Institute for Fiscal Studies

Antonella Bancalari

Can white elephants kill? Unintended consequences of infrastructure development in Peru

Working paper

20/32

Economic
and Social
Research Council

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

Antonella Bancalari [*]

September 25, 2020

Abstract

It is widely accepted that investing in public infrastructure promotes economic development. However, there is little awareness of the prevalence of unfinished infrastructure projects and their consequences. In this paper, I study the effect of unfinished sewerage infrastructure on early-life mortality in Peru. I compile several sources of administrative panel data for 1,400 districts spanning 2005–2015, and I rely on the budgetary plans and timing of expenditure for 6,000 projects to measure unfinished projects and those completed in a given district. I document that mid-construction abandonment and delays are highly prevalent. I exploit geographical features and partisan alignment to instrument for project implementation. Surprisingly, I find that unfinished sewerage projects increased early-life mortality, driven by lack of water availability, water-borne diseases and accidents. I also show that while unfinished projects pose hazards to the population, completed sewerage projects decrease early-life mortality, in line with public health studies in advanced economies during the previous centuries.

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 and Berkay Ozcan for their useful comments. I also thank participants at BREAD/CEPR/STICERD/TDC Development Economics, CSAE Oxford, RES, ESPE and PAA Conferences and at the workshops in numerous institutions such as LSE, IFS-UCL, Oxford University, EUDN Development Economics, NOVAFRICA-NOVA SBE and Universidad del Pacifico. All work and opinions in this paper are my own and do not necessarily represent the views of the Government of Peru nor other institutions. My work as a PhD student was supported by the London School of Economics Doctoral Scholarship.

1 Introduction

It is widely accepted that investing in large infrastructure promotes economic growth and development (Aschauer, 1989; Isham and Kaufmann, 1999). In fact, the World Bank directs 40 % of its lending portfolio to the development of large infrastructure in the water and sanitation, transportation and energy sectors as a means to alleviate poverty (World Bank, 2017).

However, to date, much more emphasis has been placed on the volume of infrastructure expenditure, rather than the quality of that expenditure (Besley and Ghatak, 2006). Recent evidence suggests that over one-third of the infrastructure projects started in low- and middle-income countries (LMICs) are not completed (Williams, 2017; Rasul and Rogger, 2018). Unfinished infrastructure projects are, however, not an exclusive problem of LMICs, as roads without tarmac and bridges to nowhere, for example, are commonly seen in advanced economies. Economic research has been very useful at identifying the effectiveness of infrastructure projects (e.g. sewerage, dams, roads and electricity networks) once they are completed and in use (Duflo and Pande, 2007; Dinkelman, 2011; Rud, 2012; Lipscomb et al., 2013; Alsan and Goldin, 2019; Donaldson, 2018; Banerjee et al., 2020). It is less clear what the consequences of such projects are while they are still unfinished (i.e. underway, delayed or abandoned half-way). It is regrettable that the literature has ignored the effects of unfinished infrastructure projects, given the important implications for a sound cost-effectiveness analysis.

In this paper, I seek to fill this gap in the literature. In particular, I study the effect of unfinished sewerage projects on the mortality rate of infants and children under the age of five (hereafter under-five) in Peru. This is the outcome that sewerage infrastructure has improved in advanced economies during the previous centuries (Watson, 2006; Alsan and Goldin, 2019). 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 invested three billion US dollars (USD) to start more than 6,000 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 spatial data at a grid-cell level. Specifically, I rely on detailed data on budgetary plans and the timing of expenditures to identify the number of unfinished projects and those completed in a given district. I exploit variation in unfinished projects generated by the high prevalence of mid-construction abandonment and delays in project completion. 60 % of the projects started between 2005 and 2015 were abandoned for at least one year and up to the whole decade of study. Moreover, I find large variation in project duration, with projects lasting for up to eight years, mostly because of cost overruns. Thus, districts have a combination of unfinished projects that have been abandoned (temporarily or indefinitely) and that are still underway (in time or delayed).

In order to deal with project placement bias, as richer districts with different mortality trends started and completed more projects, I rely on an instrumental variable strategy that exploits Peru's natural geographic variation. I use as an instrument a prediction of how the diffusion of sewerage would have evolved over time had project placement been based solely on cost considerations. I rely on the fact that a combination of geographic characteristics (i.e. land slope, elevation and river density) affects a district's technical suitability for low-cost sewerage projects. A time-variant project allocation is predicted with an algorithmic approach, subject to a nationwide budget constraint and maximum threshold allocation. The instrument predicts that a central planner would have allocated more projects to "cheaper" districts in terms of developing sewerage, and would have done so earlier in the period of study.

The identification assumption is that no other factors affecting mortality rates (e.g. a citizen's preference for preventive health care and other infrastructure and policies) changed over time along the same spatial lines as the predicted allocation of projects. The panel dimension of the data allows the inclusion of district and year fixed effects that control for time-invariant effects of geography on health and common shocks, respectively. A number of tests support the validity of my identification. I find that my instrument is not correlated with mortality before the start of projects. Furthermore, the results are not driven by other types of infrastructure development, geography-specific mortality trends or sorting.

I find that unfinished infrastructure projects — the so-called "white elephants" — can cause high social costs: they can kill children. With every additional unfinished sewerage project, infant mortality increased by 5 % and under-five mortality by 6 %, over the initial average mortality rate.

The mechanisms behind these non-trivial effects are threefold. First, water cuts are needed during the installation of sewerage lines. I find evidence that water and sanitation practices deteriorated as a result. While there is no effect on the connectivity to piped water, I find that an additional unfinished project increased the percentage of households relying on unsafe sources of water by 4 % over the initial averages. The limited access to safe water resulted in a decrease of the share of households relying on latrines and an increase in those practising open defecation, both by 10 % over the initial averages. Second, in order to install public sewers, extensive excavations are required, which leave open ditches that become filled with stagnant water and become pools of infections. Third, sewerage works pose hazards to the population. This entails large building sites that, for instance, divert traffic chaotically into previously quiet residential areas where children roam freely. In line with these mechanisms, I find that every additional unfinished project increased the infant and under-five mortality caused by water-borne diseases by 11 and 9.8 % from the initial rate, respectively. An additional unfinished project also increased the under-five mortality caused by accidents by 7.2 % from the initial rate. The results are consistent with the fact that older and more mobile children are more exposed to outdoor risks. Notably, I find no effects of unfinished projects on the mortality caused by other diseases and complications unrelated to infections or external hazards.

In order to get a full picture of project implementation and to understand better the counterfactual scenario, I also estimate the effect of completed projects. For a just-identified specification, I use as an additional instrument the interaction between a district's geographical suitability for low-cost sewerage projects and the partisan alignment between the district mayor and central government. Mayors politically connected to the Parliament are better able to secure funds to complete projects, conditional on starting them because of the district's geographic characteristics.

I find that early-life mortality increased with unfinished projects and decreased with completed projects, compared with no projects started. The estimated effect of unfinished infrastructure on mortality remains robust even after including project completion. Furthermore, infant and under-five mortality decreased with every additional completed project by 33 and 25 % over the initial averages, respectively.

Finally, I document that providing access to public sewers does not ensure a universal connectivity rate or sludge treatment, at least in the short run. This finding serves as evidence of the "last-mile" problem — the inability of governments to connect costly infrastructure to the final user (Ashraf et al., 2016) — and suggests that the social benefits from sewerage systems take time to be fully manifested.

The contributions of this paper are threefold. 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 and misallocation, highlighting the role of democratic institutions, political dynamics, governance structures and local managerial practices (Bandiera et al., 2009; Burgess et al., 2015; Williams, 2017; Rasul and Rogger, 2018). However, there is a need to gain a better understanding of how inefficiencies in the provision of public goods jeopardise economic development and well-being. For example, Burgess et al. (2015) acknowledge this need in the context of a misallocation of public resources in Kenyan road building, where they quantify the extent of ethnic favouritism and document how it disappears during periods of democracy, stating that: "linking [our] findings to aggregate economic outcomes represents a key priority for future research".

Second, this paper contributes to the literature on large public infrastructure effectiveness by extending the scope of analysis to the potential risks generated by projects that are still in progress or abandoned. 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 (Duflo and Pande, 2007; Dinkelman, 2011; Lipscomb et al., 2013). The literature also provides evidence that transport infrastructure increases productivity, inter-regional trade and welfare (Donaldson, 2018; Banerjee et al., 2020). More closely related papers find that environmental hazards from large infrastructure affect early-life mortality (Cesur et al., 2017; Gupta and Spears, 2017; Mettetal, 2019).

Finally, this study informs the literature on public health, which has mainly focused on water technologies (Cutler and Miller, 2005; Bhalotra et al., 2018), by exploring the effects of sewerage at scale in a contemporary setting (Watson, 2006; Kesztenbaum and Rosenthal, 2017; Alsan and Goldin, 2019). Recent studies in LMICs have mainly focused on the effectiveness of private sanitation infrastructure (Geruso and Spears, 2018) or have provided evidence from experimental studies with a limited time-horizon and geographical setting (Duflo et al., 2015). My study, by contrast, focuses on a nationwide setting and a longer temporal focus.

The rest of the paper proceeds as follows. In Section 2, I provide the context. I explain the data and present descriptive statistics in Section 3. In Section 4, I provide details of the instrumental variable strategy. In Sections 5 and 6, I present the results of the effect of unfinished and completed projects, respectively. In each of these sections I describe the mechanisms driving the results. I conclude in Section 7 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 (World Bank, 2020). To remedy this, the National Sanitation Plan 2006–2015 set the goal of increasing access to sewerage in urban areas, representing the first nationwide effort towards sewerage diffusion in Peru. In this period, the Government of Peru invested more than USD 3 billion to start 6,090 sewerage projects[1] in 80 % of the districts.[2]

The roll-out of sewerage projects across districts was not random. The starting of sewerage projects depended on two crucial factors: (i) the willingness and capabilities of the implementing agent; and (ii) the allocation of funds.

Between 2005 and 2015, most projects were implemented by local municipalities: more than 56 were implemented by district municipalities and almost 30 % by province municipalities (see Figure A1, Panel A). District municipalities can implement sewerage projects if they are incorporated into the National System of Public Investment (SNIP, Spanish acronym), which requires the following: (i) access to the 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 the Internet by 2005 started a greater number of sewerage projects (see Figures A2, Panel A and B, respectively).

For the portfolio of projects implemented by the Ministry of Sanitation, the National Sanitation Plan 2006–2015 states that previously unattended and poor areas should be prioritised when expanding access to sewerage. This was not the case as 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 (see Figures A2 Panel C and D, respectively).

In addition, sewerage diffusion depends on the cost of implementing a given project. The National Sanitation Plan 2006–2015 states that projects must achieve economic and technical viability to be implemented, which depends crucially on project costs. Projects using cheaper technologies are more likely to be declared viable. This criterium is crucial for the instrumental

---

[1] Out of these, 4,783 were construction and expansion of new systems and 1,307 were improvement of existing lines.

[2] 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.

variable strategy, explained in the next section.

Sewerage diffusion, and more specifically the completion of projects, depends on funds allocation. The largest sources of funding were transfers from the central government: 40 % of sewerage projects were funded by royalties and 30 % by direct transfers (see Figure A1, Panel B). District municipalities do not have full discretion over the use of these funds. In the case of royalties, for instance, funds can only be used in social infrastructure. Only 22 % of started projects were funded by local tax revenue, and municipalities have more discretion over the use of this revenue.

The allocation of funds to projects is conducted by an annual budgeting process in which agents with different incentives interact. Understanding these interactions is important for the instrumental variable strategy used in Section 6. For projects financed by the central government (executed directly by the Ministry of Housing, Construction and Sanitation or through transfers to local municipalities), funds are allocated through an annual budgeting process approved by the Parliament. For projects financed by local revenues, funds are allocated from the budgeting process done by Municipal Councils, which are chaired by the mayor and council members. Given that most sewerage projects are implemented by the local municipality, but financed by the central government, partisan alignment between local majors and members of the Parliament makes it easier to attract funds to complete projects.

Once projects are selected for funds, the government agency that formulates the project starts the procurement process to hire private contractors to develop the works. During the construction phase, the Enterprises of Provision of Sanitation Services (EPS) are in charge of supervising and evaluating the technical quality of sanitation works in urban areas. Once public sewers are installed, it is compulsory for landlords to connect the dwelling's waste-water pipes to the public sewerage lines. The EPS are in charge of regulating and supervising the connectivity of dwellings to the public sewerage lines. Understanding the limitations of the work conducted by the EPS will be crucial to understand the mechanisms behind the results of this paper. These limitations are discussed in Section 6.1.

3 Data and descriptives

3.1 Data

I construct a district-level panel data set of more than 1,400 districts in Peru from 2005 to 2015 by combining data from several novel sources. I compute infant and under-five mortality using vital statistics registries and population forecasts. For the core data set measuring sewerage diffusion, I compile and combine project-level data from viability studies and annual budget reports, which allows me to identify unfinished projects and those completed. To construct the instrumental variable, I use spatial data at grid-cell level, 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 data set is an unbalanced panel of 1,408 districts spanning 2005–2015, with a total of 10,494 district–year observations.

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, 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 the number of deaths of infants (under one year old) and children under five years old. The mortality data are disaggregated by cause of death following the International Classification of Diseases – ICD10. The population forecast provides data on the number of children under five years old. I construct the infant mortality rate (IMR) and the under-five mortality rate (U5MR) for each district d and year t, using as the denominator the population at risk, as described by Preston et al. (2001):

$$IMR_{dt} = \frac{\text{Deaths of infants aged 0–11 months}_{dt}}{\text{Population aged 0–59 months/5}_{dt}} \times 1,000;$$

$$U5MR_{dt} = \frac{\text{Deaths of children aged 0–59 months}_{dt}}{\text{Population aged 0–59 months}_{dt}} \times 1,000.$$

The IMR is generally computed as the ratio of infant deaths over live births. However, because of the incompleteness of birth registries in Peru, where the coverage was 93 % by 2005 (UNICEF, 2005), I use an alternative approach. I use as a denominator the total population of children aged between 0 and 5, divided by 5 (assuming that the distribution across ages is similar).

To alleviate concerns linked to the quality of the vital registers in Peru, I compare nationwide mortality trends using the vital statistics data versus data from several nationally representative surveys. I find that vital statistics generate mortality rates that are slightly lower in level, but the trends do not differ greatly (see Figure A4).

To measure sewerage diffusion, I use raw data from viability studies registered in the SNIP and budget reports from the Integrated System of Financial Administration (SIAF, 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 year 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 — including cost updates — is accrued by at least 90 %. The Ministry claims that, at this level, construction works are completed (i.e. excavation works finished and open ditches closed) and the last leg consists of paperwork. I set the years in which projects are unfinished as the years between start and completion. Projects without a completion year but with a start year are defined as unfinished until the end of the study period.

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) the cumulative number of sewerage projects started; (ii) the number of unfinished sewerage projects; (iii) the 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 or geo-codes) and there are no early-life mortality data at the same level. For projects formulated at a higher governmental level that lacks data on the number of projects per district, I assign one project to each district within the corresponding province or region. This approach does not capture the intensity of sewerage diffusion within each of the districts, but it is done in only 3.7 % of the districts that ever implemented projects.

I use spatial data provided by the Ministry of Environment to compute geographic characteristics influencing the cost of sewerage development. I rely on these data to construct an instrumental variable. The spatial data include information on surface elevation for multiple cells (1×1 km^2), which I match to district boundaries in 2015. I construct indicators for four main geographical characteristics: elevation, gradient, area and river density. First, I compute the total area within the boundaries of each district. Second, 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] metres above mean sea level (mamsl), {250–500] mamsl, {500–1,000] mamsl and above 1,000 mamsl. Third, I compute gradient using surface elevation at each cell and neighbouring cells. I construct indicators capturing the fraction of district area falling into four gradient categories: [0–0.8] %, {0.8, 4.19] %, {4.19–13] % and above 13 %. The first category captures flat areas below or equal to 0.8 % 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 the capture of differences in elevation and gradient within districts (compared with, say, using the mean per district). Finally, I compute river density as the fraction of the district area that falls in inland waters. The maps shown in Figure A6 show that districts in Peru vary greatly in their ruggedness, altitude and river density. I draw on data from the National Register of Municipalities (RENAMU, Spanish acronym) to measure municipal characteristics. As explained in Section 2, only districts that had access to the Internet, numerous 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 obtain reports concerning whether water and faecal sludge is treated in the district. I use these variables to explore whether sewerage diffusion had any effect on the removal of bacteria and contaminants from the sources of drinking water and waste water. Data on the treatment of water are available only between 2008 and 2014, and data on the treatment of sludge are 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 these data to evaluate whether sewerage diffusion increased the percentage of households connected to the public sewers. I also use these data to compute the percentage of households that have a head of household who attained education above the secondary school level and the percentage of households that are connected to the electricity network in each district. These variables are alternative outcomes used to evaluate whether 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. These data are available at the district level between 2007 and 2014 (2015 only available for transport expenditure).

3.2 Descriptive statistics

Between 2005 and 2015, both infant and under-five mortality fell by 35 % (see Table A1 in the Appendix for descriptive statistics for the beginning and end periods of analysis). Both early-life deaths and the population of children under the age of five decreased, but the decrease in the number of deaths was greater. Meanwhile, the number of started and completed sewerage projects grew dramatically.

Municipalities became richer during the period of study. The average revenue of a district municipality quadrupled — from 4 million to 15 million soles ($\sim$ USD 4.5 million) — and many municipalities gained access to the Internet. The share of municipalities registered as requiring technical assistance for the formulation of investment projects decreased, while those managing a health centre increased.

Districts improved their access to public services greatly in the decade of analysis. Water connectivity and treatment increased, while the share of households relying on unsafe sources of water decreased. As expected, sewerage connectivity and treatment increased, as well as the share of households relying on on-site sanitation increased, while those practising open defecation decreased. Districts also improved regarding the share of households that had heads who had completed secondary education and households that had electricity connectivity. Furthermore, public expenditure increased over the period of analysis in the transportation, energy and health sectors.

Peru has a great geographical diversity, which I am able to exploit in my instrumental variable strategy. On average, the largest share of area of districts falls in the highest elevation category (74 %), followed by the lowest category (15 %) and all categories have a relatively high standard deviation (20 %). Districts in the sample tend to have rugged terrains. The lowest share of area, on average, falls in the flattest gradient category (only 10 %) and the largest share in the steepest category (37 %). River density is, on average, 53 km per km^2 and there is great variation across districts (124 standard deviations).

3.3 Project characteristics

Two factors are linked to the variation over time in the number of unfinished projects: mid-construction abandonment and project duration. First, there is a high prevalence of projects that stopped receiving funds while they were still underway. Figure 1, Panel A, shows the distribution of the number of years that a project was abandoned. Strikingly, more than 75 % of the started projects in the period 2005-2012 were "white elephants" at least one year. There is large variation in the number of years that projects were abandoned, ranging from two year to indefinitely. While only half of the projects started between 2013 and 2015 were abandoned at some point, one could argue that it is only a matter of time for these newer projects to become "white elephants". [3].

Second, there is great variation in the time to complete projects. I find that half of projects took more than one year to be completed (see Figure 1, Panel B). As expected, larger projects, proxied by the number of potential beneficiaries, take longer to be completed. However, even amongst larger projects, half took three years or more (up to eight years) to be completed. This variation in project duration, even after taking into account project complexity, suggests that delays are common. The prevalence of cost overruns serves as additional evidence in support of delays. Figure A3 shows that larger projects have greater cost overruns, as high as five times the planned cost. Only 5 % of large projects had no cost overrun. Bureaucratic procedures to update costs can delay project completion.

The measure of unfinished projects is thus a combination of projects still underway (on time or delays) and abandoned (temporarily or indefinitely) in a given district.[4] Between 2005 and 2015, on average, districts started four sewerage projects. Strikingly, by 2015, districts completed fewer than one project, on average. The low rate of completion results in districts having, on average, more than one unfinished project between 2009 and 2012 and more than two unfinished projects in later years (see Figure A5 in the Appendix).

4 Empirical strategy

In order to understand the consequences of unfinished sewerage projects on early-life mortality, I rely on an instrumental variable approach.

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. I exploit the fact that

---

[3] There is no difference in the prevalence of mid-construction abandonment across small and large projects

[4] Although the majority of projects are "white elephants" or on the verge to be, I would ideally disentangle the effects of a project underway versus one that was abandoned. Because of the aggregate nature of the mortality data, I would have to focus on districts with only one project being developed. Unfortunately, I do not have the statistical power in this paper to conduct such an analysis. By 2015, only 20 % of districts have started only one project, equivalent to 2,069 district–year observations. The statistical power is reduced even further if I focus on districts developing only one project in previous years.

a combination of geographic characteristics (i.e. elevation, land gradient and river density) affects the suitability of districts to low-cost sewerage projects. I use an algorithmic approach to generate variation over time in predicted sewerage diffusion, subject to a nationwide budget constraint and a threshold of maximum project allocation.

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 affect 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. The panel dimension of the data allows the inclusion of district and year fixed effects that control for time-invariant effects of geography on health and common shocks, respectively. Lipscomb et al. (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.

Relying on the technical suitability of a district makes the instrument comply with the monotonicity assumption. While the instrument may have no effect on the launch of sewerage projects in some districts — that is, 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 on. It is sensible to assume that no district decreased its likelihood of experiencing sewerage diffusion by being more technically suitable (defiers).

The predicted sewerage diffusion is constructed following three steps.

(1) **District's technical suitability for low-cost sewerage projects**

For each district, an index is constructed capturing the 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 cost of developing sewerage infrastructure is affected by a unique combination of geographic 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 waste water 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 the costs even further (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). Elevation above the level of the sea is another topographic factor that affects districts' suitability for low-cost sewerage projects. The cheapest waste-water 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 high-altitude areas. 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 that cover large areas of land requires the installation of 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-squares (OLS) regression:

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

Here, S_d is the total number of started projects 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), R_d is the district's river density (river length in km per area in km^2) and A_d is the total area of land within district boundaries.

Table A2 in the Appendix shows that, as predicted by the engineering literature, steep gradient categories and river density favour sewerage diffusion, while elevation and district area are negatively associated with project placement. Steep gradient and elevation predicts the allocation of sewerage projects non-monotonically: the largest coefficient is the lower-middle ({0.8, 4.19] %) gradient category and the highest elevation category (above 1,000 mamsl).

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.

(2) **Nationwide budget as a constraint**

The nationwide budget for projects to construct new sewerage systems and to expand and improve existing 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 nationwide budget for sewerage projects increased year to year and this generates variation over time on the expenditure on sewerage projects. To get an idea of the over-time variation in budget spent, see Figure A7.

(3) **Time-variant allocation of projects**

The final phase consists of an algorithmic approach to construct a time-variant instrument.

> Ranking all districts in Peru based on the technical suitability index, the algorithm predicts how a central planner would allocate one project to each district until the nationwide budget is exhausted (considering the average cost of a sewerage project). The highest-ranking districts are forecast to receive sewerage projects earlier and with more projects across the years. For instance, for 2005, the prediction allocates one project for each of the 20 highest-ranking districts because the budget spent that year amounts to the average cost of 20 projects. The prediction follows the same procedure for the following years until a district receives a maximum of five projects, which is the median of the distribution of projects allocated to districts that developed sewerage between 2005 and 2015. This threshold of maximum project allocation leaves extra generation capacity that is subsequently relocated to other districts further down the ranking. 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 five sewerage projects, while the lowest-ranked districts would have received none. This creates an allocation roll-out that provides variation across districts and years.

Description of the instrumental variable

Figure 2 shows a map of Peru, plotting the diffusion of sewerage from 2005 to 2015. The early development of sewerage projects was focused on the affluent and populous north coast as well as on the relatively less affluent centre region of the Andes. The intensity of sewerage diffusion increases in these regions and expands eastward every year, until the Amazon region is covered. By 2015, there is great variation in the number of sewerage projects across districts. The regions that experienced relatively lower diffusion of sewerage are the north-east region of the Amazon and the south of Peru.

Figure 3 plots the districts predicted to receive sewerage projects by year. Between 2005 and 2015, districts were predicted to receive up to five projects. Water-rich areas with steeper gradients and lower altitudes are predicted to receive sewerage infrastructure earlier, but the dynamics are mediated by the budget constraints and the restriction that districts that received five projects in previous years do not receive more projects. Ignoring the demand-side drivers of sewerage diffusion forces the prediction to over-allocate projects to unattended places, such as the north-east Amazon area and the south coast. This weakens the relevance of the instrument, but allows the extraction of exogenous variation linked to geographical characteristics. The strength of the spatial correlation between Figures 2 and 3 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.

4.2 Empirical model

I estimate the effect of unfinished sewerage projects on the IMR and U5MR rates between 2005 and 2015 relying on variation in the intensity of sewerage projects across districts and years and using predicted sewerage projects as an instrument. The instrumental variable strategy corrects for the bias introduced by the endogenous placement of projects. To formally evaluate the relationship between actual and predicted projects, I estimate the following first-stage regression:

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

Here, S_{dt} denotes the number of unfinished sewerage projects and Z_{dt} is the number of projects predicted 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 the IMR and U5MR using the following two-stage least-squares (2SLS) model:

$$MR_{dt} = \alpha_2 \hat{S}_{dt} + \gamma_2 d + \delta_2 t + \xi_{dt}. \tag{3}$$

Here, MR_{dt} denotes infant ($1q_o$) or under-five ($5q_o$) mortality rates and $\hat{S_{dt}}$ is the instrumented number of unfinished sewerage projects 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 zero to one project, from one to two 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 realised 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 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. Standard errors are clustered at the district level to deal with serial correlation due to the panel characteristics of the data and the fact that the intra-cluster correlation is lower within higher spatial levels.

Table 1 shows that the predicted sewerage diffusion is a relevant instrument for the number of unfinished sewerage projects. This table presents the first-stage results, where the dependent variable in column (1) is the number of unfinished sewerage projects. I find that, on average, an additional project predicted to be allocated in a district is associated with 0.4 unfinished projects. The Sanderson–Windmeijer F-statistic of excluded instruments is high and above the rule of thumb of Stock and Yogo (2002) (an F-statistic equal to or higher than 10), which confirms the relevance of the instrument.

In support of the identification assumption, columns (2) and (3) show that the instrument is not associated with infant and under-five mortality before the start of sewerage projects. The dependent variable in columns (2) and (3) is the infant and under-five mortality rate, respectively. I find that, on average, an additional project predicted to be allocated in a district has no effect on infant or under-five mortality in the years prior to the start of the first sewerage project.

5 Effect of unfinished projects on early-life mortality

The main result of this paper is that unfinished projects increased early-life mortality. Table 1 presents the estimated effect of the number of unfinished sewerage projects on a district's IMR and U5MR. Columns (4) and (5) show OLS estimates and columns (6) and (7) show 2SLS estimates. All specifications include district and year fixed effects. Both the OLS and 2SLS estimates show that sewerage diffusion increased the IMR and U5MR, though the 2SLS estimates are larger in magnitude.

On average, an additional unfinished sewerage project increased the IMR by 0.001 deaths per 1,000 infants and the U5MR by 0.299 deaths per 1,000 children. These results translate into a 5 and 6.2 % increase, respectively, from initial average mortality rates.

Figures 4, which plot the mortality trends of districts predicted and not predicted to receive projects by the instrument, provide three insights. First, in support of the identification strategy, infant (Panel A) and under-five (Panel B) mortality trends are parallel before the start of the very first sewerage project. Second, after the start of the first project, mortality increases. Third, over time, early-life mortality decreases at a slower rate in districts that started a sewerage project because they were predicted to (i.e. compliers),[5] compared with districts that started a project although they were not predicted to (i.e. always-takers).

The fact that the mortality of "always-takers" decreases at a steeper rate after the start of the very first project is evidence that these districts were better able to mitigate hazards during the construction works and to take advantage of the social benefits of sewerage infrastructure. This explains partially why the OLS estimates are larger than the 2SLS estimates. The compliers in the instrumental variable strategy (based on a district's technical suitability for low-cost sewerage projects) are different from the average district whose placement of projects was affected by

---

[5] Compliers are also those that did not start a project because they were not predicted to, captured by the blue dot in the red line.

socio-economic and political considerations or other demand-side factors. “Always-takers” are likely richer districts, better politically connected and with greater willingness to improve living standards.

The OLS downward biased estimates also reveal the expected project placement bias, as richer municipalities with lower mortality experienced greater diffusion. Finally, the 2SLS estimates are larger than the OLS estimates likely because the 2SLS model corrects measurement error. While the actual number of unfinished projects constructed using a combination of administrative records likely suffers from classical measurement error, the geographical variables used to predict the placement of projects are measured quite precisely (based on 1×1 km^2 satellite maps). The 2SLS model may be addressing the associated attenuation bias.

5.1 Robustness checks

A variety of checks bolster the robustness of the main results. I estimate the 2SLS model with district and year fixed effects with a series of modifications.

First, I control for time-varying lagged population density. This addresses the concern that the instrument may be capturing variation in population density.

Second, I control for municipal characteristics that were correlated with actual sewerage diffusion (as discussed in Section 2). These include indicators for whether the district municipality has access to the Internet and needs technical assistance to formulate investment projects and municipal revenue, in order to control for public investment capabilities. I also add as a covariate an indicator for whether the municipality manages at least one health centre, in order to control for political will on health policy. If the instrumental variable strategy is as good as random when predicting unfinished projects, then I expect that controlling for these factors will affect the point estimates only slightly.

Third, I add an indicator for whether the district is located in the Amazon region, given that peculiar factors of this area could be driving the results.

Fourth, I restrict the sample of analysis to districts that started at least one sewerage project, in order to make the sample of study more comparable. Also, this test clarifies the counterfactual scenario better: the effect of more versus fewer unfinished projects, as opposed to also considering as counterfactual starting no projects.

Fifth, I exclude the capital and main province of Peru, Lima, to check that this different region is not driving the results.

Moreover, 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. Finally, I replace the independent variable with one capturing unfinished project density, measured as projects per 10,000 people per km^2. This transformation helps us to understand the extent to which population density is a mediator of the effect.

Table 2 shows the different robustness checks (or specifications) in each row. The magnitude and precision of the estimated effect of unfinished projects on IMR (column 1) and U5MR (column 2) remain robust and highly significant. The Sanderson–Windmeijer F-statistic of excluded instruments (column 3) remains similar in most cases (it drops to 8 for the project density transformation).

5.2 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 early-life mortality only through actual sewerage diffusion. In this section, I provide evidence that supports the validity of the exclusion restriction and, hence, 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. 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). Furthermore, 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. Another concern could be if investing in sewerage systems crowds out investment in other type of infrastructure beneficial for early-life health.

To alleviate these concerns, I first control for district expenditure on transportation, energy and health. Next, I explore if the alternative infrastructure investments can 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, whether the predicted variation can explain the increase in mortality rates.

Table 3 presents the estimates of the effect of unfinished projects on IMR (column 1) and U5MR (column 2) when controlling for expenditure in transport, energy and health projects (specifications 1–3). This exercise confirms the main results: the magnitude of the estimates remain similar. The Sanderson–Windmeijer F-statistic of excluded instruments (column 3) also remains similar

Table 3 also presents 2SLS estimates of transport, energy and health expenditure on early-life mortality rates using the predicted sewerage diffusion as an instrument (specifications 4–6). None of the three alternative infrastructure developments explains the estimated effects in mortality. The transportation and energy expenditure channels are not statistically significant. If anything, the health expenditure channel has a negative effect on early-life mortality. Because this effect is opposite to the one estimated, if anything my results would be downward biased. Yet, in all cases, the first-stage is weak, as shown by the low Sanderson–Windmeijer F-statistic of excluded instruments (column 3).

Another concern would be if the instrument is capturing variation driven by specific geographic characteristics or regions with greater suitability for low-cost sewerage projects. In Table 4, I test the robustness of the estimated effect of unfinished projects on early-life mortality when controlling for geography-specific trends and interactions with the annual budget. I include as controls the following components interacted with year and annual budget: the flat gradient category (specifications 1 and 5); the low elevation category (specifications 2 and 6); the district area in km^2 (specifications 3 and 7); an indicator for the Amazon region (specifications 4 and 8); and population density per km^2 (specifications 9 and 10). The estimated effects of unfinished sewerage projects on IMR (column 1) and U5MR (column 2) remain robust. The different specifications also have little effect on the first-stage power (column 3), in some cases even increasing it (as with population density controls). When controlling for elevation-specific trends and its interaction with nationwide budget, the magnitude remains similar, but the precision and F-statistic of the 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 the possibility of my instrument being correlated with the distribution of rural population across districts. Because the instrument is computed using geographic factors, such as gradient and elevation, which 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 might attract households with farming as their main occupation. This sorting could explain the main results as rural life has long been associated with higher mortality rates (Hathi et al., 2017). Figure A8 shows that, while the actual sewerage diffusion is correlated with the percentage of rural population (upper plot), this is not the case for predicted sewerage diffusion (lower plot). Districts with a percentage of rural population above the median by 2005 have an identical distribution of predicted sewerage projects as those with a percentage of rural population below the median.

5.3 Mechanisms

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

I first investigate whether sewerage diffusion affected early-life mortality rates through systematic demographic changes. The observed increase in mortality rates could be a result of a decrease in the denominator, namely the number of infants (IMR denominator) and the number of children aged under 5 (U5MR denominator). For instance, a decrease in births and population could be a result of families moving away from disruptive infrastructure works. I find that this is not the case (see Table A3 in the Appendix, columns 1 and 2): the estimated effects on live births and the under-five population go in the opposite direction. The coefficients of the effect of unfinished projects on early-life mortality are, if anything, underestimated. The increase in the under-five population could be explained by the increase in mortality, as the death of a young child may motivate families to have more children in order to achieve their desired fertility.

Another channel explaining the estimated positive effect on early-life mortality is selective emigration of the most well-off households and immigration of poorer households. Disruptive sewerage works may create incentives for well-off households to move away, reducing housing prices and rent and hence attracting poorer households. I find no evidence of sorting across districts (see Table A3 in the Appendix, columns 3 and 4). The effect of sewerage diffusion on the number of household heads with completed secondary education is not statistically significant. There is a negative and statistically significant effect on households that have electricity connectivity, but this could be because the results are restricted to a small subsample (data are only available for 50 % of the districts of analysis and for two years). Table A4 alleviates concerns that the results may be driven by education and electricity trends picked up by the instrument: the estimates remain robust when controlling for education and electricity-specific trends.

Next, I argue that the main mechanisms behind the estimated increase in mortality are linked to the disruptions posed by the construction works to install sewerage lines. Interviews with local engineers reveal that water cuts are needed in order to install sewerage pipes. Cases of unfinished sewerage projects leaving the population without access to piped water have attracted media attention (RPP Noticias, 2018). I find evidence that piped-water cuts affected the water and sanitation behaviour in affected districts.

Table 5 shows the coefficients of a 2SLS model of the effect of unfinished projects on water and sanitation practices. The dependent variables in columns (1)–(5) are, respectively, an indicator capturing whether the district has high connectivity to piped water (between 75 and 100 %), an indicator capturing whether the municipality treats water, the share of households that rely on unsafe sources of water, the share of households that use a latrine and the share of households that practise open defecation. Although, as expected, there is no effect on the connectivity to piped water, I find a negative effect of the likelihood of the municipality treating the piped water to make it safe (though not statistically significant). Notably, I find that an additional unfinished project increased the percentage of households relying on unsafe sources of water by 3 percentage points (ppts), which translates into a 4 % increase over the initial average. The limited access to safe water resulted in a decrease in the share of households relying on latrines by 0.04 ppts and an increase in those practising open defecation by 0.05 ppts. These results are exactly opposite and equivalent to a 10 % change over the initial average.

Further disruptions are linked to the excavation works. Open ditches required to install sewerage pipes pose a number of hazards to children. [6] Environmental dangers documented in Peru are linked to dust particles, stagnated ground water that creates sources of vector-borne diseases and the use of ditches as landfill sites (Malpartida Tabuchi, 2018). Shockingly, there is evidence of children falling and drowning in ditches from sewerage works that were as deep as 2 m, became 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 quiet residential areas. An interview with an engineering expert on the implementation of sewerage projects disclosed that contractors frequently divert traffic in an unorganised matter (i.e. failing to put in place effective signaling systems), which leads to traffic accidents.

---

[6] Figures A10 and A11 show how sewerage works look while underway and abandoned, respectively. Both show how a sewerage project underway leaves equally dangerous open ditches as one abandoned.

Table 6 investigates the effect of unfinished sewerage projects on different measures of mortality depending on the diseases and health-related problems that caused the death. Mortality data are disaggregated for general pathological groups following the World Health Organization's International Classification of Diseases (ICD 10). The outcome in the first row 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 the second row is the mortality rate linked to external causes (ICD-10 category XX), which mostly includes deaths caused by falls, drowning and traffic-related accidents. The following rows estimate the effect of sewerage works on deaths unrelated to sanitation and external hazards. The outcome in the third row is the mortality rate resulting from diseases of the respiratory system (category X) and the fourth row shows the mortality rate due to congenital malformations (ICD-10 category XVII). The outcome in the last row is the mortality rate linked to other unrelated factors, including diseases of the nervous system (ICD-10 category VI), circulatory system (ICD-10 category IX) and neoplasms (ICD-10 category II).

I find estimates in line with unfinished projects affecting mortality due to hazards from the excavation works, in addition to potential infectious diseases from deteriorations in water and sanitation behaviour. An additional unfinished sewerage project increased the mortality caused by water-borne diseases by 0.001 deaths per 1,000 infants and by 0.2 deaths per 1,000 children (11 and 9.8 % increases from the initial rate, respectively). Furthermore, an additional unfinished project increased the U5MR caused by accidents by 0.09 deaths per 1,000 children (7.2 % increase from the initial rate). Both infants and children are exposed to infectious diseases, directly as a result of the pools of infection that open ditches become, or indirectly because of the greater use of unsafe sources of water and the increase in faecal exposure. As expected, there is no effect on the infant mortality caused by accidents, as only older children are exposed to outdoor hazards.

If my estimates are well identified, then only an increase in mortality caused by pathogenic infections and accidents would be observed. In line with this prediction, I find no statistically significant effect of unfinished sewerage projects on mortality caused by other diseases or unrelated to external hazards from the construction works. Encouragingly, this means that the instrumental variable methodology is not picking up a general difference in mortality trends by all causes.

6 Effect of completed projects on early-life mortality

In order to get a full picture of the project implementation and to understand better the counterfactual scenario, I additionally estimate the effect of completed sewerage projects. It is necessary to consider project completion, given its potential confounding effect. On the one hand, one may expect the social benefits of sewerage systems to manifest upon project completion. On the other hand, mortality might not decrease if users do not connect to the infrastructure, and it might even increase if systems become a collection of sludge that contaminates the environment due to unsafe disposal.

To estimate both the effect on early-life mortality from unfinished sewerage projects and those completed, I use two instruments. The first instrument is the low-cost prediction of sewerage diffusion used in the main analysis. The second instrument is the interaction between the geographical suitability for low-cost sewerage projects with an indicator capturing partisan alignment between the municipal mayor and the central government.

I define partisan alignment as the case when the district mayor is from the same political party as the party forming the Parliament. In Peru, there is a great percentage of municipal mayors whose affiliation is to a new political party or an independent movement that has no representation at the central level. Given that there were three municipal elections and two central elections for the Parliament and President, there is variation over time in the percentage of districts aligned (see Figure A9).

Table 7 presents the first-stage results. The dependent variables in columns (1) and (2) are, respectively, the number of unfinished projects and the number of projects completed. An additional predicted sewerage project increases the number of unfinished projects by 0.26 and those completed by 0.11. This result corroborates the fact that completed projects confound the effect of unfinished projects, as my original instrument predicts both unfinished and completed projects.

Notably, the geographic suitability for low-cost projects increases by 0.88 the number of projects completed in districts with partisan alignment. Mayors politically connected to the Parliament are better able to secure funds to complete projects, conditional on starting them due to the district's geographic characteristics. This interaction has no statistically significant effect on the number of unfinished projects.

The first-stage is weak, but there is no concern with this generating a bias. Following the recommendation of Sanderson and Windmeijer (2016) for applied work with multiple endogenous variables, I report the Kleibergen–Paap rk Wald F-statistic (a robust version of the Cragg–Donald statistic) and the Stock and Yogo (2002) weak ID test critical values. The latter essentially tests if the bias of the instrumental variable estimator (IV), relative to the bias of OLS, could exceed a certain threshold. For example, if one were willing to tolerate a maximal size of 15 %, the size of the IV–OLS distortion would be 10 % for the 5 % level test. The 10 % maximal IV size for my instrumental variable estimation just identified is 7.03. Given that the Kleibergen–Paap rk Wald F-statistic is less than all critical values, the instruments are weakly identifying the number of unfinished and completed projects. Yet, the estimated first-stage coefficient above 0.1 and the exactly identified model alleviate concerns linked to the low F-statistic generating a bias in the 2SLS coefficients (Bound et al., 1995).

The omission of completed projects generates a downward bias of the estimated effect of unfinished projects. Table 7 also presents the effect of unfinished projects and those completed on IMR (columns 3 and 5) and U5MR (columns 4 and 6). The 2SLS estimates reveal the expected results. While an additional unfinished project increased mortality, an additional completed project decreased it, compared with not starting a project. Although I am unable to estimate statistically significant effects here, this exercise serves as a “sanity check”. Once completed projects are included in the estimation strategy, the effect of unfinished projects is slightly larger than the original estimation.

The magnitude of the negative effect of a completed project is greater than the positive effects of an unfinished project; that is, an increase of 0.004 infant deaths versus a decrease of 0.006 infant deaths and an increase of 0.862 child deaths versus a decrease of 1.214 child deaths. In line with my main hypothesis, early-life mortality increased during the construction phase but these unintended consequences dissipate once projects are completed (e.g. when water supply is resumed and open ditches are closed).

Given that the interaction between the geographical suitability for low-cost sewerage projects and partisan alignment only predicts completed projects, we could use this as an instrument in a specification where we exclude unfinished projects. The results of this alternative specification are shown columns (5) and (6). The estimated effect of completed projects on mortality remains robust, though slightly lower in magnitude. The Sanderson–Windmeijer F-statistic is now higher (3.62) and in the margin of the 25 % maximal IV size. Hence, I assume a size distortion (bias of the IV estimator related to the OLS) of 20 % for the 5 % level test.[7]

6.1 Mechanisms

Even when projects are completed, the health benefits associated with sewerage systems may not fully materialise in the short run for two main reasons. First, if less than universal connectivity is achieved, then this means that neighbours are still contaminating the environment. There are negative externalities from using rudimentary sanitation prone to leakages (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, which is known as the “last mile problem” (Ashraf et al., 2016).

Second, even if universal connectivity is achieved, untreated faecal sludge can contaminate bodies of water used for drinking or irrigation purposes. A study has revealed that in Latin American, particularly in Peru, only about 30 % of waste water is treated, with the remaining sludge being discharged in open waters (Fay et al., 2017).

The sustainability of sewerage systems depends on the effectiveness of government agencies to operate and maintain the systems. A diagnosis 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 % perform poorly, measured by transparency, customer support, institutional management, financial and operational sustainability and work environment (Von Hesse, 2016).

Although the Peruvian norm establishes that it is compulsory for landlords to connect house-

---

[7] The results are not statistically significant likely because this study does not have the statistical power to estimate the effects of completed projects. Recall from Figure A5 that, on average, a district completed only one project over ten years. The lack of variation in the intensive margin restricts the analysis of the effects of completed projects. However, the purpose of this paper is to fill the gap in the literature on the effects of unfinished projects, rather than completed.

holds to public sewers when available, the enforcement of this norm is weak (Von Hesse, 2016). Furthermore, there is evidence that the bad performance of public firms leads to inoperative treatment plants, which contaminate local sources of water and agricultural fields, and to a deterioration in the environment, which causes disease (Vega Ysela, 2015).

To quantify the extent to which sewerage diffusion was accompanied by an improvement in the operation of sewerage systems, I use census data on the percentage of households connected to sewerage (connectivity) and municipal reports indicating if water and sludge is treated (treatment).

I estimate the effects of completed projects (using the same 2SLS specification as in Panel B of Table 7, due to the higher first-stage F-statistic) on sewerage connectivity and the likelihood of treating water and sludge. I find that an additional completed project increases connectivity by 23 ppts and sludge treatment by 15 ppts (see Table A5 in the Appendix), though the effects are not statistically significant[8]. Although public sewers are introduced, the district's average connectivity rate and prevalence of sludge treatment are still less than universal (i.e. 46 and 39 %, respectively). While a higher number of completed projects may lead to universal connectivity and treatment, recall that during the period of study, on average, a district completed one project. Moreover, water treatment —i.e. administering chlorine and other minerals— decreases by 6 ppts (a decrease of 7 %), perhaps because better sludge management is a substitute for supplying safer water, though not statistically significant.

7 Conclusions

Large public infrastructure can be a driver of development, setting LMICs on track to achieve sustainable development goals (SGDs) by 2030. However, the implementation of large public infrastructure can be highly disruptive, resulting in negative unintended consequences. In this paper, I examine the logic of this trade-off by asking the following question. What are the consequences of unfinished infrastructure projects? To answer this question, I focus on the diffusion of sewerage infrastructure across district municipalities between 2005 and 2015 in Peru. The aim of this public intervention was to improve early-life mortality, as was the case in advanced economies during the previous centuries.

There is a large prevalence of unfinished projects across years due to mid-construction abandonment and delays. The majority of projects are "white elephants" (i.e. expensive infrastructure projects that are useless or troublesome) at some point, and the rest are projects at the verge of becoming one (i.e. experiencing delays). By the end of this study, 40 % of the projects were still abandoned, with an average 40 % of the contractual sum disbursed. If these projects are never completed, then a back-of-the-envelope calculation suggests that this would generate a waste equal to 5 % of the public expenditure on education or 4 % of the expenditure on health in 2015 in Peru (World Bank, 2020). These figures reflect the high social opportunity cost of the non-completion of public infrastructure.

---

[8] Again, this study does not have the statistical power to estimate significant effects from the completion of projects.

In this paper, I document that unfinished infrastructure projects could not only be a wasteful use of public resources, but could also generate high social costs (i.e. kill children). Infant and under-five mortality increased with every additional unfinished sewerage project, as opposed to not launching a project. The estimated effect is equivalent to $\sim$6–7 % over the initial averages. Considering that, on average, districts in Peru started four projects, the estimated increase in under-five mortality is equivalent to almost half the mortality rate in 2005 (3.08 deaths per 1,000 children). Because mortality decreased over the period of study, these results can be interpreted as mortality decreasing at a lower rate because of the unfinished infrastructure than it would have otherwise.

I find that water cuts forced the population to rely on unsafe sources of water and jeopardised sanitation practices. Furthermore, the construction works exposed the population to hazards, generating pools of infection from open ditches and increasing accidents among older children. The estimated effect on infant mortality is mostly driven by water-borne diseases, while the effect on under-five mortality can be separated into water-borne diseases (0.20 deaths per 1,000 children) and accidents (0.1 deaths per 1,000 children).

I also show that an additional completed project decreases early-life mortality, as opposed to not starting a project. The estimated negative effects of completed projects are comparable to those of Alsan and Goldin (2019) in the United States during the late 19th century ($\sim$30–40 % from the initial averages). Completing one project did not ensure universal connectivity or sludge treatment, and it crowded out water treatment, preventing the social benefits of sewerage systems from fully manifesting.

By no means is the policy implication of the results that governments should not provide public infrastructure, but its delivery should be complemented with other policies that can mitigate the negative effects. Stricter health and safety measures, improvements in the quality of primary health care and the provision of alternative safe sources of water and sanitation can prevent child deaths during the construction phase.

There is a need to understand better if the social costs of infrastructure development are a result of the monopolistic nature of the institutional arrangement. Galiani et al. (2005), for example, find large gains in connectivity and performance linked to the privatisation of sewerage services in Argentina, which decreased child mortality. The estimated negative effect is of a similar magnitude to the estimated positive effect of an unfinished sewerage project in this paper. Post-construction privatisation could be as good as offsetting the negative effects of the implementation phase of sewerage systems. Nonetheless, Granados and Sánchez (2014) find that municipalities that privatised sewerage services exhibited a slower reduction of child mortality rates and lower increases in coverage.

Regardless of ownership, however, any institutional arrangement will have to deal with the lumpy nature of finance and construction of infrastructure. A reform of the contractual system can help finish projects that are started, such as leaving a high lump sum of the contractual payment for when projects are finalised and including a penalty for not completing infrastructure. The literature has pointed to other policy alternatives to attain project completion and universal connectivity, but there is room to explore further. Rasul and Rogger (2018) suggest that managerial practices of local bureaucrats, such as incentive schemes, increase the probability of completing infrastructure projects. Williams (2017) suggests the inclusion of inter-governmental rules for completing a project before starting a new one, as a way to deal with unstable local political dynamics that deter project completion. Ashraf et al. (2016) suggests finding a “sweet spot” between fines and subsidies to promote connectivity to public sewers.

Another avenue of future research is to identify whether sewerage is unique in triggering early-life mortality or if such an adverse effect can also be seen with other forms of infrastructure. Equally, it is vital to quantify other negative consequences of unfinished public infrastructure projects on well-being and economic outcomes. In short, we must gain a better understanding of how dangerous “white elephants” can be.

References

Alsan, Marcella and Claudia Goldin (2019) "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880 to 1920". *Journal of Political Economy*, 127:586–638.

Aschauer, David Alan (1989) "Is Public Expenditure Productive?". *Journal of Monetary Economics* 23:177–200.

Ashraf, Nava, Edward L. Glaeser, and Giacomo A. M. Ponzetto (2016) "Infrastructure, Incentives, and Institutions". *American Economic Review* 106(5):77–82.

Augsburg, Britta and Paul A. Rodríguez-Lesmes (2018) "Sanitation and Child Health in India". *World Development* 107:22–39.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti (2009) "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment". *American Economic Review* 99(4):1278–1308.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2020). "On the Road: Access to Transportation Infrastructure and Economic Growth in China". *Journal of Development Economics* 145:102442.

Besley, Timothy and Maitreesh Ghatak (2006) "Public Goods and Economic Development" in Abhijit V. Banerjee, Roland Bénabou, and Dilip Mookherjee, eds, *Understanding Poverty*. Oxford: Oxford Scholarship Online.

Bhalotra, Sonia R., Alberto Diaz-Cayeros, Grant Miller, Alfonso Miranda, and Atheendar S. Venkataramani (2018) "Urban Water Disinfection and Mortality Decline in Developing Countries". NBER Working Paper No. 23239.

Bound, John, David A. Jaeger, and Regina M. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak". *Journal of the American Statistical Association* 90(430):443–450.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró Miquel (2015) "The Value of Democracy: Evidence from Road Building in Kenya". *American Economic Review* 105(6):1817–1851.

Cesur, Resul, Erdal Tekin, and Aydogan Ulker (2017). "Air Pollution and Infant Mortality: Evidence from the Expansion of Natural Gas Infrastructure". *Economic Journal* 127(600):330–362.

Cutler, David and Grant Miller (2005) "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States". *Demography* 42:1–22.

Dinkelman, Taryn (2011) "The Effects of Rural Electrification on Employment: New Evidence from South Africa". *American Economic Review* 101(7):3078–3108.

Donaldson, Dave (2018) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure". *American Economic Review* 108(5):899–934.

Duflo, Esther, Michael Greenstone, Raymond Guiteras, and Thomas Clasen (2015) "Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation". NBER Working Paper No. 21521.

Duflo, Esther and Rohini Pande (2007) "Dams". *Quarterly Journal of Economics* 122:601–646.

Fay, Marianne, Luis Alberto Andres, Charles Fox, Ulf Narloch, Stephane Straub, and Michael Slawson (2017) "Rethinking Infrastructure in Latin America and the Caribbean: Spending Better to Achieve More". Technical report, World Bank, Washington DC.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky (2005) "Water for Life: The Impact of the Privatization of Water Services on Child Mortality". *Journal of Political Economy* 113: 83–120.

Geruso, Michael and Dean Spears (2018) "Neighborhood Sanitation and Infant Mortality". *American Economic Journal: Applied Economics* 10:125–162.

Granados, Claudia and Fabio Sánchez (2014) "Water Reforms, Decentralization and Child Mortality in Colombia, 1990–2005". *World Development* 53:68–79.

Gupta, Aashish and Dean Spears (2017) "Health Externalities of India's Expansion of Coal Plants: Evidence from a National Panel of 40,000 Households". *Journal of Environmental Economics and Management* 86:262–276.

Hammer, Mark (1986) *Water and Wastewater Technology* (2nd edn). New York: Wiley.

Hathi, Payal, Sabrina Haque, Lovey Pant, Diane Coffey, and Dean Spears (2017) "Place and Child Health: The Interaction of Population Density and Sanitation in Developing Countries". *Demography* 54:337–360.

Isham, Jonathan and Daniel Kaufmann (1999). "The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects". *Quarterly Journal of Economics* 114:149–184.

Kesztenbaum, Lionel and Jean-Laurent Rosenthal (2017) "Sewers' Diffusion and the Decline of Mortality: The Case of Paris, 1880–1914". *Journal of Urban Economics* 98: 174–186.

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

Malpartida, Tabuchi, Jorge (2018) Mosquitoes due to abandoned sewerage project in Asia, Lima. *El Comercio*, 17 February, https://elcomercio.pe/lima/sucesos/criadero-zancudos-asia-obra-paralizada-noticia-497970-noticia/.

Marcus, Michelle (2017) "On the Road to Recovery: Gasoline Content Regulations and Child Health". *Journal of Health Economics* 54:98–123.

Mettetal, Elizabeth (2019) "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) and Cooperacion Suiza para el Desarrollo (COSUDE), Lima, Peru.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot (2001) *Demography: Measuring and Modeling Population Processes*. Oxford: Wiley-Blackwell.

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

Romero Rojas, Jairo Alberto (2000) *Treatment of Waste Water: Theory and Principles of Design*. Bogota: Editorial Escuela Colombiana de Ingenieria.

RPP Noticias (2018) In Lambayeque, 55 abandoned sewerage projects, 11 January, https://rpp.pe/peru/lambayeque/en-lambayeque-55-obras-de-saneamiento-estan-paralizadas-noticia-1099167.

Rud, Juan Pablo (2012) "Electricity Provision and Industrial Development: Evidence from India". *Journal of Development Economics* 97:352–367.

Sanderson, Eleanor and Frank Windmeijer (2016) "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables". *Journal of Econometrics* 190:212–221.

Serquen, Walter (2018) Children drown in an open ditch from sewerage project. *Correo*, 09 February, https://diariocorreo.pe/edicion/lambayeque/lambayeque-dos-hermanitos-mueren-ahogados-al-caer-una-zanja-de-una-obra-801963.

Stock, James H. and Motohiro Yogo (2002) Testing for Weak Instruments in Linear IV Regression. NBER Technical Working Paper No. 284.

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

Vega, Ysela (2015) "Treatment plant EPSEL is a source of infection". *La Republica*, 30 June, https://larepublica.pe/sociedad/199549-planta-de-tratamiento-de-epsel-es-un-foco-contaminante.

Vega Luna, Eduardo (2015) "Why are there so many sanitation projects abandoned? This affects your rights", *Defensoria del Pueblo*, 09 June, https://www.defensoria.gob.pe/blog/obras-de-agua-paralizadas/.

Von Hesse, Milton (2016) Guidelines for a National Sanitation Policy. Technical report, Peruvian Ministry of Housing, Construction and Sanitation, Lima.

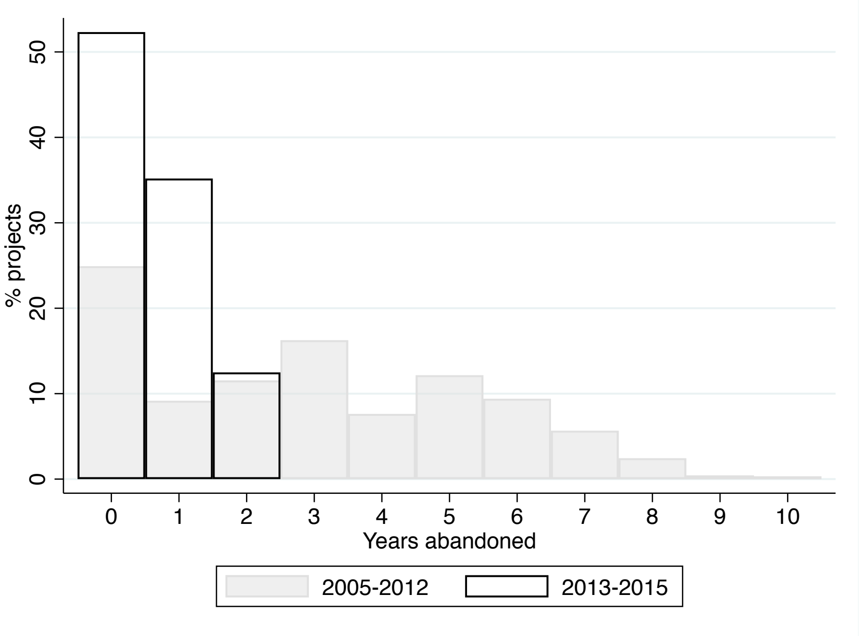
Watson, Tara (2006) "Public Health Investments and the Infant Mortality Gap: Evidence from Federal Sanitation Interventions on U.S. Indian Reservations". *Journal of Public Economics* 90:1537–1560.

Williams, Martin J. (2017) "The Political Economy of Unfinished Development Projects: Corruption, Clientelism, or Collective Choice?". *American Political Science Review* 111:705–723.

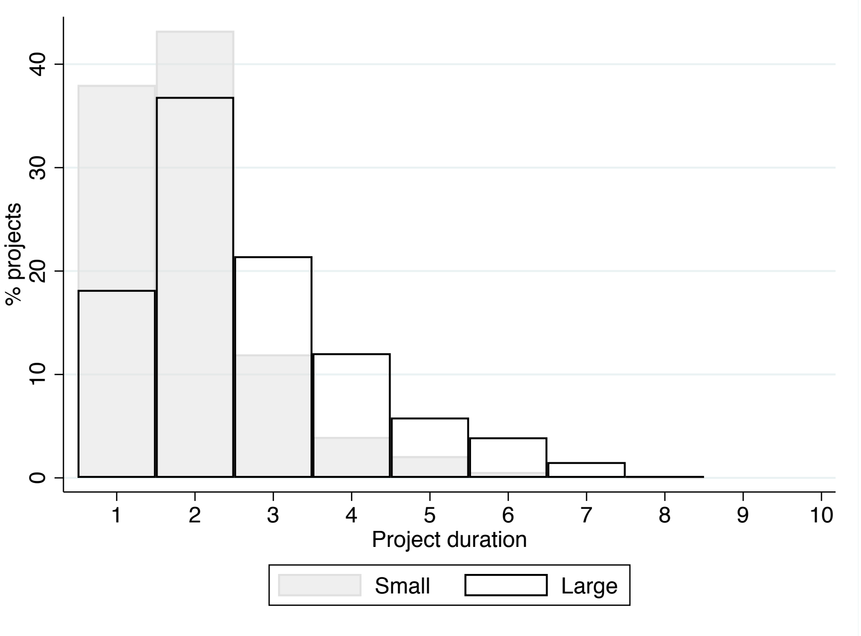
World Bank (2017) "World Bank Lending (Fiscal 2017)". Technical report. Washington DC: World Bank.

World Bank (2020) World Development Indicators. Washington DC: World Bank.

Figure 1: Mid-construction abandonment and project duration, between 2005 and 2015



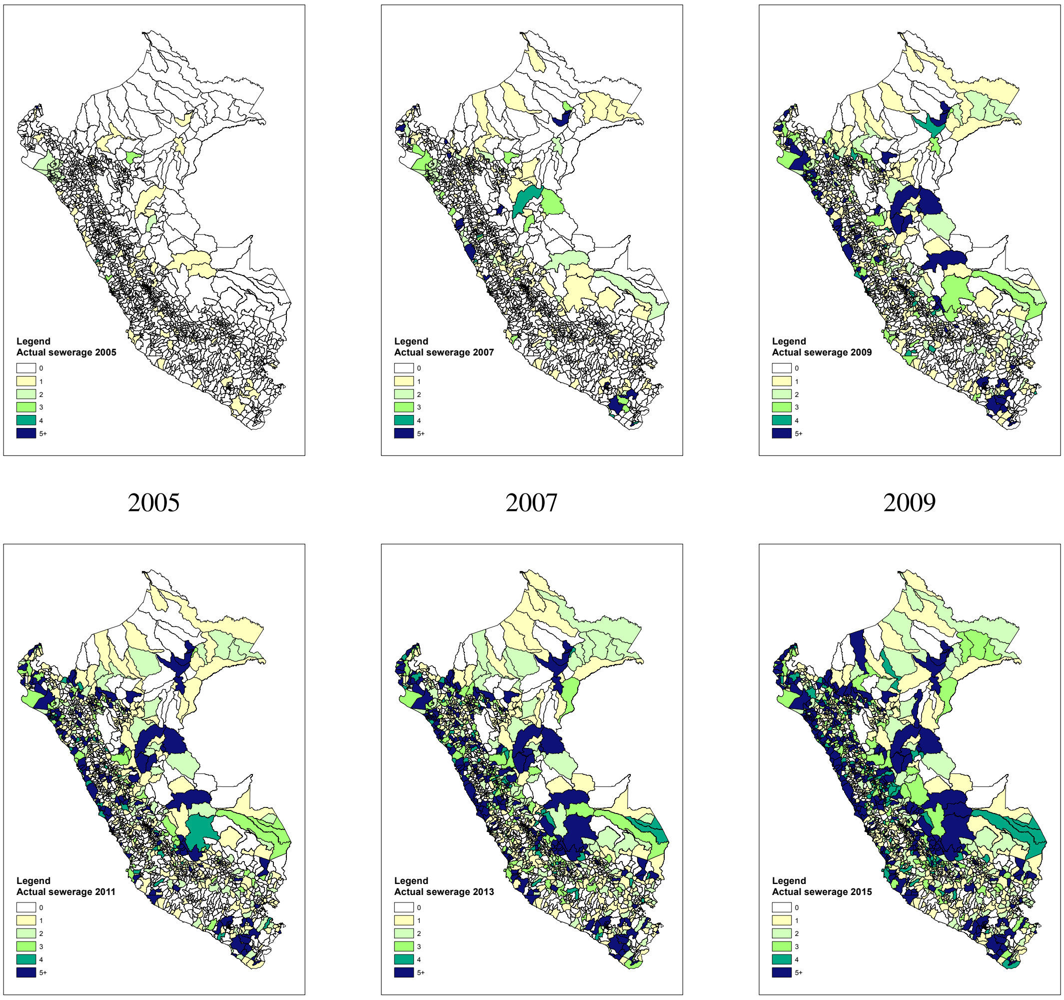
Panel A. Abandonment



Panel B. Duration

Notes: Abandonment is computed as the number of years that no additional funds are disbursed even when a project is still underway. Project duration is computed as the number of years it takes for a project to be completed (if it ever accrued more than 90 % of the budgeted investment). Projects are considered small if they are planned to affect below the median of the distribution of beneficiaries, and large projects otherwise. Sample is restricted to projects that were ever started between 2005 and 2015 in Panel A and also to those completed between 2005 and 2015 in Panel B.

Figure 2: Actual projects across districts in Peru, 2005–2015



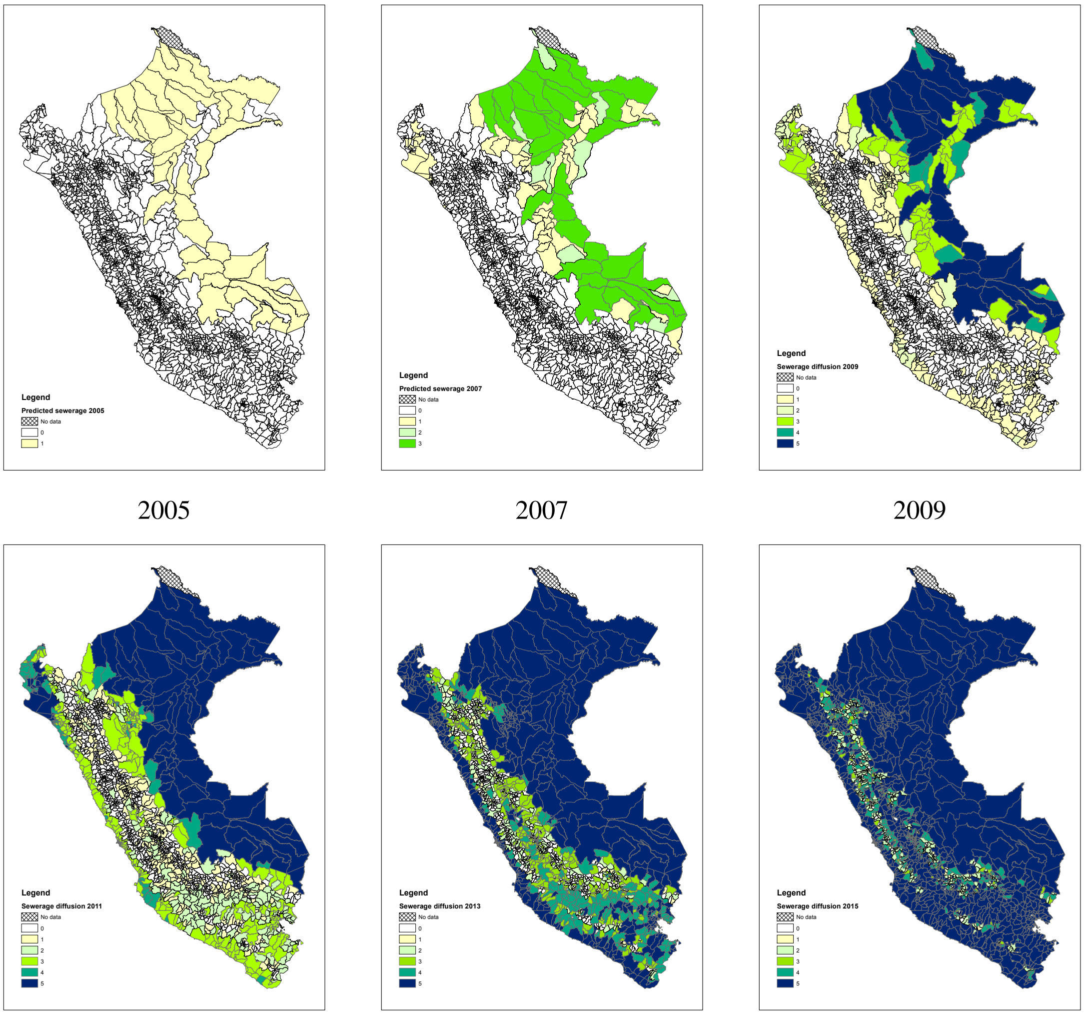
2005 2007 2009

2011 2013 2015

Notes: These maps show the district boundaries of Peru and the distribution across districts of the actual number of sewerage projects started 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.

Source: Author's calculations using data on the number of sewerage projects started between 2005 and 2015 from the SNIP and the SIAF.

Figure 3: Predicted projects across districts in Peru, 2005–2015



2005 2007 2009

2011 2013 2015

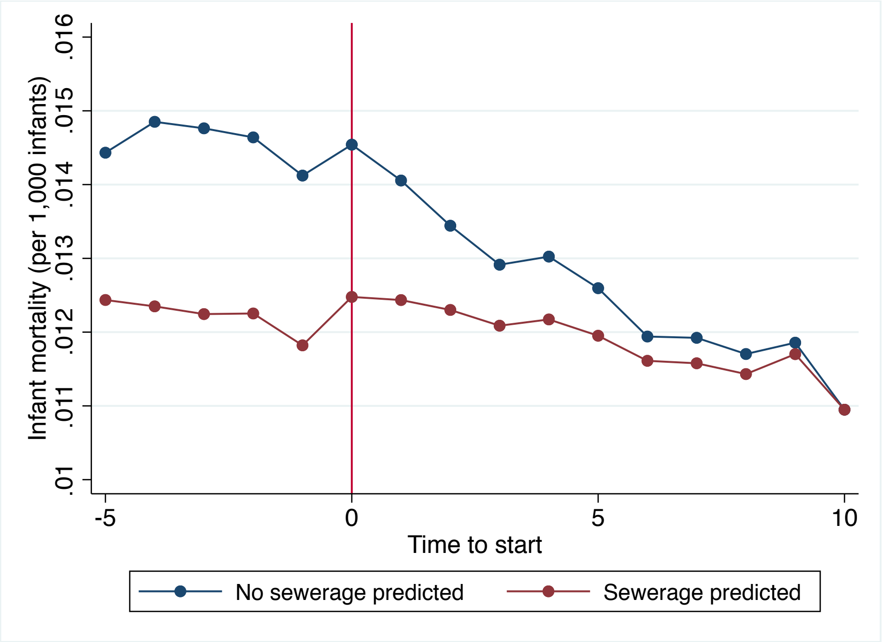
Notes: These maps show the district boundaries of Peru and the distribution across districts of the predicted number of sewerage projects to be started and completed between 2005 and 2015. Light-shaded districts are those in which no or few sewerage projects were allocated and dark-shaded districts are those in which several sewerage projects were allocated.
Source: Author's calculations using data on the number of sewerage projects started between 2005 and 2015 from the SNIP and the SIAF.

Table 1: Effect of unfinished projects on early-life mortality

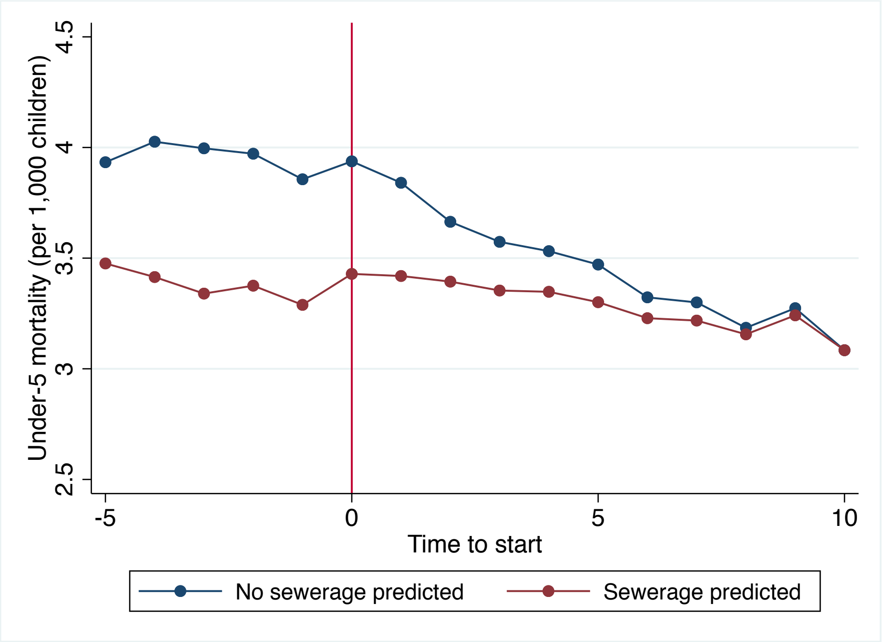
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1st stage	Reduced form		OLS		2SLS	
Dependent variable	Unfinished projects	IMR	U5MR	IMR	U5MR	IMR	U5MR
Predicted projects	0.356	0.000	0.069				
	(0.069)	(0.000)	(0.065)				
	[0.000]	[0.176]	[0.284]				
Unfinished projects				0.000	0.030	0.001	0.299
				(0.000)	(0.007)	(0.000)	(0.105)
				[0.000]	[0.000]	[0.008]	[0.004]
Mean (initial)				0.018	4.816	0.018	4.816
F-stat (Sanderson-Windmeijer)				.	.	26.996	26.996
Sample		Before start	Before start				
District-year	10,494	5,443	5,443	10,494	10,494	10,494	10,494
Districts	1,408	1,234	1,234	1,408	1,408	1,408	1,408

Notes: This table presents the main results of the effect of the number of unfinished sewerage projects on early-life mortality rates. The dependant variable in column (1) is the number of unfinished projects; IMR in columns (2), (4) and (6) and U5MR in (3), (5) and (7). Columns (2) and (3) restrict the sample to years before the start of sewerage projects. Columns (4) and (5) show OLS estimates and columns (6) and (7) show 2SLS estimates. The table also shows the Sanderson–Windmeijer F-statistic and the average initial (2005) mortality rates. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets.

Figure 4: Before and after starting projects: compliers and always-takers



Panel A. Infant Mortality Rate



Panel B. Under-five Mortality Rate

Notes: The plots show trends for mortality before and after the first project started in districts. The red vertical line denotes the time in which the first sewerage project in a given district was started. The average mortality rate of districts that never started a sewerage project is also placed in time to start equal to zero. The analysis is split into districts predicted to receive sewerage projects and those not predicted to.

Table 2: Sensitivity analysis

Specification	(1) IMR	(2) U5MR	(3) F-statistic (SW)
Additional controls			
1. Pop-density (t-1)	0.001	0.212	21.430
	(0.001)	(0.116)	
	[0.109]	[0.069]	
2. Municipal characteristics	0.001	0.322	24.438
	(0.001)	(0.115)	
	[0.010]	[0.005]	
3. Amazon location dummy	0.001	0.299	26.996
	(0.000)	(0.105)	
	[0.008]	[0.004]	
Changing sample			
4. Intervened districts	0.001	0.267	27.474
	(0.000)	(0.094)	
	[0.008]	[0.004]	
5. Lima excluded	0.001	0.250	24.238
	(0.000)	(0.106)	
	[0.030]	[0.018]	
Transformation			
6. Projects top-coded	0.001	0.198	18.135
	(0.000)	(0.087)	
	[0.036]	[0.023]	
7. Projects density (10000 people per sq km)	0.002	0.427	8.807
	(0.001)	(0.207)	
	[0.062]	[0.039]	

Notes: Each row represents a different sensitivity test on the specifications reported in columns (3) and (4) in Table 1. Columns (1) and (2) in this table report the coefficient and standard error on unfinished projects where the dependent variable is the infant mortality rate (1) and under-five mortality rate (2). Column (3) reports the associated first-stage F-statistic (Sanderson–Windmeijer). The different specifications in each row are reported in the left-hand column. Coefficients correspond to 2SLS estimations. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets. Specifications are as follows: 1, controls for lagged population density; 2, controls for municipal characteristics, including indicators capturing whether district municipality has access to the Internet, needs technical assistance to formulate investment projects, and manages at least one health centre, and municipal income (ln), where missing values are replaced by the district's average value; 3, controls for a dummy capturing whether the district is located in the Amazon region; 4, restricts the sample of analysis to those districts that ever had an intervention (at least one sewerage project ever started); 5, excludes the region of the capital of Peru, Lima, from the sample of analysis; 6, transforms endogenous variable (unfinished projects) to a version top-coded at the top 10 percentile; 7, transforms endogenous variable (unfinished projects) and instrumental variable (predicted projects) to a version interacted with population density (10,000 people per km^2).

Table 3: Validity of IV: projects in other sectors

Specification	(1) IMR	(2) U5MR	(3) F-statistic (SW)
Expenditure controls			
1. Transportation	0.002	0.352	14.436
	(0.001)	(0.164)	
	[0.036]	[0.032]	
2. Energy	0.001	0.223	12.485
	(0.001)	(0.159)	
	[0.105]	[0.161]	
3. Health	0.002	0.346	14.227
	(0.001)	(0.165)	
	[0.040]	[0.036]	
Alternative endogenous variable			
4. IV for Transportation	0.009	2.056	0.751
	(0.011)	(2.579)	
	[0.428]	[0.425]	
5. IV for Energy	0.002	0.376	3.673
	(0.002)	(0.322)	
	[0.199]	[0.243]	
6. IV for Health	-0.002	-0.497	9.337
	(0.001)	(0.266)	
	[0.066]	[0.061]	

Notes: Each row represents a different sensitivity test on the specifications reported in columns (3) and (4) in Table 1. Columns (1) and (2) in this table report the coefficient and standard error on unfinished projects where the dependent variable is the infant mortality rate (1) and under-five mortality rate (2). Column (3) reports the associated first-stage F-statistic (Sanderson–Windmeijer). The different specifications in each row are reported in the left-hand column. Coefficients correspond to 2SLS estimations. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets. Specifications are as follows: 1, controls for district's expenditure in transportation projects (log); 2, controls for district's expenditure in energy projects (log); 3, controls for district's expenditure in health projects (log); 4, alternative endogenous variable – district's expenditure in transportation projects (log); 5, alternative endogenous variable – district's expenditure in energy projects (log); 6, alternative endogenous variable – district's expenditure in health projects (log).

Table 4: Validity of IV: geographic controls

Specification	(1) IMR	(2) U5MR	(3) F-statistic (SW)
1. Gradient x year dummies	0.001	0.297	21.147
	(0.001)	(0.125)	
	[0.026]	[0.018]	
2. Elevation x year dummies	0.001	0.230	13.412
	(0.001)	(0.164)	
	[0.171]	[0.160]	
3. Area x year dummies	0.001	0.293	26.944
	(0.000)	(0.104)	
	[0.009]	[0.005]	
4. Amazon x year dummies	0.001	0.295	26.892
	(0.000)	(0.106)	
	[0.009]	[0.005]	
5. Gradient x annual budget	0.001	0.279	15.698
	(0.001)	(0.141)	
	[0.059]	[0.048]	
6. Elevation x annual budget	0.001	0.130	7.100
	(0.001)	(0.219)	
	[0.484]	[0.553]	
7. Area x annual budget	0.001	0.284	26.699
	(0.000)	(0.104)	
	[0.011]	[0.006]	
8. Amazon x annual budget	0.001	0.288	26.506
	(0.000)	(0.107)	
	[0.013]	[0.007]	
9. Pop density x year dummies	0.001	0.231	29.851
	(0.000)	(0.092)	
	[0.019]	[0.012]	
10. Pop density x annual budget	0.001	0.212	30.465
	(0.000)	(0.089)	
	[0.027]	[0.017]	

Notes: Each row represents a different sensitivity test on the specifications reported in columns (3) and (4) in Table 1. Columns (1) and (2) in this table report the coefficient and standard error on unfinished projects where the dependent variable is the infant mortality rate (1) and under-five mortality rate (2). Column (3) reports the associated first-stage F-statistic (Sanderson–Windmeijer). The different specifications in each row are reported in the left-hand column. Coefficients correspond to 2SLS estimations. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets. Specifications are as follows: 1 and 5, gradient is the percentage of area falling in the lowest gradient category (0–0.8 %); 2 and 6, elevation is the percentage of area falling in the lowest elevation category (below 250 mamsl); 3 and 7, area is km^2; 4 and 8, Amazon location dummy is one if the region is in the Amazon; 9 and 10, population density corresponding to the initial year (2005).

Table 5: Effects on water and sanitation behaviour

	(1)	(2)	(3)	(4)	(5)
		Water		Sanitation	
	Connectivity	Treated	% Unsafe	% Latrine	% OD
Unfinished projects	0.031	-0.067	0.030	-0.039	0.049
	(0.024)	(0.037)	(0.012)	(0.012)	(0.013)
	[0.200]	[0.068]	[0.015]	[0.001]	[0.000]
F-stat (SW)	26.137	7.981	19.464	19.464	19.464
Mean (initial)	0.973	0.853	0.459	0.342	0.407
District-year	3326	6355	2630	2630	2630
Districts	1054	1277	1014	1014	1014

The dependent variables are the following: the district has high connectivity to piped water (column 1); the municipality treats water (column 2); the percentage of households that rely on unsafe sources of water (column 3); the percentage of households that use a latrine (column 4); the percentage of households that practise open defecation (OD; column 5). Coefficients correspond to a 2SLS estimation. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets.

Table 6: Effect of unfinished projects on mortality by cause of death

	(1)	(2)	(3)	(4)
	Coeff.		Initial mean	
Specification	IMR	U5MR	IMR	U5MR
Water borne	0.001	0.221	0.009	2.265
	(0.000)	(0.078)		
	[0.012]	[0.005]		
Accidents	0.000	0.091	0.004	1.248
	(0.000)	(0.051)		
	[0.182]	[0.077]		
Respiratory	-0.000	-0.003	0.003	0.736
	(0.000)	(0.044)		
	[0.564]	[0.941]		
Malformation	0.000	0.030	0.002	0.388
	(0.000)	(0.034)		
	[0.222]	[0.376]		
Other	0.000	-0.012	0.005	1.303
	(0.000)	(0.056)		
	[0.580]	[0.827]		

Notes: Each row represents a different cause of death. Columns (1) and (2) report the coefficient and standard error on unfinished projects where the dependent variable is the IMR and U5MR. Columns (3) and (4) report the mean mortality of the initial year (2005). Coefficients correspond to 2SLS estimations. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets.

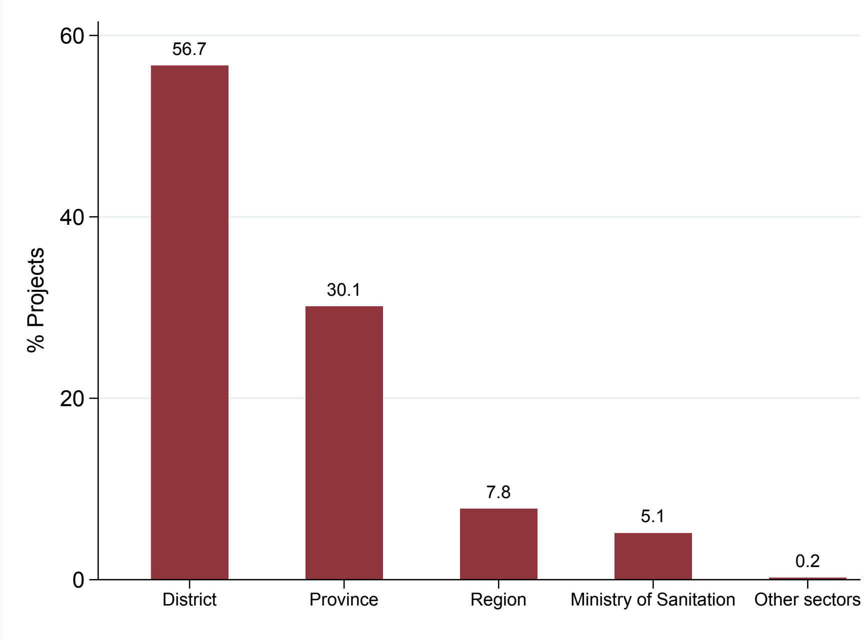
Table 7: Effect of unfinished and completed projects on early-life mortality

	(1)	(2)	(3)	(4)	(5)	(6)
	1st stage		2SLS			
Dependent variable	Unfinished	Completed	IMR	U5MR	IMR	U5MR
Predicted projects	0.259	0.113				
	(0.067)	(0.038)				
	[0.000]	[0.003]				
Geography * Partisan alignment	1.027	0.887				
	(0.705)	(0.452)				
	[0.145]	[0.050]				
Unfinished project			0.004	0.862		
			(0.004)	(0.830)		
			[0.288]	[0.299]		
Completed project			-0.006	-1.214	-0.001	-0.244
			(0.008)	(1.700)	(0.003)	(0.558)
			[0.455]	[0.475]	[0.602]	[0.662]
Fstat (Kleibergen-Paap)			0.603	0.603	3.621	3.621
District-year	8517	8517	8517	8517	8517	8517
Districts	1212	1212	1212	1212	1212	1212

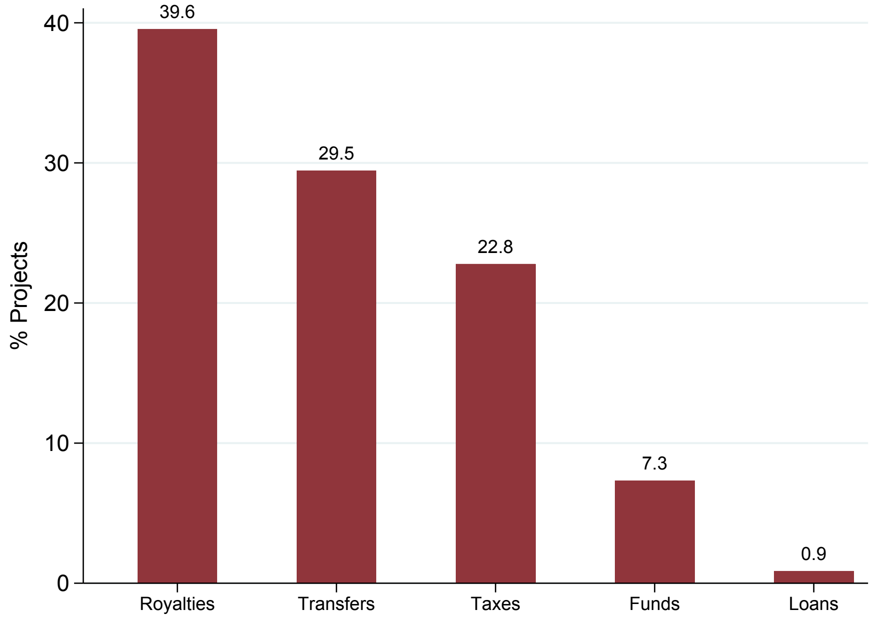
Notes: This table presents the first-stage results and the estimated effect of the number of unfinished sewerage projects and cumulative completed projects on early-life mortality rates. The dependant variable in columns (1) and (2) is the number of unfinished projects and of completed projects, respectively; in columns (3) and (5) the IMR and in columns (4) and (6) the U5MR. All regressions control for partisan alignment and include district and year fixed effects. The lower number of observations is due to missings in the partisan alignment variable. The table also shows the Kleibergen–Paap rk Wald F-statistic for the regression with multiple endogenous variables and the Sanderson–Windmeijer F-statistic for the regression with a single endogenous variable. As reference: the 10% Maximal IV size is 7.03; the 15% Maximal IV size is 4.58; the 20% Maximal IV size is 3.95; and the 25% Maximal IV size is 3.63. Clustered standard errors by district are given in parentheses and p-values in brackets.

Appendix

Figure A1: Agency formulating and funding sources, sewerage projects 2005–2015



Panel A. Agency

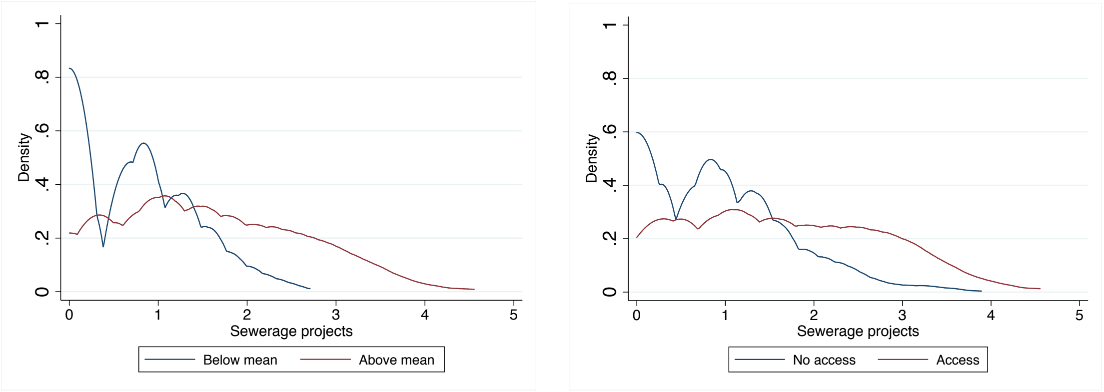


Panel B. Funding

Notes: These figures shows the percentage of sewerage projects formulated by each government agency and funded by different sources. The percentage is calculated from the pool of projects declared viable and started between 2005 and 2015.

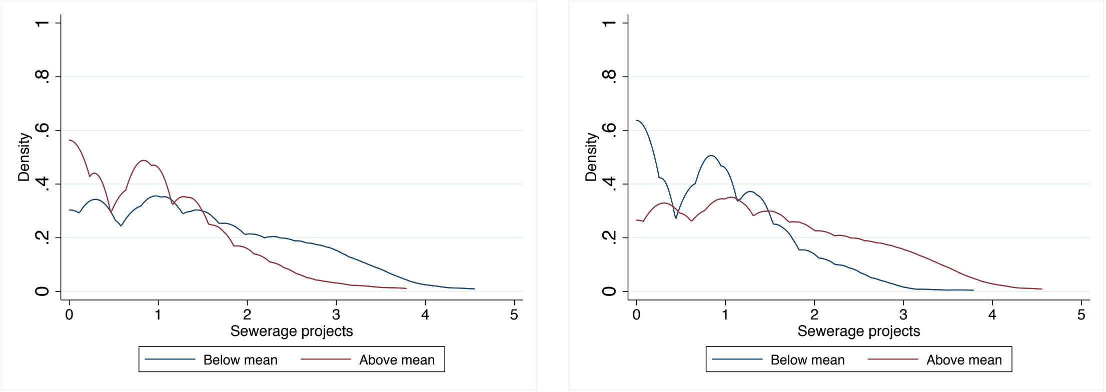
Source: Author's calculations using data from the SNIP and SIAF.

Figure A2: Project allocation by initial municipal characteristics



Panel A. Municipal revenue

Panel B. Internet access

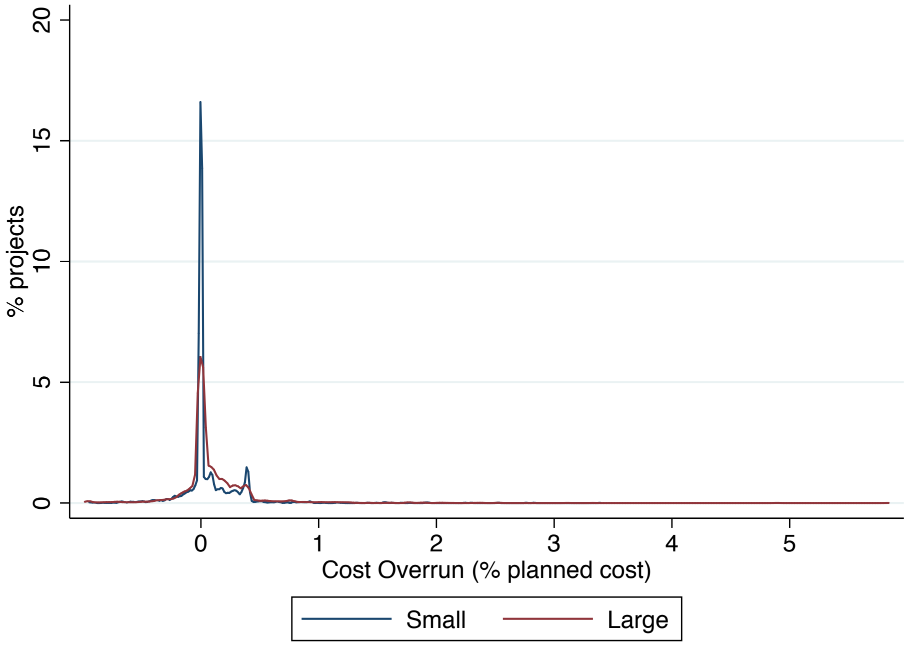


Panel C. Unmet basic needs

Panel D. Sewerage connectivity

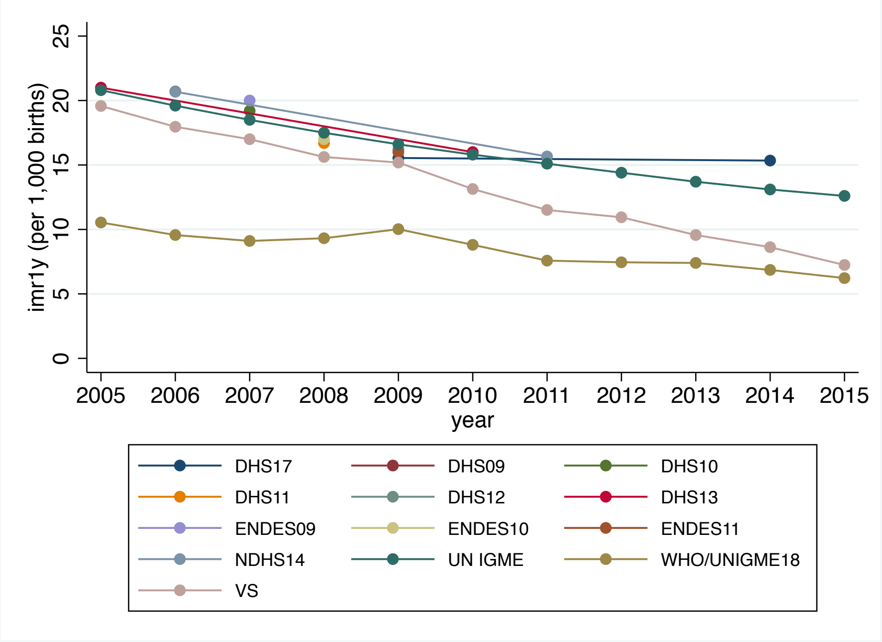
Notes: These figures show the distribution of started sewerage projects by initial municipal characteristics. The blue distribution corresponds to municipalities with characteristics below the median and the red distribution corresponds to those above the median of the distribution by 2005.

Figure A3: Distribution of cost overrun for sewerage projects, 2005–2015

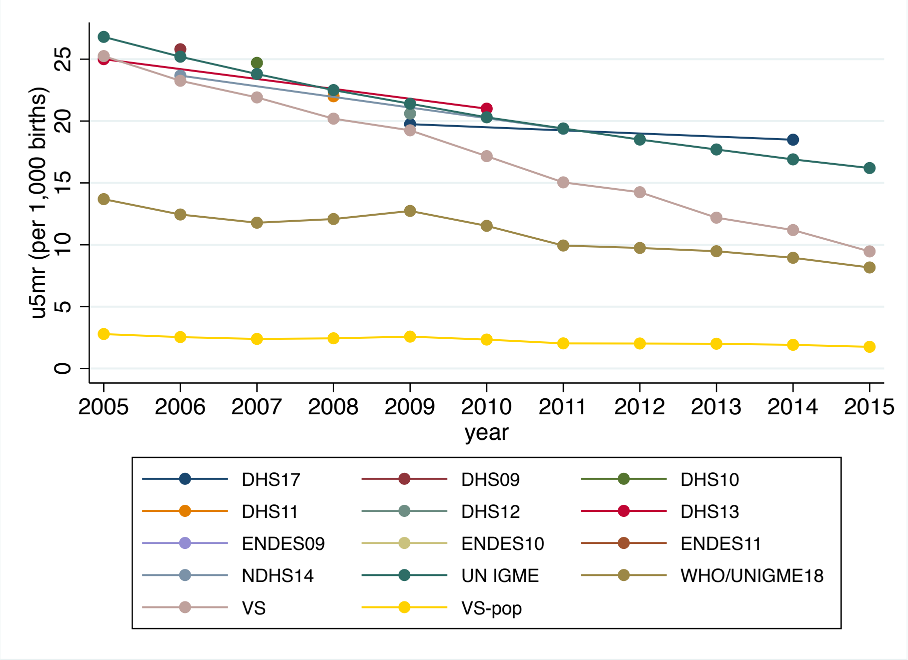


Notes: This figure shows the distribution of cost overrun as a percentage of the planned cost. It is calculated as the difference between actual and planned costs, divided by the planned cost.

Figure A4: MR from vital statistics compared with other data sources



Panel A. Infant Mortality Rate



Panel B. Under-five Mortality Rate

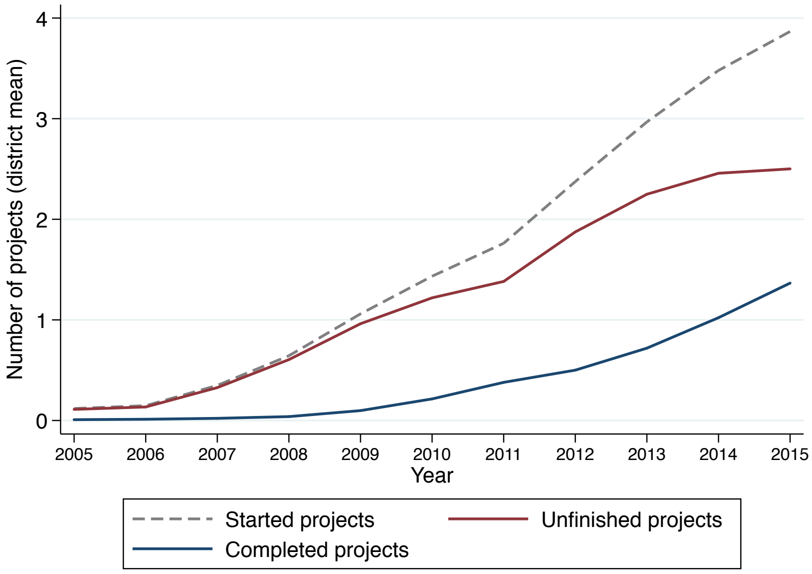
Notes: Alternative data obtained from the Health and Demographic Surveys (DHS), the National Survey of Health and Demography (ENDES) and Inter-Agency Group for Child Mortality Estimation (UN IGME).

Table A1: Summary statistics and data sources

	(1)	(2)	(3)	(4)	(5)
	Beginning period		End period		Source
	Sum		Sum		
1. Outcomes					
Deaths under 1y	6,404		3,820		Vital records
Deaths under 5y	8,256		4,987		
Population under 5y	2,672,357		2,481,908		INEI Pop forecast
Infant mortality (per 1,000 infants)	11.98		7.70		
Under-five mortality (per 1,000 children)	3.08		2.01		
2. Sewerage diffusion					
Started projects	161		4,873		SNIP and SIAF reports
Completed projects	11		1,754		
	Mean	SD	Mean	SD	
2. District characteristics					
Population density (pop/sq km)	642.91	2837.77	847.34	3188.96	Census and Spatial data
Population	23,403.32	57,020.49	32,947.11	75,973.03	Census
Municipal revenue (millions)	4.84	21.82	15.50	55.47	Municipal Registry
Internet access	0.38	0.48	0.93	0.26	
TA in formulation of investment projects	0.66	0.46	0.58	0.49	
Manages health centers	0.22	0.41	0.32	0.47	
Water connectivity	0.61	0.49	0.86	0.34	
Water treated	0.85	0.36	0.99	0.10	
Sewerage treated	0.23	0.42	0.57	0.50	
Share HH unsafe water	0.46	0.27	0.28	0.23	Census
Share HH sewer	0.25	0.27	0.46	0.29	
Share HH on-site	0.34	0.23	0.40	0.25	
Share HH open defecation	0.41	0.26	0.13	0.13	
Share HH head secondary	0.22	0.15	0.34	0.16	
Share HH electrified	0.56	0.26	0.79	0.16	
Transport expenditure (millions)	1.50	7.62	1.92	7.94	SIAF reports
Energy expenditure (millions)	0.04	0.22	0.19	1.13	
Health expenditure (millions)	0.71	2.53	0.36	1.49	
Major affiliated to the government party	0.11	0.31	0.13	0.33	Electoral data
3. Geography					
Fraction district gradient $\leq$ 0.8%	0.10	0.23			Spatial data
Fraction district gradient {0.8-4.19]%	0.19	0.22			
Fraction district gradient {4.19-13]%	0.34	0.20			
Fraction district gradient above 13%	0.37	0.29			
Fraction district elevation $\leq$ 250 mamls.	0.15	0.33			
Fraction district elevation {250-500] mamls.	0.05	0.14			
Fraction district elevation {500-1000] mamls.	0.06	0.15			
Fraction district elevation above 1000 mamls.	0.74	0.41			
River density (km/sq km)	53.32	124.30			
District area (sq. km)	635.93	1,655.50			

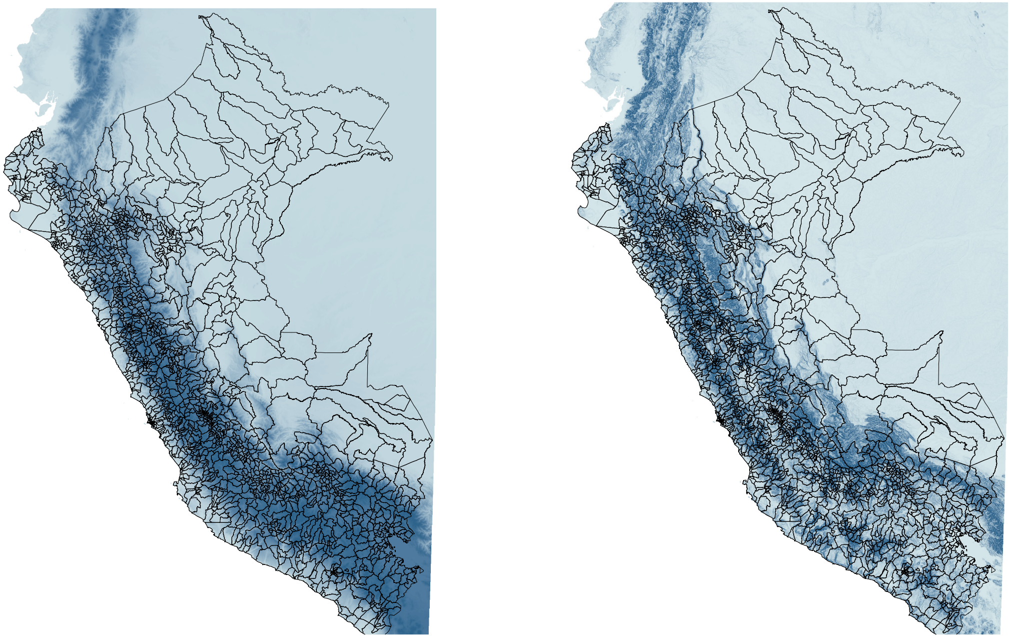
Notes: The beginning period is 2005 and the end period is 2015. Columns (1) and (3) provide the sum for the variables of interest and the mean for the geographical and control variables for 2005 and 2015, respectively. Columns (2) and (4) provide the standard deviation for control variables for 2005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

Figure A5: Number of projects between 2005 and 2015 (district average)



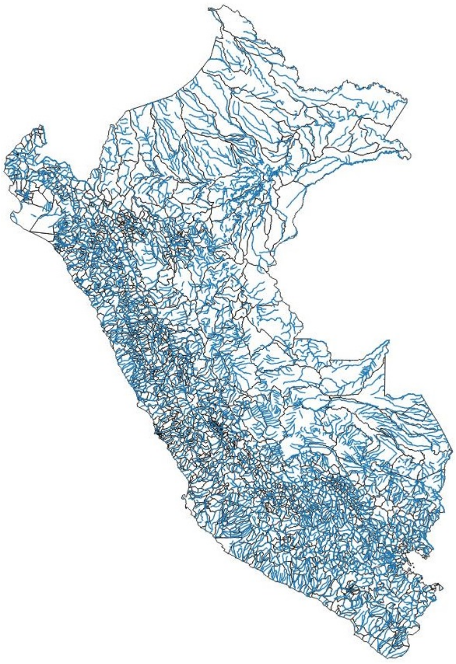
Notes: The grey dashed line shows the average cumulative number of projects started, the red line shows the average number of projects unfinished and the blue line shows the average cumulative number of projects completed.

Figure A6: Geography in Peru



Panel A. Elevation

Panel B. Slope



Panel C. Rivers

Notes: Darker shaded grid cells are at a higher altitude.

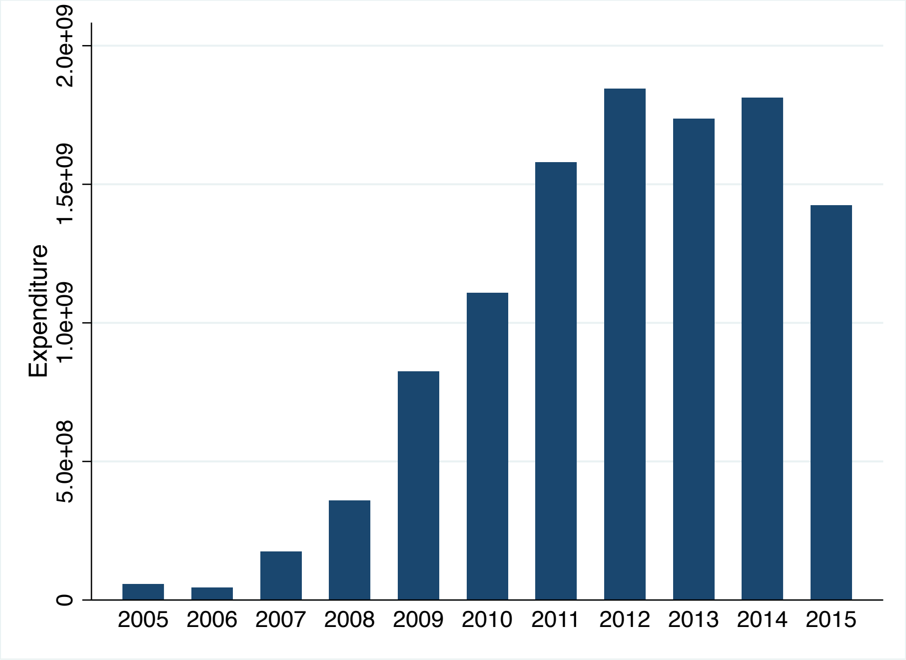
Source: Digital elevation maps provided by the Peruvian Ministry of Environment with information on multiple cells (1×1 km^2).

Table A2: Geographic cost parameters for sewerage projects

	(1)	(2)
Dependent variable	Sewerage projects 2005–2015	
	OLS coeff.	Beta coeff.
Fraction district gradient {0.8-4.19]%	0.833 (2.047) [0.684]	0.022
Fraction district gradient {4.19-13]%	2.315 (1.785) [0.195]	0.064
Fraction district gradient above 13%	0.903 (1.542) [0.558]	0.038
Fraction district elevation {250-500] mamls	-5.015 (1.475) [0.001]	-0.103
Fraction district elevation {500-1000] mamls	-1.425 (1.818) [0.433]	-0.029
Fraction district elevation above 1000 mamls	-6.710 (1.233) [0.000]	-0.369
River density (km/sq km)	0.005 (0.003) [0.090]	0.096
District area (sq. km)	-0.001 (0.000) [0.016]	-0.134
Observations	1832	

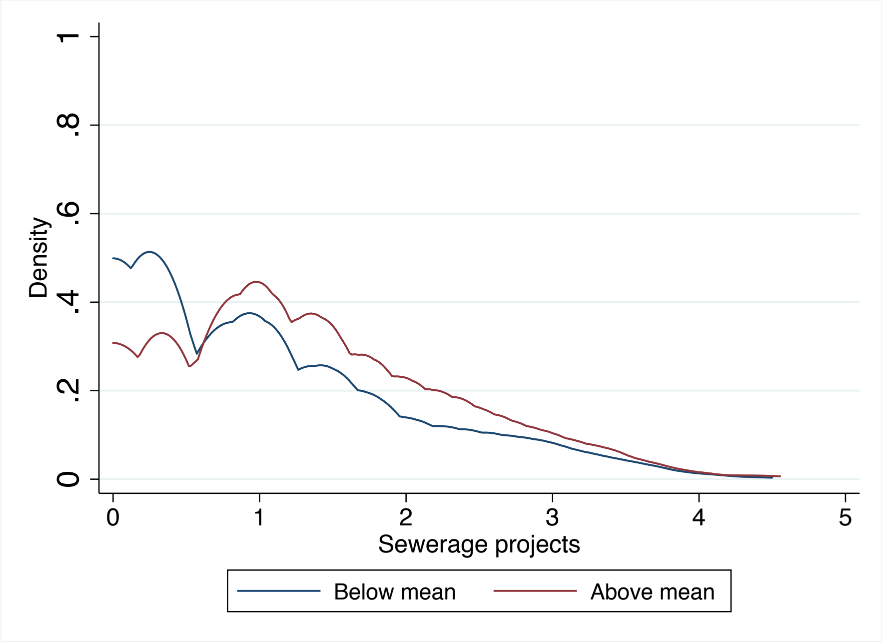
Notes: The dependent variable is the number of sewerage projects started between 2005 and 2015. Column (1) shows the coefficients of an OLS regression and column (2) shows the standardised beta coefficients. The omitted gradient category is the fraction of district area in the flat category (below 0.8 %) and the omitted elevation category is the fraction of district area in the low-altitude category (below 250 mamsl). Robust standard errors are given in parentheses.

Figure A7: Annual Budget Spent in Sewerage Projects

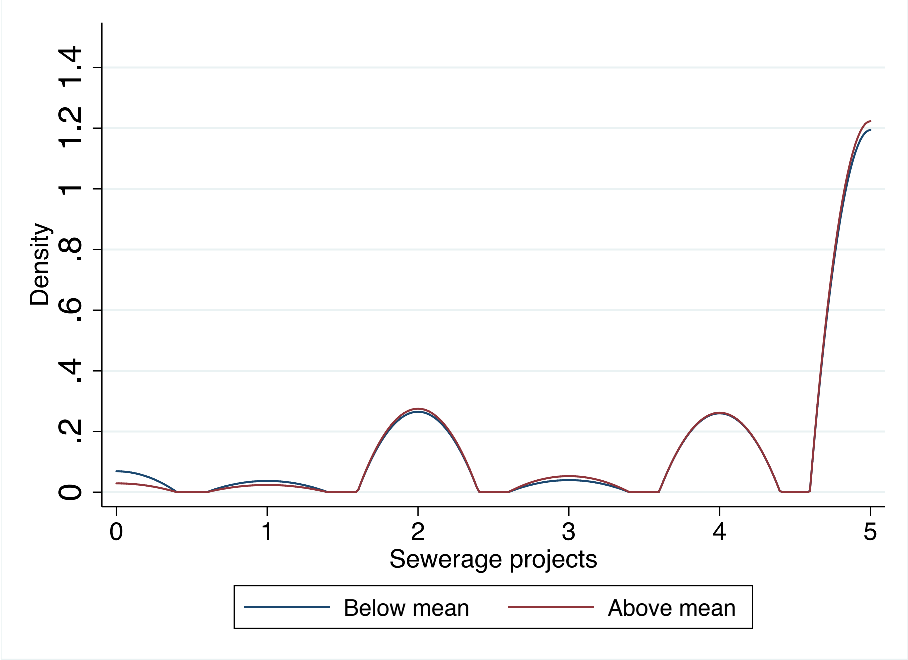


Note: Author's calculation using data from the National System of Public Investment (SNIP for its Spanish acronyms) and the Integrated System of Financial Administration (SIAF for its Spanish acronyms).

Figure A8: Distribution of actual and predicted projects by percentage of rural population



Panel A. Actual



Panel B. Predicted

Notes: These figures show the distribution of actual (Panel A) and predicted (Panel B) sewerage projects by the district's percentage of rural population. The blue distribution corresponds to districts with a 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 A3: Effects on fertility and migration

	(1)	(2)	(3)	(4)
Dependent variable	Births	Pop U5	Educ sec	Electricity
Unfinished projects	0.369	0.018	-0.002	-0.036
	(0.256)	(0.005)	(0.002)	(0.011)
	[0.149]	[0.001]	[0.420]	[0.001]
F-stat (SW)	27.034	26.996	19.464	21.296
Mean (initial)	2.917	6.988	0.219	0.557
District-year	10495	10494	2630	1406
Districts	1408	1408	1014	703

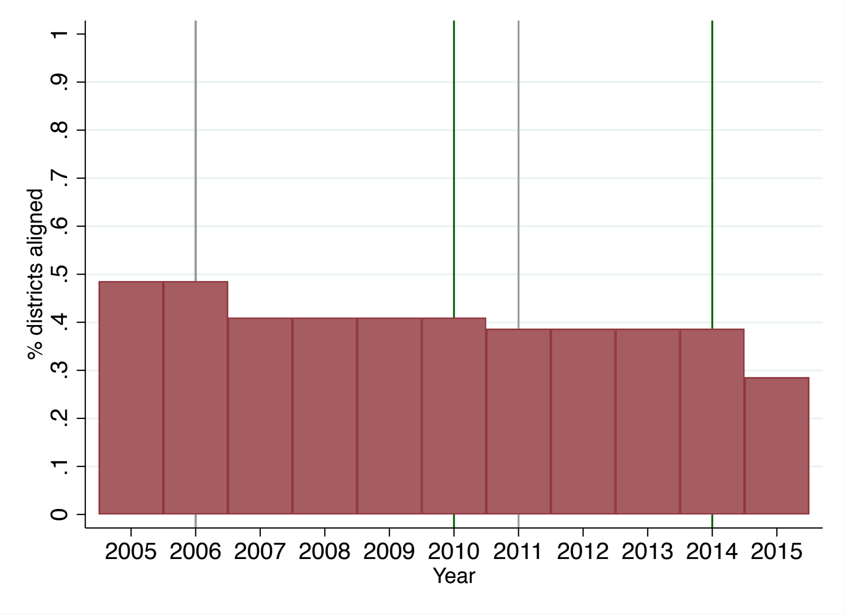
The dependent variable in columns (1)–(4), respectively, is the number of live births (ln), the under-five population (ln), the percentage of household heads with secondary education completed and the percentage of households connected to the electricity network. Coefficients correspond to a 2SLS estimation. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets.

Table A4: Robustness check: socio-economic trends

	(1)	(2)	(3)
Specification	IMR	U5MR	F-stat (SW)
1. Education x year dummies	0.001	0.300	21.509
	(0.001)	(0.126)	
	[0.026]	[0.017]	
2. Electrification x year dummies	0.001	0.328	16.458
	(0.001)	(0.153)	
	[0.037]	[0.032]	

Notes: Each row represents a different sensitivity test on the specifications reported in columns (3) and (4) in Table 1. Columns (2) and (3) in this table report the coefficient and standard error on unfinished projects where the dependent variable is the infant mortality rate (1) and under-five mortality rate (2). Column (3) reports the associated first-stage F-statistic (Sanderson–Windmeijer). The different specifications in each row are reported in column (1). Coefficients correspond to 2SLS estimations. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets. 1. Share of households with the head having completed secondary education. 2. Share of households connected to the electricity grid.

Figure A9: Partisan alignment between district major and central government



Notes: Blue lines for municipal elections and green lines for central elections.

Table A5: Effects on connectivity and treatment

	(1) Sewerage	(2)	(3) Water
	Connectivity	Treatment	Treatment
Completed projects	0.231	0.158	-0.066
	(0.372)	(0.155)	(0.080)
	[0.535]	[0.308]	[0.406]
F-stat (SW)	0.413	3.281	4.173
Mean (initial)			
District-year	3,583	10,395	8,209
Districts	1,211	1,212	1,212

The dependent variables in columns (1)–(3), respectively are the percentage of households connected to sewerage, an indicator for whether the municipality reports that sludge is treated in the district and an indicator for whether water is treated in the district. All regressions include district and year fixed effects. Clustered standard errors by district are given in parentheses and p-values in brackets.

Figure A10: Sewerage project abandoned in Piura with a completion rate below 60 %

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

Figure A11: Sewerage project abandoned in Huanuco

Source: Photograph taken in Huanuco for the technical report of the Defensoria del Pueblo (Vega Luna, 2015) exploring mid-construction abandonment of sewerage projects.