribution corresponds to districts with a percentage of household connected to sewerage below the median and the red distribution corresponds to districts above the median of the distribution of sewerage connectivity by 2005.

Figure 12: Agency Formulating Sewerage Projects, 2005-2015



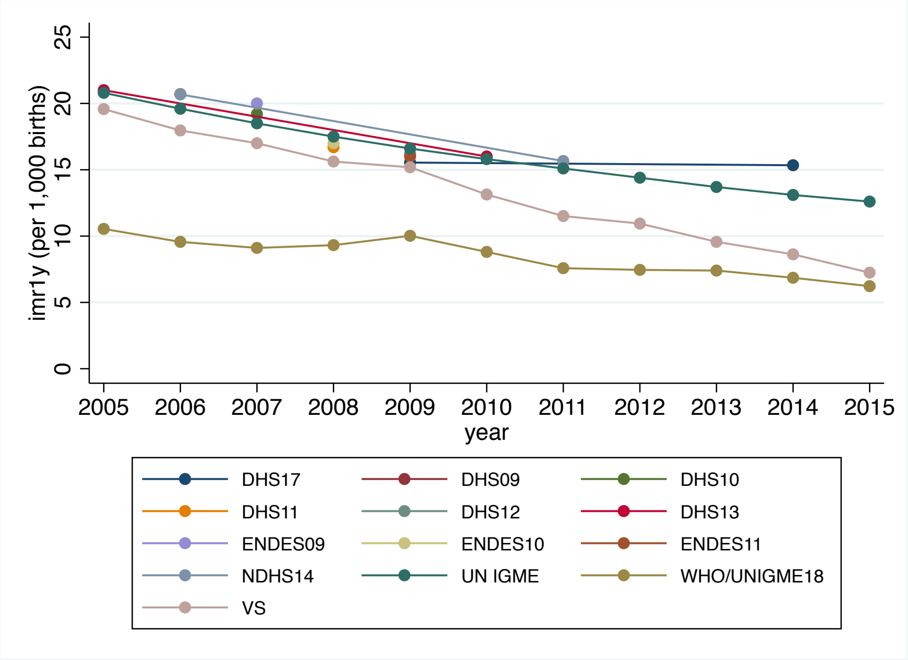
Note: This figure shows the percentage of sewerage projects formulated by each government agency. The percentage is calculated from the pool of projects declared viable and started between 2005 and 2015. Author's calculation using data from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Figure 13: Financing Sources for Sewerage Projects, 2005-2015



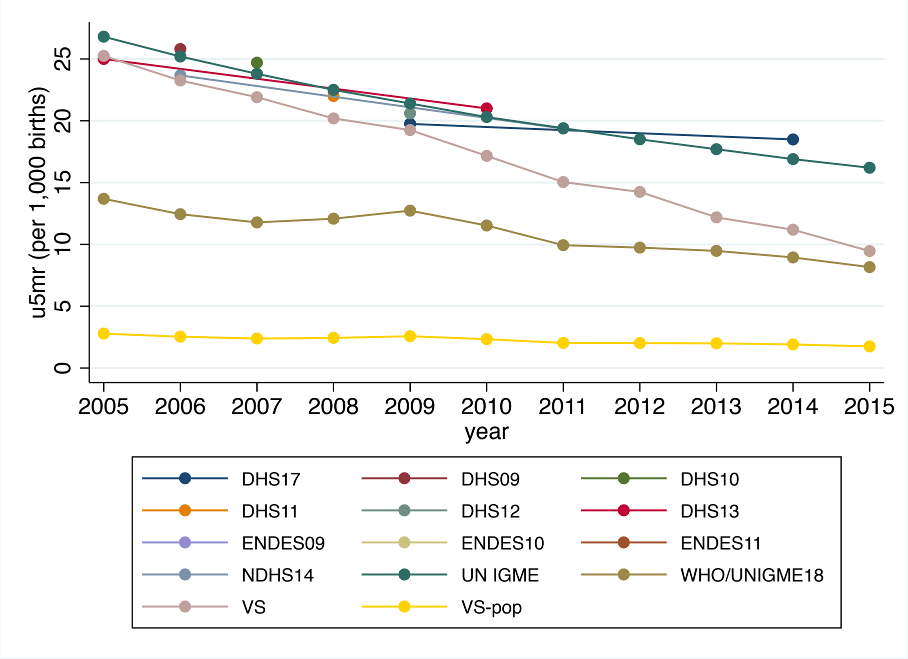
Note: This figure shows the percentage of sewerage projects financed by each of the different public resources. The percentage is calculated from the pool of projects declared viable and started between 2005 and 2015. Author's calculation using data from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Figure 14: IMR from Vital Statistics Compared to Other Data Sources



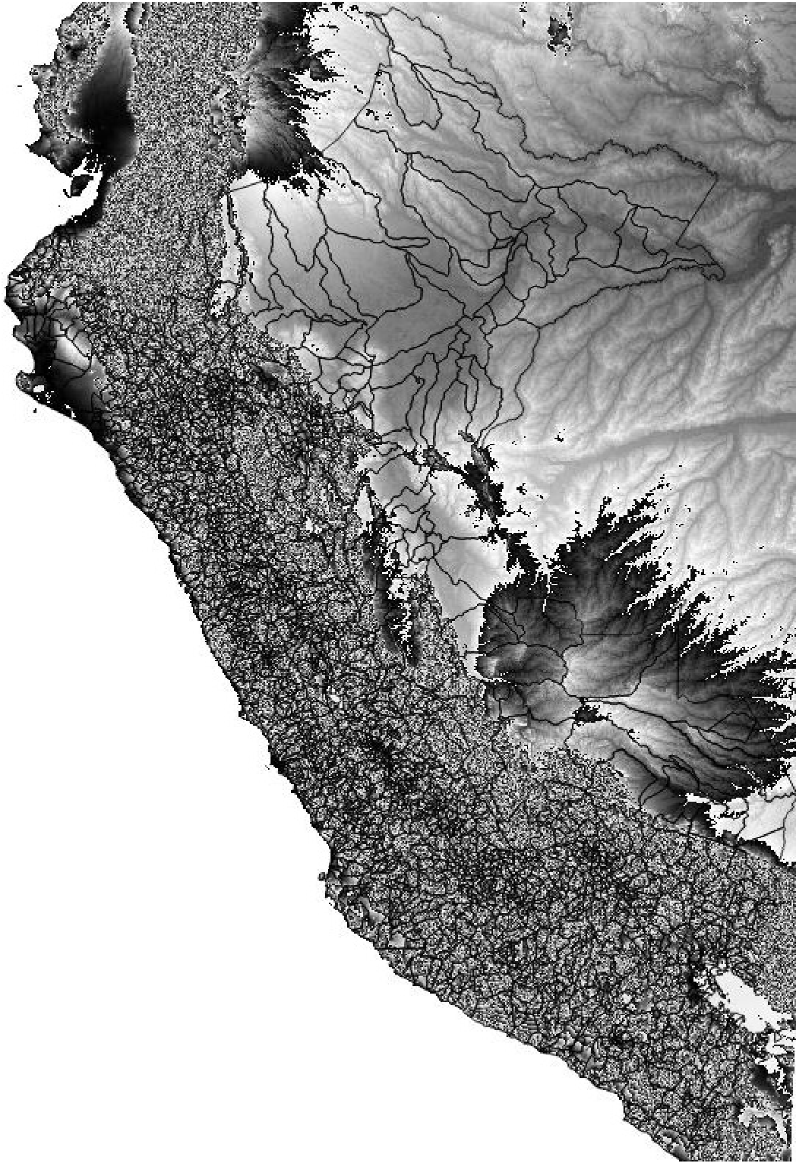
Note: Alternative data obtained from the Health and Demographic Surveys (DHS), National Survey of Health and Demography (ENDES) and Inter-agency Group for Child Mortality Estimation (UN IGME).

Figure 15: U5MR from Vital Statistics Compared to Other Data Sources



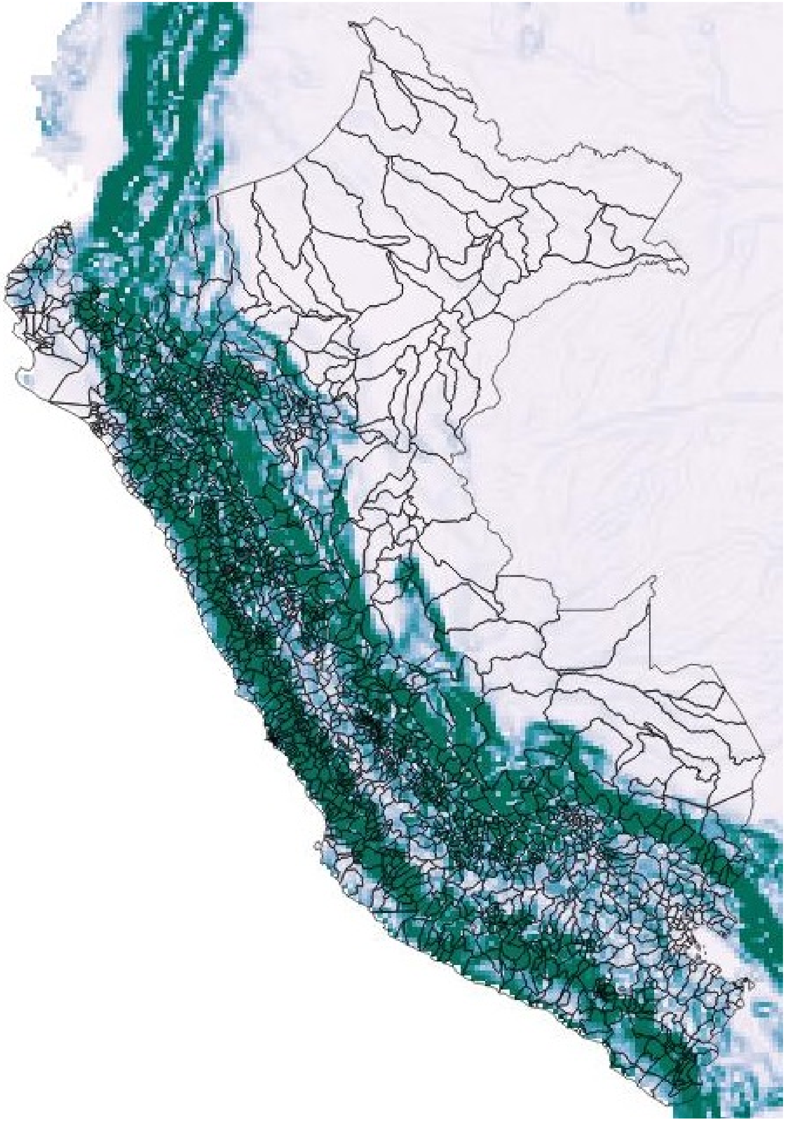
Note: Alternative data obtained from the Health and Demographic Surveys (DHS), National Survey of Health and Demography (ENDES) and Inter-agency Group for Child Mortality Estimation (UN IGME).

Figure 16: Elevation in Peru



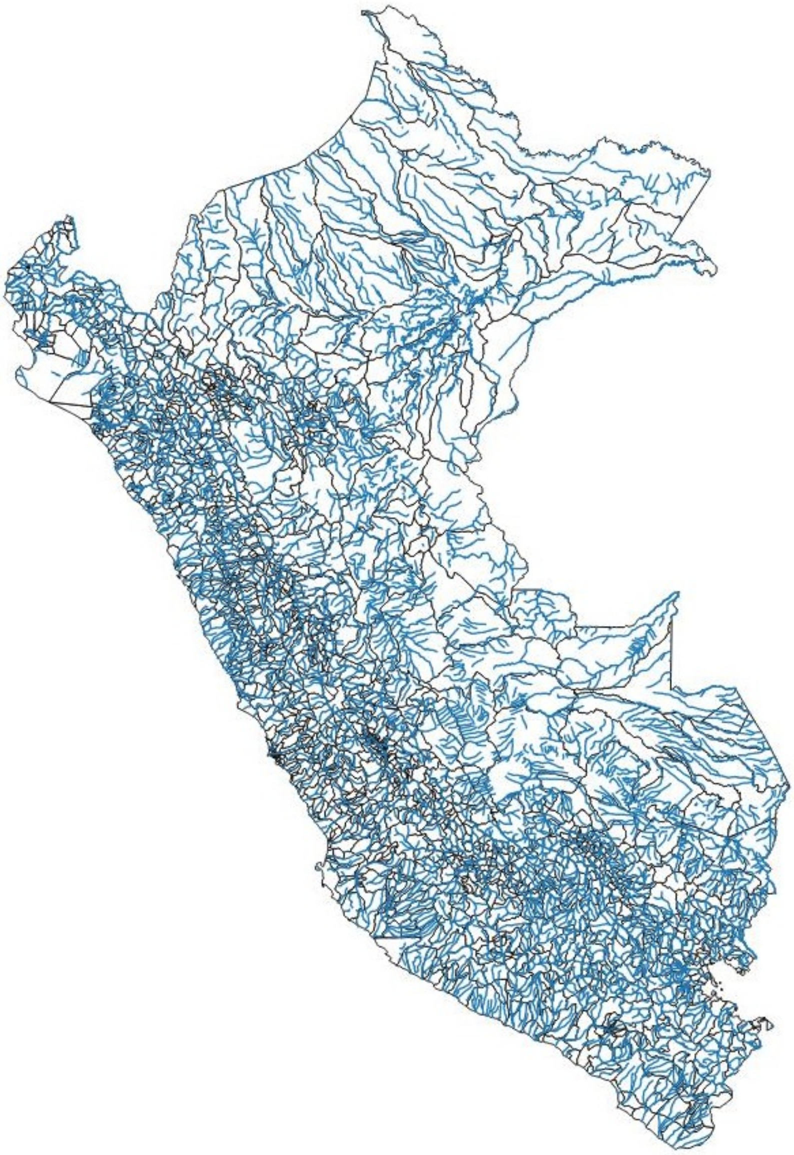
Note: Darker shaded grid-cells are in higher altitude. Source: digital elevation map provided by the Peruvian Ministry of Environment with information on surface elevation for multiple cells (1x1 km).

Figure 17: Ruggedness in Peru



Note: Darker shaded grid cells are steeper. Gradient is a measure of the steepness of the ground surface calculated with ArcMap using elevation in each cell and neighboring cells. Source: digital elevation map provided by the Peruvian Ministry of Environment with information on surface elevation for multiple cells (1x1 km).

Figure 18: River density in Peru



Note: Rivers are denoted by the blue lines within district boundaries (black lines). Source: digital elevation map provided by the Peruvian Ministry of Environment with information on surface elevation for multiple cells (1x1 km).

Table 13: Effect of sewerage diffusion on MR - logs

Dependent variable	IMR		
	(1)	(2)	(3)
Cum. number of started sewerage	0.437 [0.281]		
Num. sewerage in construction		0.610 [0.392]	
Cum. number of completed sewerage			1.538 [1.050]
F-stat	22.04	22.42	10.16
District-year	10032	10032	10032
Districts	1408	1408	1408
Dependent variable	**U5MR**		
	(1)	(2)	(3)
Cum. number of started sewerage	0.054^{**} [0.021]		
Num. sewerage in construction		0.076^{**} [0.030]	
Cum. number of completed sewerage			0.186^{**} [0.081]
F-stat	24.99	24.14	12.68
District-year	10390	10390	10390
Districts	1405	1405	1405

Note: All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

8.1 Effect of sewerage diffusion on the extensive margin

To estimate the effect of a district receiving at least 1 sewerage project, I rely on a modified version of my instrument. First, I identify how many districts can be intervened each year based on the sewerage expenditure per district between 2005 and 2015 and the annual budget. Second, I build an alternative technical suitability index from a logistic regression of the probability of a district being intervened between 2005 and 2015 on the geographic factors and I use this index to rank all districts. Finally, I categorize as intervened the highest-ranking districts until the annual national budget is exhausted. This process is repeated every year with districts not previously intervened. The alternative instrument is a binary variable equal to 1 in all years since a district is predicted to be intervened and 0 otherwise. The 2SLS estimates of the effect of the extensive margin of sewerage diffusion on mortality rates are presented in Appendix Table 14. These estimates should be interpreted as the local average treatment effect (LATE) or average causal effect of starting at least one sewerage project on districts that did so because of their technical suitability (i.e. compliers). On average, a district's IMR increased by 23 and U5MR increased by 4.3 if the district starts sewerage projects. This represents a 39 and 86 percent increase, respectively, from initial average mortality rates. The predictive power of the alternative instrument is still high (Fstat higher than 40) and the effect is only precisely estimated for U5MR at a 1 percent statistical significance.

8.2 Sensitivity analysis

Table 15: Sensitivity Analysis - U5MR

Dependent variable	Under-five mortality rate (U5MR)			
	Intervened (1)	Exc. Lima (2)	Amazon (3)	Top-coded (4)
Started sewerage	0.233^{***}	0.342^{**}	0.219^{***}	0.236^{***}
	[0.081]	[0.160]	[0.080]	[0.084]
F-stat	20.73	8.402	23.96	28.64
Sewerage in construction	0.325^{***}	0.499^{**}	0.307^{***}	0.229^{***}
	[0.112]	[0.232]	[0.112]	[0.088]
F-stat	21.11	8.544	23.97	18.87
District-year	8595	9725	10494	10494
Districts	1108	1317	1408	1408

Note: The dependent variable is the under-five mortality rate. Alternative specifications of the 2SLS results of Table 4 without including municipal controls. In column (1) I restrict the sample of analysis to districts that started at least one sewerage project. In column (2) I exclude Lima and in column (3) I add an indicator of whether the district is located in the Amazon region. In column (4) I replace the independent variable with a version top-coded at the 90th percentile of the distribution of sewerage projects. All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

8.3 Alternative IV

My main results also remain robust when using as an instrumental variable approach a modified version of Duflo and Pande (2007) strategy. I use the interaction of sewerage diffusion at an aggregate level with the steep gradient categories (keeping the flat gradient category as reference) to instrument sewerage diffusion at the district level. I use two alternative measures of sewerage diffusion at an aggregate level. First, I use the cumulative number of projects started or the number of projects in construction in the province that the district belongs to. The identification assumption of this instrument is that absent sewerage diffusion, the evolution of early-life mortality across districts located in the same province but with different gradient would not have systematically differ across provinces with high and low sewerage diffusion. The main limitation of this aggregate measure is that it is not independent of province-level shocks that also affect districts. Second, I use the cumulative number of projects started or the number of projects in construction in adjoining district. To do so, I conduct a spatial analysis of district boundaries. For each district, I calculate the average number of sewerage projects implemented by districts sharing the same borders. By construction, this instrument is independent of district-specific mortality shocks. The exclusion restriction of this instrument is satisfied if sewerage diffusion in adjoining districts did induce

Table 14: Effect of sewerage on MR - Extensive Margin

	OLS				2SLS			
Dependent variable	IMR		U5MR		IMR		U5MR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Started district	1.820	3.780	0.0417	0.136	23.84	15.79	4.323***	3.973***
	[2.441]	[3.213]	[0.0807]	[0.102]	[20.82]	[27.31]	[0.993]	[1.062]
Fstat					47.04	39.47	47.72	40.24
Muni	No	Yes	No	Yes	No	Yes	No	Yes
District-year	8150	5152	8536	5384	8150	5115	8506	5331
Districts	1092	913	1129	942	1092	876	1099	889

Note: All regressions control for population density and total population and include district and year fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

sewerage diffusion in the own district, conditional on gradient, but was not related to demand-side factors affecting mortality rates. The spatial correlation is thus expected to happen through the supply-side. The results are shown in Appendix Table 16. Using the interaction of gradient with sewerage diffusion at an aggregate level, I also find that sewerage diffusion increased IMR and U5MR. It is reassuring to find positive and statistically significant estimates when using alternative instrumental variables. This bolsters the confidence to interpret my main results as causally identified estimates of the effect of sewerage diffusion on early-life mortality rates.

Table 16: Effect of sewerage diffusion on MR - alternative IV

Dependent variable	IMR		U5MR	
	(1)	(2)	(3)	(4)
Panel A: IV with province sewerage diffusion				
Started sewerage	0.794	2.079*	0.072**	0.075**
	[0.610]	[1.116]	[0.037]	[0.038]
F-stat	10.28	11.61	12.86	12.69
Sewerage in construction	1.552	4.286**	0.099*	0.102*
	[0.982]	[1.893]	[0.056]	[0.054]
F-stat	10.02	11.62	11.01	12.53
Panel B: IV with province sewerage diffusion				
Started sewerage	1.029*	3.058**	0.0452	0.0537
	[0.600]	[1.333]	[0.037]	[0.039]
F-stat	14.44	12.80	16.08	13.86
Sewerage in construction	1.921*	6.236**	0.079	0.083
	[1.120]	[2.504]	[0.066]	[0.071]
F-stat	20.93	17.85	23.83	21.87
Muni controls	No	Yes	No	Yes
Baseline mean	57.74	57.74	4.954	4.954
District-year	10021	6474	10621	6852
Districts	1407	1149	1466	1202

Note: All regressions include interactions of province sewerage diffusion with district elevation categories and land area, district fixed effects, a full set of province*year interactions and gradient*year interactions. Standard errors clustered by province in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

8.4 What determines project non-completion and duration?

The literature offers several reasons for delays in project execution and mid-construction abandonment. Williams (2017) models project non-completion as a dynamic inconsistent outcome of a collective choice process in the context of limited funds. In the budgeting phase, a group of political actors with different distributive preferences select a set of projects to allocate funds to. Vote-trading coalitions are formed, but these require intertemporal bargains that are difficult to maintain because legislators whose projects were favoured may violate the promise ex-post. These commitment failures make any coalition vulnerable to an alternative proposal during the next budgeting process. Political negotiations over expenditure priorities are ongoing throughout the year, as plans and budgets are not strictly executed. These ongoing negotiations occur frequently through informal channels. In Peru, informal lobbying in an attempt to influence expenditure decisions is common . The result is that the governments collective expenditure priorities are frequently changing and hence the government is often starting new projects while other half-built projects decay. The more unstable the collective choice from the budgeting phase is, the more likely a project is interrupted with no guarantee of completion. Corruption and clientelism are other factors that can affect project completion. In a novel approach to clientelistic models, Robinson and Verdier (2013) argue that public delivery can take an inefficient form because of a two-sided credibility problem. On the one hand, self-interested politicians face a commitment problem because it is not in their interests to implement ex-post the policies that would induce people to vote for them. On the other hand, citizens might vote for a politician if offered sufficient incentives. Therefore both politicians and citizens must commit. For politicians to ensure that they have the support of voters, they must be able to use policies that tie the continuation utility of a voter to their political success, or alternatively, voters will be punished. Leaving intentionally unfinished projects right before the electoral year is a way that politicians can ensure votes, as project completion is at risk with a change in power. Another factor highlighted by the literature is managerial practices. Rasul and Rogger (2018) find that that the quality of management matters for completion rates. Increasing bureaucrats autonomy is positively associated with completion rates, but using incentives and monitoring mechanisms as bureaucratic managerial practices is negatively associated with completion rates.

To shed lights on the determinants of project non-completion and duration, I formally estimate a discrete-time hazard model of the probability of completing sewerage projects using the sample of all started sewerage projects between 2005 and 2015 (19,106 project-year observations, 4,246 projects in 929 districts). I model the probability that a sewerage project in a given municipality and year after starting the project is completed as a function of project time-invariant and district time-varying covariates and controlling non-parametrically for duration and district fixed effects. In addition, I model the duration of projects ever completed as a function of time-invariant project characteristics and district characteristics of the year in which the project started.

The results reported in Table 17 show several interesting insights. The outcome in (1) and (2) is an indicator when a project is completed and the outcome in column (3) is the duration of projects from start until completion. The likelihood of project completion is lower if the district municipality executes the project with its own funds (i.e. mining royalties and/or tax revenue), but this is offset if the mayor is affiliated to the central government political party (column 2). These findings are consistent with Williams (2017b) theory of collective choice model. If a district municipality has full discretion over the resources funding a sewerage project, unstable local political bargains may reallocate the money intended to finish the project to start another project (even in a different sector). Having the mayor affiliated to the governments political party may create top-down political pressures that align the incentives of local politicians to complete started projects. Furthermore, the likelihood of project completion decreases in years in which there is a General Election (presidential and parliamentarian). This is consistent with the clientelistic model of politicians deliberately leaving projects incomplete before elections to tie the continuation utility of a voter to their political success (Robinson and Verdier 2013). Incumbents are more likely than challengers to complete projects started by them to take credit from it. Thus, leaving unfinished projects could incentivize voters to re-elect incumbents.

Project completion is also associated with various project and municipal characteristics. The likelihood of project completion decreases with budgeted investment, a proxy for project complexity, and with the number of other sewerage projects in construction, perhaps due to resource dispersion. The likelihood of project completion decreases with population density, though it increases with total population. Notably, the likelihood of project completion decreases if the district needs technical assistance in the formulation of public investment projects. This is in line with Rasul and Roggers (2018) finding that the quality of management matter for project completion.

Table 17: Discrete-Time Hazard Model: Project Completion

	Likelihood		Duration
	(1)	(2)	(3)
District and own funds	-0.057***	-0.057***	0.212*
	[0.019]	[0.019]	[0.126]
Major affiliated to the gov party	-0.006	-0.001	
	[0.017]	[0.017]	
District and own funds × Major affiliated to gov party	0.042	0.0546*	
	[0.032]	[0.0295]	
District and gov funded	0.001	0.002	-0.049
	[0.028]	[0.028]	[0.158]
District and government funded × Major affiliated to gov party	0.020	0.024	
	[0.047]	[0.047]	
General elections	-0.023***	-0.019**	
	[0.009]	[0.009]	
Local elections	0.003	0.009	
	[0.004]	[0.007]	
Budgeted investment (ln)	-0.019***	-0.019***	0.381***
	[0.005]	[0.005]	[0.026]
Num. sewerage in construction	-0.006***	-0.008***	-0.048***
	[0.001]	[0.001]	[0.009]
Population density (sq kms)		-0.000**	0.002***
		[0.000]	[0.001]
Population		0.000***	-0.000***
		[0.000]	[0.000]
Total income (ln)		0.019	-0.176*
		[0.012]	[0.096]
Manages health centers		0.005	-0.040
		[0.007]	[0.090]
TA in formulation of investment projects		-0.011*	-0.007
		[0.006]	[0.084]
Internet access		-0.007	-0.466*
		[0.013]	[0.256]
Project-year	19106	19106	2488
Projects	4246	4246	1755
Districts	929	929	745

Note: The dependent variable in columns (1) and (2) is equal to 1 the period in which the project is completed. The dependent variable in column (3) is the duration of projects from start until completion. Sample in column (3) restricted to projects ever completed. All estimations include time-invariant project characteristics. Specifications in column (1) and (2) control non-parametrically for duration (i.e. periods between start and completion). All estimations include district fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

8.5 Effect of years exposed to construction

Table 18: First stage - Number of years exposed to construction

	(1)	(2)	(3)	(4)
Suitability index	0.471^{***}	0.388^{***}	0.454^{***}	0.373^{***}
	[0.064]	[0.063]	[0.061]	[0.061]
Fstat	54.15	37.63	54.01	37.08
Predicted cum. projects	3.846^{***}	3.105^{***}	3.772^{***}	3.021^{***}
	[0.428]	[0.455]	[0.420]	[0.447]
Fstat	80.75	46.63	80.63	45.73
Muni/project controls	No	Yes	No	Yes
Districts	952	952	1015	1015

Note: All regressions control for population density, total population and the number of sewerage projects started and include province fixed effects. Standard errors clustered by district in brackets. Statistical significance denoted by $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 19: Sewerage project abandoned in Piura with a completion rate below 60 perct.

Source: Picture taken in Piura from Google streets on 2013, the year the project was started.

Figure 20: Sewerage project abandoned in Huanuco

Source: Picture taken in Huanuco for the technical report of the Defensoria del Pueblo (2015) exploring mid-construction abandonement of sewerage projects.