Essays on public service delivery in India

Dissertation

zur

Erlangung des Doktorgrades des

Fachbereichs Wirtschaftswissenschaften der

Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Paula von Haaren

Heidelberg

September 2022

"By assuming that the machine either runs on its own or does not run at all, we avoid having to go looking for where the wheels are getting caught and figuring out what small adjustments it would take to get the machine to run properly."

Abhijit Banerjee. "Inside the Machine." *Making Aid Work* (2007): 123-166.

Acknowledgements

This thesis could not have been written without the help of a great number of people. In particular, I am greatly indebted and thankful to the following:

Stefan Klonner, for his supervision during the last years, in particular for giving me the opportunity to be part of his research group, for teaching me the value of detail, for his dedicated work as a co-author, for always being present and never hesitating to answer my numerous questions (sometimes repeatedly), and for generously allowing me time to gather new strength when I needed it.

Christine Binzel, for taking on the additional responsibility and time to be second referee of this thesis, taking genuine interest in my career and her very helpful feedback and suggestions.

Rahul Mukherjee for making time to be part of my dissertation committee and bringing me into contact with Sowmya Kidambi, who generously shared valuable information with me.

Axel Dreher, for insisting that I write a quantitative bachelor thesis, which sparked my interest in research.

My dear colleagues, Christina, Michael, Christian, Min, Kafeel and Cristina, for welcoming me with open arms, translating and explaining everything, supporting and nudging me when needed, and Sumantra for lending me his hands and hunting down social audit reports.

Sarah Langlotz, Angelika Budjan, Marion Krämer and Jana Kuhnt for many helpful conversations, being such power women and proving that it is possible to do relevant research during your PhD. You are my role models! Thank you Marion for telling me that my time will come.

The numerous workshop and conference participants and faculty members who all contributed feedback and emotional support through many conversations and lunch times and became good friends. In particular, I thank Johannes Matzat, Zain Chaudhry, Cristina Cibin, Matthias Quinckhardt, Marion, Michaela Theil and Ute Seitz for taking the time to give me some very detailed feedback, and / or proofreading my drafts.

My students, for brightening up my workdays by being so interested and appreciative.

My flatmates Steffen Müthing and Clara Cornaro, for being there for me throughout all these years, and patiently listening to and supporting me at all times.

My friends outside of university and my band, for being so faithful, fun and caring and for ensuring my work-life balance.

Ms. Ameis, for helping me make my decision to pursue this PhD.

Finally, I owe limitless gratitude to my parents and brother for always believing in me but loving me unconditionally, showing me that learning, work and research can be fun, and sharing their wisdom and experience. Finally, Großmama, for your many stories and your love. I am proud to have such an intelligent independent grandma.

Table of contents

List of tables

List of figures

List of abbreviations

$	Dollar
₹	Indian Rupee
ASHA	Accredited Social Health Activist
AWC	Anganwadi Center
AWW	Anganwadi Worker
BCG	Bacillus Calmette-Guérin
BJP	Bharatiya Janata Party
BMI	Body Mass Index
CCT	Conditional Cash Transfer
DHS	Demographic and Health Survey
DLHS	District Level Household and Facility Survey
DPT	Diphtheria, Pertussis and Tetanus
g/dl	Gram per Decilitre
GPU	Gram Panchayat Unit
HAZ	Height-for-Age Z-score
ICDS	Integrated Child Development Scheme
IFA	Iron and Folic Acid
IGMSY	Indira Gandhi Matritva Sahyog Yojana
ITT	Intent-To-Treat
IYCF	Infant and Young Child Feeding
JSY	Janani Suraksha Yojana
LPM	Linear Probability Model
MDG	Millennium Development goal
MKSS	Mazdoor Kisan Shakti Sangathan
NFHS	National Family and Health Survey
NGO	Non-Governmental Organization
NREGA	National Rural Employment Guarantee Act
PMMVY	Pradhan Mantri Matru Vandana Yojana
RCT	Randomized Controlled Trial
SA	Social Audit
SAU	Social Audit Unit
SC	Scheduled Caste
SC/ST	Scheduled Castes and Tribes
SD	Standard Deviation
SLX	Spatial Lag of X
ST	Scheduled Tribe
UT	Union Territory
VHAS	Voluntary Health Association of Sikkim

1 Introduction

This dissertation consists of three essays which evaluate the effects of two government interventions that aim to improve public service delivery in Indian maternity and employment programs (chapter 2-4). Maternity and unemployment are two key risks for health and income (ILO 2022). Maternal and child health and work hold intrinsic value of their own right by preserving physical integrity and social participation, but are also important determinants of income and thus poverty (Mwabu 2007; Janvry and Sadoulet 2016; ILO 2022). In addition, health and employment are closely interlinked, as wage income can be used to buy food and medical care and good health can enhance productivity (Mwabu 2007). In the context of maternal and child health, this is illustrated by Atkin (2009), who finds that access of Mexican mothers to comparatively good job opportunities is associated with greater height of their children, indicating a better health and nutrition status. Similarly, Rosenzweig and Schultz (1983) find that low-income families in the United States postpone prenatal care, which results in lower birth weight of children. Conversely, the health of mothers during pregnancy and lactation can determine their children's physical and cognitive human capital at birth and later in life, and thus their income-earning capability as adults (Currie 2009; Maluccio and Flores 2005; Maluccio et al. 2009; Miguel and Kremer 2004). Maternal cash transfer and work programs both aim to buffer short-term consumption shocks caused by pregnancy and unemployment, while at the same time laying the foundation for long-term welfare improvements through the prevention of poverty traps (ILO 2022; Janvry and Sadoulet 2016). Notwithstanding this, only 18.6 percent of workers worldwide who become unemployed have access to unemployment benefits (ILO 2022) and only 44.9 percent of women with newborns worldwide receive maternity benefits (ILO 2022; Ghosh and Kochar 2018).

The focus of this dissertation lies on two public interventions in the context of maternity and unemployment. These encompass, firstly, a maternity benefit program which aims to improve maternal and child health and avoid unintended fertility increases, and secondly, a community monitoring tool called external social audits which aims to improve service delivery by reducing corruption and mismanagement within a public work program. In the introduction (chapter 1), I discuss the general importance of public programs for development and poverty reduction, and how the aim to ensure their effectiveness and efficiency calls for an iterative learning cycle in which public policies are continuously re-evaluated and adapted (section 1.1). To underline the importance of public policy in India, and the current maternity and work programs in particular, I provide in section 1.2 a brief history of public social programs in India. In section 1.3, I compare and summarize the three essays and discuss how they contribute to the scientific literature and the iterative learning cycle introduced in section 1.1. At the end of the introduction (section 1.4), I draw general conclusions from the findings of the three essays.

1.1 Importance of public programs for development and poverty reduction

Public programs in the areas of health and unemployment are typically part of a greater state agenda to improve the wellbeing of its subjects. Amartya Sen's famous capability approach stresses that development is an increase in individual wellbeing and should be evaluated in terms of the enhanced ability of individuals to choose, pursue and reach individual life goals (Sen 1999, 1988). These life goals (so-called functionings) can refer to aspects as diverse as health and nutrition, education or a high standard of living. Having the capabilities to reach a greater number of these goals translates into a greater number of options to choose from and accordingly greater individual freedom, an important goal in itself. State intervention to improve individual welfare is typically justified by the presence of market failures and poverty traps, which hinder individuals from pursuing or reaching their individual wellbeing goals. In Sen's framework, income and ownership of private goods can be a means to achieve individual life goals. However, income is of little use to reach one's life goals if markets for fundamental private and public goods and services are absent or if access to existing markets and institutions is restricted (Sen 1988). Accordingly, the state can foster development by reducing income constraints as one determinant of wellbeing, but also by addressing the above-mentioned market inefficiencies through programs which produce and provide access to public and private goods and services (Sen 1999). Policies which explicitly pursue the goal to foster welfare are called social policies (Scott and Marshall 2009). However, as individual needs vary according to context, definitions of the concrete development and welfare[1] goals which social policies should pursue vary in practice by country (Pinker 2022). This thesis follows the broad definition of the World Bank (1999), according to which social policies aim to ensure "access to basic social services" (for instance education and health care), "secure and sustainable livelihoods, and decent working conditions", "social protection" (against the adverse effects of risks, for example to income and health), and "social integration". Although public social programs do not necessarily explicitly target the income poor, the latter are by definition particularly financially constrained and vulnerable to risks, and their access to some private markets is even more limited because it can be unprofitable for private providers to offer goods and services in areas where the poor live. Hence, the poor often depend to a disproportionate extent on publicly provided goods and services (Besley and Ghatak 2006; Sen 1999). Conversely, expanding capabilities in non-income areas, such as health or education, can help to reduce income poverty (Sen 1999). Public social programs can contribute to the reduction of income poverty in three ways (Janvry and Sadoulet 2016): first, decrease *chronic poverty,* second, decrease *vulnerability of non-poor* to become poor through income shocks; and third, prevent *deepening poverty of the poor* through shocks.

In the first poverty reduction mechanism, public social programs aim to decrease *chronic poverty* directly through non-contributory social assistance in the form of cash and in kind

[1] The terms welfare policy and social policy are used interchangeably in the European and Indian context. However, in some contexts, for instance in the U.S., the term welfare policies is restricted to policies solely targeting the poor and vulnerable, and not the general population (Pinker 2022).

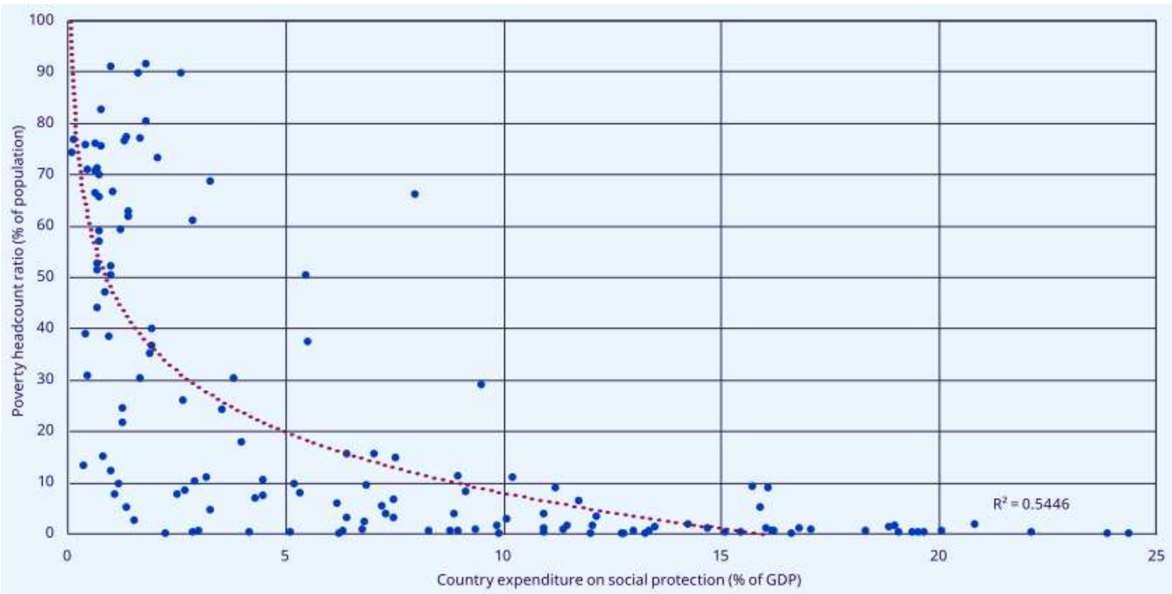
transfers, or indirectly by creating income-earning opportunities. Earning opportunities can be generated in the following ways (Janvry and Sadoulet 2016): (i) through the building of private productive assets. These assets can encompass physical capital such as certain infrastructure (e.g. private irrigation infrastructure for farmers), tools needed to pursue a certain occupation, financial capital, land, social capital (status and social networks) and human capital (health or education and skills). (ii) By changing the context in which the poor operate. For instance, providing access to institutions or certain public goods such as schools, roads or markets can be a prerequisite for the poor to access employment opportunities, transport and sell produced goods, and to open or expand their own business. Public infrastructure is also an important driver of economic growth and thus indirectly of individual incomes. Finally, state provision of infrastructure is particularly important for the poor as they often live in remote areas where it is not profitable for private providers to operate at all or to invest in the most expensive "last mile" of infrastructure networks. (iii) By incentivizing preferences and behavior which are conducive for income generation. For example, they can encourage the poor to make riskier but more profitable investments by informing them about project returns or by mitigating the penalty for risk-taking through social protection programs (described below). Public programs can also nudge behavior which increases the human capital of the poor (e.g. to engage in healthy behavior or enroll in school) and in this way expand their ability to work productively.

Both the second and third poverty reduction components aim to protect citizens from the poverty effects of temporary income shocks (through so-called social protection programs), but differ in whether they target the non-poor or poor population. In the second component, public programs aim to reduce *vulnerability of the non-poor to poverty* caused by shocks over the life cycle, so that they do not become poor when shocks hit. Social protection programs for the non-poor usually consist of social insurance, which can insure against risks such as unemployment, health, disability or even old age. It is usually at least partially contributory and set up before shocks hit (ILO 2022; Janvry and Sadoulet 2016). The example of the COVID-19 pandemic, which raised the number of children living in poor households from 356 million in 2017 to almost 725 million today (ILO 2022), illustrates how quickly external shocks can draw individuals into poverty. However, the pandemic also demonstrated how effective social protection programs, in particular in the areas of health and employment, can mitigate effects of such shocks and prevent people from falling into destitution (ILO 2022).

The third component addresses poverty through public programs by aiming to increase the *coping capability of the poor* to protect themselves when shocks hit from deepening poverty or falling into poverty traps (Janvry and Sadoulet 2016). These types of programs typically do not involve ex-ante contributions by the beneficiaries. For example, in the case of such diverse shocks like drought and ensuing famine, employment shocks or pregnancy, the government can set up social assistance programs which distribute free food, offer employment in work programs and pay non-contributory maternity benefits to compensate the income and consumption loss, respectively (Janvry and Sadoulet 2016; ILO 2022).

In these ways, poverty reduction through public programs can not only increase individual welfare of the poor, but also contribute to national economic growth by mitigating market failures and by reducing risk-taking penalties that hamper profitable investments (Janvry and Sadoulet 2016). Many countries acknowledge the benefits of public social programs and most recently, the COVID pandemic led many governments to expand or introduce new social protection programs (ILO 2022). Notwithstanding this, the existing systems suffer from deficits in quality and outreach (ILO 2022): as of 2020, 46.9 percent of the population worldwide benefit from social protection programs while 53.1 percent lack adequate protection (ILO 2022). Furthermore, low income countries typically have higher absolute poverty rates and missing markets are more common (for example in health, education and insurance against shocks) (World Bank 2022c; Dupas 2011; Mwabu 2007; Rodrik and Rosenzweig 2010), but this is not mirrored in higher social spending (ILO 2022; World Bank 2022a, 2022b). For instance, the percentage of the budget that low income countries dedicate to social protection is fifteen times lower than that of high income countries (ILO 2022). Figure 1.1 illustrates how higher absolute poverty is associated with a *lower* share of the budget being spent on social protection (causality can of course go either way).

Figure 1.1: Public social protection expenditure (excluding health), percentage of GDP, and poverty rates, 2020 or latest available year



Notes: Source: ILO (2022).

In addition to imperfect coverage and financing, the benefits of public programs can be impaired by

(i) ineffectiveness or unintended adverse effects – programs may fail to improve the envisaged outcomes (e.g. improvements in health or employment) for the targeted population (Duflo 2017), or have detrimental side effects (e.g. crowding out of private health suppliers) (Powell-Jackson et al. 2015).

(ii) waste of resources – corruption and mismanagement may lead to leakage of resources which could alternatively be employed to make the program more effective or increase outreach (Olken and Pande 2012).

Thus, policy makers that design public programs have to juggle several goals at the same time: to realize the intended welfare effects and minimize adverse effects, while at the same time using the existing government funds as efficiently as possible. The last point is particularly important for low-income countries, which operate on particularly tight budgets. In order to reach these goals, public programs need to be continuously evaluated, honed and re-evaluated in an iterative process (henceforth called iterative learning cycle) with regard to their effects in the field (Duflo 2017). To grasp all consequences of an intervention, such evaluations should take into account both intended and unintended effects (Angelucci and Di Maro 2016). Evaluations of intended effects assess whether the program's intended objectives have been reached. However, the benefits of public programs may be greater than they appear at first glance if there are positive spillover effects on untargeted population groups and localities, or effects on outcomes which were not directly incentivized. In contrast, benefits will be overestimated if adverse effects on such groups and outcomes are not taken into account (Angelucci and Di Maro 2016).

Notwithstanding this, Banerjee and Duflo (2011) argue that policy makers often fail to learn and improve policies due to what Banerjee and Duflo (2011) coin the "three Is": ideology, ignorance and inertia. As Duflo (2017) puts it: "Policy makers tend to design schemes based on the ideology of the time, in complete ignorance of the reality of the field, and once these policies are in place, they just stay in place". While it may be difficult for researchers to change policy makers' ideology, research can suggest new, innovative program designs and generate information on the effectiveness and efficiency of existing programs and specific program components. With regards to the latter, research can contribute to the iterative learning cycle outlined in the previous paragraph in two ways: first, by rigorously evaluating after each program design loop the effects of a new program or program component on its intended objectives, and second, by identifying potential unintended inefficiencies and weaknesses in the current program design. In this way, research only expands policy makers' information base, but could potentially also overcome bureaucratic inertia by generating public support for change.

Section 1.3 will explain in detail how the research in this dissertation contributes to the iterative learning cycle. However, first, in order to provide a better understanding of the historical context from which the evaluated programs evolved, and the relevance of public social programs in India, section 1.2 provides a short history of such programs.

1.2 History of public social programs in India

The foundations of India's national social policy were laid already before its independence, during its colonial past. The first cornerstone was the installation of early social protection laws during colonial rule, as a response to the increasing organization of laborers, which were continued in amended form after independence. The most important of these early policies were the Workmen's Compensation Act in 1923 and the Maternity Benefit Act in 1929, which mandated employers to compensate employees in case of occupational disease and injury, and pregnancy, respectively. However, these early benefits were restricted to the formal labor force, which till today constitutes only a fraction of the Indian work force (93 percent of the Indian labor force is employed in the informal sector (Ahuja 2020)). Labor protection of formal laborers was continued after India's independence in 1947 with greater involvement of the state in program financing and administration in the form of a public maternity, health and disability insurance (the Employees' State Insurance Act 1948). However, coverage was restricted to key industries perceived as relevant for economic development (Ahuja 2020).

The second set of cornerstones were the extractive nature of the colonial rule and the inaction of the regime in face of a series of famines "of genocidal proportions" (Ahuja 2020) at the end of the nineteenth century. These influenced the priorities set by the state after India's independence: policies in the 1950s and 60s focused on the reconstruction and growth of the Indian economy, as well as the attainment of food security (Pelissery 2020). Concrete policies to attain these goals were laid out in so-called five year plans. Food security was pursued through the introduction of new technologies in agriculture (which led to the 'green revolution' in the 1960s), through output and input price subsidies for farmers, and public food programs such as the Public Distribution System (Pelissery 2020). Moreover, as a major source of poverty-related hunger and inequality was seen in the prevailing caste system, the new constitution stipulated quotas which gave disadvantaged castes and tribes preferred access to public employment and public goods such as education (Ahuja 2020).

However, during this phase dominated the conception that social programs were a luxury which the state could extend to its citizens only within the scope of its financial abilities and India was viewed as too poor to afford extended welfare and poverty alleviation programs (Kumar 2005). Accordingly, poverty alleviation only appeared as a goal in the fifth development plan, at the end of the 1970s (Pelissery 2020). Similarly, public health programs initially focused only on the control of epidemics (and later population growth), and access of the poor and rural areas to health infrastructure was only tackled in the 1983 through the *National Health Policy* (Duggal 2001). Apart from this, state provision of public goods and services was often inefficient due to corruption and misalignment of state and national policies (Pelissery 2020).

The government's negligence and failure to provide public services and reduce poverty led in the late 1960s to 90s to a rise of non-governmental organizations (NGOs) who provided welfare-related goods and services (Pelissery 2020). The increase in NGOs was partially driven by foreign, mainly Western, donors who started to see NGOs as more effective agents of change than the state. However, in the 1970s, the increase in NGOs was also fueled by Indian citizens and ex-bureaucrats who were disappointed by the inability of the political system to implement the demands of citizens. Until today, NGOs remain important actors in India: in 2015, India had with 3.1 million registered NGOs more NGOs than hospital beds or schools, respectively (Anand 2015).

Notwithstanding this, the Indian welfare state expanded in the late nineties, following economic liberalization (Kumar 2005). Indian social sector spending tripled between 1990/91 and 2013/14 (Pelissery 2020). The perception that social programs were a luxury was increasingly replaced by the notion that citizens had a *right* to receive benefits (Kumar 2005). This culminated for instance in the nationwide *National Rural Employment Guarantee Act* (NREGA) which guarantees each citizen the right to a hundred days of employment through the state per year. Accordingly, although the present party in power, the Bharatiya Janata Party (BJP), is not known for its particular support of the poor, attempts by prime minister Narendra Modhi to cut social spending in 2015 were successfully overthrown after opposition from functionaries in both his own and other parties (Kapur and Nangia 2015; Kalra and MacAskill 2015). However, Kapur and Nangia (2015) argue that contemporary Indian social policy focuses on social assistance and protection policies rather than the provision of basic public goods and services because of the government's continued institutional incapacity to effectively provide quality public goods, and because large social protection programs are more visible to voters.

In line with these different economic possibilities, interests and demands, nationwide social programs have been set up since the late 1990s which increasingly benefit rural areas and target citizens which do not enjoy a formal employment status (Ahuja 2020). These programs include for instance the *Targeted Public Distribution System*, a food distribution system which, in contrast to its predecessor program, is restricted to the poor but extended to more rural regions; the *National Social Assistance Program*, a non-contributory assistance for the elderly, disabled and widowed poor; and *Rashtriya Swasthya Bima Yojana* and *Aam Aadmi Beema Yojana,* two subsidized health and life insurance programs for the poor, although expenditure on the last two programs is comparatively low (Kapur and Nangia 2015). In addition, the ILO (2022) highlights two areas in which the Indian government has made particularly marked progress in the last two decades: work programs and maternity benefits. This refers on the one hand to the above-discussed NREGA work program, which, in addition to the generation of income earning opportunities, also creates public infrastructure and individual assets for the poor which are meant to improve their living conditions and productivity. On the other hand, India introduced two maternity benefit schemes which for the first time extend maternity benefits to mothers not working in the formal sector at a national scale: first the *Janani Suraksha Yojana* (JSY), a one-time maternity benefit which is conditional on institutional delivery, and later the *Indira Gandhi Matritva Sahyog Yojana* (IGMSY), a repeated cash transfer during pregnancy and lactation which is conditional on the fulfillment of several maternal- and child-health-related requirements.

1.3 Summary and contribution of the three essays

As discussed in the previous section, India has made particular progress in expanding its employment and maternity benefit programs in the last decades. This dissertation assembles three essays, summarized in Table 1.1, which document and evaluate attempts of the Indian government to improve public service delivery in its maternity and work programs. The focus lies on two public interventions: the introduction of the maternity benefit program IGMSY and the introduction of external social audits - a form of community monitoring - in the work program NREGA in the Indian state of Sikkim. The first intervention, the IGMSY maternity benefit program, was designed to overcome an adverse effect - a fertility increase - and to achieve broader impacts on maternal and child health than the pre- (and now co-) existing similar JSY program[2]. In order to achieve this, policy makers changed the targeting and timing in the new IGMSY benefit, and added conditions which aimed to incentivize cheap preventive health and health seeking behavior. The second intervention, the introduction of external social audits in the NREGA work program in Sikkim, can be understood as an answer to several shortcomings in the NREGA program and to social audit designs in other states. First, NREGA suffered from widespread corruption and inefficiencies (Niehaus and Sukhtankar 2013; Muralidharan et al. 2016). Sikkim aimed to improve public service delivery by decreasing corruption and inefficiencies of NREGA through social audits. These gather information on public service delivery from public providers and beneficiaries, and identify discrepancies between the two in public hearings. Second, in an attempt to overcome deficiencies in the quality of social audits and to decrease audit costs, Sikkim diverged from the at the time predominant *user-driven* social audit design in which the auditors who collect information for verification are recruited from beneficiaries (Tambe et al. 2016a). Instead, Sikkim implemented a new e*xternal* social audit design, which is defined here as a particular form of social audits in which the auditors are all permanently employed external personnel. Differences between the two interventions arise not only from the type of programs which they aim to improve (maternity versus employment), but also in the interventions' focus on *how* to improve effectiveness and efficiency in these services, and in their planning level. As indicated in Table 1.1, while IGMSY primarily targets the demand side through health-enhancing behavior and demand for health inputs (with support of health staff), external social audits primarily target the supply side by altering the behavior of bureaucrats (with support of citizens). Aside from this, the IGMSY intervention was designed exclusively by bureaucrats at a national scale, while the external social audit intervention was primarily designed by a state-level organization which is public actor and NGO at the same time.

As outlined in section 1.1, public programs need to be continuously adapted and adjusted in order to ensure the achievement of the intended program goals and minimize negative side effects and waste of resources. Whether these adjustments are successful, or whether they are accompanied by certain weaknesses needs to be verified through the continued evaluation of the effects of new interventions. In this context, I view on the one hand the above-discussed interventions (IGMSY and external social audits) as part of this iterative learning cycle, and evaluate in which ways they represent attempts of policy makers to reform existing public programs to reach a maximum of the intended beneficial goals, eliminate unintended side effects and reduce waste of resources.

[2] The fact that policy makers installed IGMSY as an improved, additional program and kept JSY as complementary program instead of implementing the policy changes within JSY sadly seems to support Banerjee and Duflo's (2011) view that policies, once established, are never abolished.

Table 1.1: Comparison of the three essays

Essay	Research question	Outcomes	Intervention	Design change	Potential actors	Sample	Method	Results
1: Lessons learned? Intended and unintended effects of India's second-generation maternal cash transfer scheme	Can conditional cash transfers paid around time of birth improve maternal and child health at a national scale, and without encouraging the unintended side effect excess fertility?	- Mother and child health indicators - Public health system contact - Breastfeeding - Child vaccination - Fertility, birth spacing	Conditional cash transfer during pregnancy and lactation (demand)	- Conditional on health checks, nutrition counseling and vaccinations - Longer payment period - Restriction to first two life births	- Supply: health workers - Demand: mothers	13,367 eligible children born 2010-2013, and their mothers in 70 districts, 24 states	Difference-in-difference with placebo test	- Increased conditioned-on and not conditioned-on vaccinations - Greater contact with state health system - No adverse fertility effects - No robust improvement in health
2: External social audits: the value of experience. Combining external and community monitoring for improved public program delivery	Which average effects have long-term external social audits on expenditure, service delivery and use (in a public work program)?	- Detected irregularities - Program expenditure - Enrollment in the program - Ongoing, suspended and completed works	External social audits in a work program (supply)	External auditors	- Supply: politicians, bureaucrats - Demand: citizens, auditors	Work program outcomes for 7 years in 170 communities in Sikkim, and detected irregularities for 391 audits in 6 years	Panel fixed effects model with community and year fixed effects (except for irregularities)	- Increased efficiency in asset creation, increased material expenditure - No/negative effects on employment - Initial increase in corruption hiding counteracted
3: The spatial spillover effects of social audits	Do effects of social audits on work program outcomes spill over to neighboring communities?	- Program expenditure - Enrollment and days worked in program - Ongoing, suspended and completed works	External social audits of the work program in neighboring communities (supply)	External auditors	Audited community: - Demand: auditors Audited and neighbor community: - Supply: politicians, bureaucrats - Demand: citizens	Work program outcomes and geospatial location for 7 years in 166 communities in Sikkim	Spatial lag of X fixed effects model with community and year fixed effects	Spillovers on percentage of labor expenditure and work projects in neighbor communities similar to effects in audited communities

Note: Design change compared to previous interventions. Auditors represent interests of demand side but can influence social audit policy design.

On the other hand, the essays in this dissertation contribute themselves to this iterative learning cycle through the generation of new information on the effects of the new interventions outlined above. The effects of the IGMSY program have so far only been evaluated for a single Indian state, and a limited set of outcomes which excludes, for instance, maternal health and health seeking behavior which is not directly conditioned on in the program. Similarly, there exists so far only one study which evaluates the effects of external social audits (set in the Indonesian context). The respective study (Olken 2007), evaluates effects of a one-time change in social audit design and of the probability to be audited, and is therefore not suitable to capture long-term effects of repeated external social audits, which might be influenced by learning of the relevant actors. Moreover, it focuses solely on the outcome leakage of funds and does not evaluate whether quality and quantity of public goods and services, as well as service use improve. Finally, no study so far evaluates spatial spillover effects of social audits on neighboring communities. Thus, the effects of the above-discussed interventions are by no means clear.

The contributions of this dissertation are thus as follows: to shed light on the effectiveness of the two interventions, I evaluate their effects on a much broad range of both directly targeted outcomes and population groups as well as not directly targeted ones. Moreover, I take stock of the effectiveness of the respective changes and discuss whether policy makers and implementers learn effectively how to improve such interventions. More specifically, the first essay tackles the question whether the newly designed IGMSY maternity benefit is effective at improving maternal as well as child health at a *national* scale, whether it successfully avoids an adverse increase in fertility, and whether it has other not directly incentivized side effects on vaccination and service use. The second essay evaluates the effects of additional external social audits on detected irregularities, but also on indicators of the extent and use of public services provided - program expenditure, employment and asset creation - in audited communities over a period of seven years, while the third essay explores whether these effects of social audits spill over to neighboring communities.

As indicated in Table 1.1, I apply three different empirical approaches to elicit the causal effects of the respective program or program component. All these methods are quasi-experimental methods, which aim to establish *plausible* exogenous variation in the main explanatory variables. They are typically used in settings where randomization of the intervention is not possible, has not been considered by policy makers before implementation of the intervention, or if doubts arise that the randomization was successful[3]. The first essay applies a matched-pair difference-in-difference approach. This approach attempts to minimize potential selection bias into districts who received the IGMSY program in the following way: districts which received the program early (program districts) are matched to districts in the same state which have similar health development characteristics but received the program later (control districts). The intention-to-treat effect on the targeted population is estimated

[3] E.g. in some cases in which randomization is officially done by policy makers but without involvement of researchers, doubts can arise if policy makers did not *unofficially* prioritize certain groups or locations.

through a double difference in outcomes between eligible and ineligible children (according to their birth year) in program and control districts. In the second essay, I exploit the panel nature of my datasets, which allows me to apply a panel fixed effects methodology with community and, in most specifications, year fixed effects to estimate the effect of repeated external social audits on program outcomes in the next year. The fixed effects methodology eliminates potential biases in the estimates which could arise from unobserved time trends in outcomes common to all communities, as well as from unobserved community-specific characteristics which remain constant over time. It can be implemented by including indicator variables for each except one year and community as control variables. The third essay employs a 'spatial lag of X' (SLX) fixed effects panel approach. This augments the panel fixed effects approach by accounting for spatial correlation in the explanatory variable and thus omitted variable bias. It also explicitly estimates spillover effects to neighboring communities. The approach is implemented by adding a spatial lag of the explanatory variable (i.e. social audit completion in the neighbor community) to the regression. The three essays are summarized in detail in the following sub-sections.

1.3.1 Summary of the first essay: "Lessons learned? Intended and unintended effects of India's second-generation maternal cash transfer scheme"

The first essay (chapter 2), co-authored with Stefan Klonner, evaluates the IGMSY conditional maternity benefit program. The program aims to improve maternal and child health by incentivizing health-promoting behaviors and providing financial means for adequate nutrition during the critical phase of pregnancy, childbirth or lactation. IGMSY was designed to correct weaknesses of a previously introduced conditional cash transfer program, JSY, which pays cash transfers conditional on institutional delivery, but failed to improve child mortality and health and featured a number of adverse effects - among others an increase in fertility. IGMSY addresses these weaknesses by extending cash transfers to the critical nine-month period around birth and conditioning on the performance of a number of health- and nutrition-enhancing actions to improve health outcomes of mothers and children holistically, and further by restricting access to only the first two livebirths to avoid an adverse effect on fertility. We approach IGMSY's geographically targeted pilot phase as a natural experiment and use data from a large national health survey to estimate its effects by a matched-pair difference-in-differences approach. We find that the program improves both health care seeking behavior which is directly incentivized by conditions and behavior which was not explicitly conditioned on. In particular, the program increases infant immunization and the long-term utilization of primary public health institutions. Moreover, the program not only manages to avoid an adverse side effect, an increase in fertility, but also increases birth intervals between eligible children by 17 percent. However, we find no or only weak evidence for improvements in breastfeeding and in child and maternal health outcomes. Thus, while the redesign of the program successfully avoids unintended side effects, the improvements in maternal and child health are at best marginal. We ascribe this lack of transformative change to insufficient outreach of the program and poor quality of health services.

1.3.2 Summary of the second essay: "External social audits: the value of experience. Combining external and community monitoring for improved public program delivery"

The second essay (chapter 3) evaluates the effects of external social audits in the NREGA public work program for the Indian state Sikkim. Social audits are a type of community monitoring which involves the verification of official information on program performance by comparing it in public hearings to information gathered from beneficiaries. They were introduced to improve service delivery in the work program by combating corruption and mismanagement. While the traditional user-conducted social audits avail themselves of service users to collect the relevant information to be presented in the public meeting, external social audits employ external auditors to collect this information. Policy makers in Sikkim deviated from the blueprint of 'user-conducted' social audits practiced by other Indian states, in the hope to achieve social audits of higher quality and at lower cost, and implemented an 'external' social audit design instead. External auditors were anticipated to be more independent, educated and experienced in auditing. I evaluate the effects of additional external social audits on several work program outcomes using a panel fixed effects model. My findings show that additional external social audits reduce mismanagement and corruption in the creation of infrastructure assets through the work program. While program officials initially hide work projects from auditors, I find evidence that auditors learn and successfully adapt the social audit process to counteract hiding behavior. I also argue that external social audits are cost-effective. Notwithstanding this, I find that external social audits have no or even potentially adverse effects on labor expenditure, program enrollment and employment. In particular, I find a significant drop in enrollment due to fewer applications. As suggestive evidence indicates that particularly disadvantaged groups reduce their labor supply to a greater extent, this potentially threatens NREGA's outreach to the targeted poor unemployed population.

1.3.3 Summary of the third essay: "The spatial spillover effects of social audits"

As shown in the second essay, external social audits can be an effective tool to combat corruption and inefficiencies in program implementation in audited communities. However, external social audits also have some (albeit sometimes only temporary) unintended effects such as hiding of corruption and a decrease in program enrollment. In addition to these direct effects, social audits could influence outcomes in *nearby communities* positively or negatively through several pathways. For instance, information about incidence and outcomes of social audits could spread to nearby communities and change behavior of local actors by increasing the perceived probability of detection and punishment, or changing local social norms equilibria. Finding *beneficial* spillover effects, e.g. increased efficiency in asset construction, could lend additional support to the continuation of this intervention since estimates which do not take spillovers into account underestimate the true effects (and thus also cost-effectiveness) of external social audits. On the other hand, knowledge about adverse spillover effects, for example the hiding of corruption, is important for policy makers to take countermeasures. In the third essay, "*The spatial spillover effects of social audits*" (chapter 4), I investigate such spillover effects in the context of the Indian NREGA program, using a panel fixed effects spatial lag of X model. My findings indicate that social audits in neighboring communities influence program outcomes in the same direction as having an audit in one's own community, leading to a multiplicator effect. In particular, I find a significant increase in ongoing and suspended works, and a significant decrease in the percentage of labor expenditure. While I cannot rule out that corruption hiding strategies of bureaucrats spill over to neighboring communities, it is worth emphasizing that a notable, potentially adverse direct effect of social audits on enrollment in NREGA does not spill over to nearby communities. I argue that this suggests that villagers not yet enrolled in the work program lack information on NREGA social audit results in neighboring communities.

1.4 General conclusion

The three essays in this dissertation evaluate the effects of two interventions within public programs which aim to mitigate the effects of shocks and improve individual welfare. These interventions represent attempts to improve public service delivery by maximizing positive effects, eliminating negative side effects and minimizing waste of resources. The innovations of this dissertation are as follows: the first essay contributes to the literature by investigating the effects of the IGMSY maternity benefit program at a national scale and for a broad range of maternal- and child-health-related outcomes. The second essay expands the literature on social audits by estimating the effects of additional *external* social audits on work program service delivery outcomes, and by estimating the average effects of external social audits which have been in place over an extended period of time. The third essay pioneers by estimating spillover effects of (external) social audits to neighboring communities. Moreover, as both interventions evaluated in this dissertation, IGMSY and external social audits, evolved from previously introduced interventions with similar aims, both the results in essay one and two are complemented with a comparative discussion of existing evidence on these alternative intervention designs.

Overall, the three essays demonstrate that the evaluated interventions often manage to improve outcomes that can be interpreted as intermediary outcomes towards the higher order program goals to improve maternal and child health, and provide high-quality infrastructure and employment to the poor. These intermediary outcomes are for instance child vaccination and long-term health seeking behavior in IGMSY. I also find evidence for a reduction in the waste of resources, such as the removal of inefficiencies in infrastructure production and cost-effective reduction of embezzlement through external social audits.

Notwithstanding this, the essays also reveal that these improvements often only lead to small or no lasting improvements in the lives of the targeted population in terms of enhanced maternal and child health, employment and a higher number of infrastructure assets produced (although I find some evidence that asset quality rises and that poor households may profit from a greater poverty focus in the selection of beneficiaries of private assets constructed under NREGA). Indeed, a conclusion in both the first and second essay is that, in order to enhance the impact on the population, policy makers need to pay more attention to reach their respective target group. In the case of IGMSY these are mothers who are currently not fulfilling the health behavior conditions of the cash transfer, in the case of external social audits in NREGA these are the rural unemployed poor and otherwise disadvantaged (in particular citizens who are not yet enrolled in the NREGA program).

The essays also provide evidence for the existence of both beneficial and adverse side effects: in the first essay, we find evidence for positive side effects of IGMSY on measles vaccination and long-term health seeking behavior, whereas I uncover in the second essay adverse unintended effects of external social audits on audit avoidance by bureaucrats and potential adverse effects on program enrollment by citizens. Finally, the third essay shows that external social audits also influence some outcomes in neighboring communities, although the decrease in employment does not spill over to neighbors. However, I argue that changing contexts, for instance greater collaboration with higher-level departments in the implementation of NREGA, may change the dynamic of spillover effects in the future.

In conclusion, when combining the evidence from the three essays, two aspects stand out as particularly important: the essential role of learning of all stakeholders, and the need to put a greater focus on outreach and ultimate welfare goals. Regarding the importance of learning, the three essays contribute to the evidence base in the iterative learning cycle outlined in section 1.1 by evaluating the new IGMSY and external social audit interventions both with regards to their effectiveness and their weaknesses, and taking into account also secondary effects such as spatial spillovers or unintended effects. A first general implication of the three essays is that careful re-design of policies manages to successfully avoid adverse side effects *known* to policy makers, for example an increase in fertility in IGMSY, and hiding behavior of bureaucrats in external social audits. This emphasizes the value of information regarding program's side effects for policy makers and shows that research and experience can provide an escape to the "ignorance" bias evoked by Banerjee and Duflo (2011)[4]. However, an important difference between the setting in the second and third essay (external social audits) compared to the first essay (IGMSY) is that policy makers in the former not only learn from previous external research, but also based on their own or their colleague's (in this case: auditors') continued observations and experience. As external social audits are repeated regularly, and new information is thus frequently generated, policies are also updated more often to incorporate this new information. In addition, the fact that the external social audit intervention is designed and implemented at a lower (state) level than the nation-wide IGMSY program, as well as the involvement of an NGO in the design, might be conducive to new information reaching policy makers quickly. Moreover, all three essays highlight the influence of learning of other stakeholders on the examined outcomes. For example, in the first essay, mothers learn to use public health services more frequently, in the second essay, community-

[4] The mere fact that policy makers try to address such weaknesses indicates that inertia is of less concern in the setting of this dissertation than predicted by Banerjee and Duflo (2011).

level bureaucrats temporarily learn how to hide corruption, and in the third essay, communities learn from external social audit experiences of neighbor communities.

The second general implication relates to the attainment of development goals. As outlined in section 1.1, the ultimate goal of social programs is to increase the welfare of citizens, in particular those with low access to private means to enhance their wellbeing, who typically also have high deficits in wellbeing. However, the three essays show that so far, re-design has been less successful at improving several important welfare objectives of the respective programs - maternal and child health, employment generation and provision of a greater number (not only higher quality) of infrastructure assets - an aspect that policy makers need to devote more attention to. In particular, they need to improve outreach to the targeted population groups (the unhealthy, poor and unemployed). The three essays identify important information gaps which need to be closed to devise further optimal policy adjustments: in particular, the behavioral and informational mechanisms which catalyze or inhibit changes in program outcomes must be explored. To this end, we need to overcome data restrictions, particularly at the community level.

2 Lessons learned? Intended and unintended effects of India's second-generation maternal cash transfer scheme

Published in *Health Economics*

With Stefan Klonner

Abstract

The maternity benefit scheme piloted as IGMSY since 2011 and recently rolled out as PMMVY incentivizes mothers to participate in infant health-promoting activities. It has become India's largest conditional cash transfer program ever, outrivaling the country's first-generation maternity benefit scheme JSY, which incentivizes institutional delivery and has been criticized for its unintended side effects on fertility. We approach IGMSY's geographically targeted pilot phase as a natural experiment and use data from a large national health survey to estimate its effects by a matched-pair difference-in-differences approach. Consistent with the program's conditions, we find increases in infant immunization. As side effect, long-term utilization of public health facilities becomes more frequent and intervals between eligible births increase by 17 percent. Our findings suggest that India's second-generation maternity benefit scheme has been more carefully designed than its predecessor, with side effects that support the program's broader objectives. But both direct and indirect effects are small and can make only a small contribution to redressing India's dismal maternal and child health record.

2.1 Introduction

Poor health, nutrition and health care in early childhood affect long-term physical and cognitive development (Currie 2009; Maluccio and Flores 2005; Maluccio et al. 2009; Miguel and Kremer 2004). Despite rapid economic growth since the 1980s, India accounted for almost one third of global infant deaths and 40 percent of low-weight births by the beginning of the 21st century (World Bank 2019; UNICEF and WHO 2004). As a response, India started exploring conditional cash transfer programs (CCTs) which incentivize health-promoting behaviors during the critical phase of pregnancy, childbirth or lactation. The first such CCT, the *Janani Suraksha Yojana* (JSY), was rolled out in 2005 with a short-term focus on birth itself. It pays a one-time cash transfer to mothers conditional on institutional delivery or skilled assistance for delivery at home. Notwithstanding this, five years later, India continued to show marked deficits in perinatal health service use and health markers among infants and young children, especially weight-related indicators (WHO 2012).

The numerous existing evaluations of JSY provide some insights into why it largely failed to redress the country's dismal maternal and child health record. While institutional deliveries went up substantially (Lim et al. 2010; Powell-Jackson et al. 2015; Rahman and Pallikadavath 2018), none of these studies finds an effect on maternal mortality and several do not document any reductions in neonatal mortality (Lim et al. 2010; Powell-Jackson et al. 2015). Moreover, the program has had several unintended effects, including a fertility increase (Powell-Jackson et al. 2015; Nandi and Laxminarayan 2016), which has previously been documented for maternal CCTs in other countries (Powell-Jackson and Hanson 2012; Stecklov et al. 2007), as well as substitution away from private health care providers (Powell-Jackson et al. 2015). On the other hand, it is broadly acknowledged that JSY has also had a number of positive side effects regarding breastfeeding (Powell-Jackson et al. 2015), infant immunization (De and Timilsina 2020) and contact with the public health system later in the child's life course (Glick, 2017). The program features held responsible for JSY's failure to improve maternal and child health and its unintended effects on fertility are its narrow focus on institutional delivery, the short time interval around delivery covered by the program, cash incentives for all live births including higher parities, and a lack of qualitatively adequate health infrastructure supply (Powell-Jackson et al. 2015; Lahariya 2014).

As a consequence, JSY has been complemented by an additional, second-generation maternal CCT. The scheme, introduced in 2011 as *Indira Gandhi Matritva Sahyog Yojana* (IGMSY) and later renamed *Pradhan Mantri Matru Vandana Yojana* (PMMVY), incentivizes a broader range of healthy behaviors around the time of birth and features a number of trainings, including family planning. It covers a longer time interval of nine months around delivery and includes additional supply-side financing (Ministry of Women and Child Development 2011). During its five-year long pilot phase, cash transfers were paid for the first two live births of women aged 19 and older while eligibility has been restricted to the first live birth since the program has been expanded to India in its entirety in 2017.

In this paper we are first to assess the effects of India's second-generation maternity benefit program with nationally representative data. While this is an endeavor well worth in its own right, we will put a particular focus on whether the design improvements of IGMSY/PMMVY relative to its predecessor have been effective in improving maternal and child health outcomes on the one hand and avoid undesired side effects on the other. In addition, the extended pilot phase of five years and the timing of the national health survey on which we draw, which was fielded four years after IGMSY's introduction, provide the unique opportunity to also uncover medium-term effects of a maternal CCT whose cash transfers end six months after a child's birth. We approach the program's pilot phase as a natural experiment employing a matched-pair differences-in-differences estimator. We exploit the feature that the 52 pilot districts were selected based on district scores computed from a previous health survey to identify 52 control districts, one for each pilot district. We estimate intent-to-treat (ITT) effects of the program by comparing the within-district difference in health outcomes between younger birth cohorts exposed to the program and older cohorts not exposed to IGMSY across pilot and control districts.

Consistent with IGMSY's incentives, we find that polio, DPT and BCG vaccinations increase. As an indirect effect, measles immunizations, which are administered well beyond the period covered by the scheme, also increase. As a consequence, complete infant immunizations increase by nine percent. We also document two positive side-effects: mothers of once eligible children report fourteen percent more contacts with the government health system three to four years later. Moreover, there are no adverse effects on fertility, and birth intervals increase by eleven percent on average and by 17 percent between the first two parities, which are covered by the program. On the other hand, similar to JSY, we find no robust evidence of increased breastfeeding and gains in health outcomes, albeit some of our results suggest improvements in breastfeeding duration, child mortality and weight-related outcomes for both children and mothers.

In sum, our main finding is that the design of IGMSY/PMMVY – with several successive cash transfers paid over nine months, eligibility restrictions regarding parities and a broader range of conditions incentivizing health service use and trainings – results in direct and indirect program effects that are aligned with the scheme's intentions. In particular, there are no adverse side effects on fertility. We attribute this to repeated contacts with health workers, trainings that include family planning and the restriction of eligibility to the first two parities. On the downside, all our estimated program effects including health-services use are small and certainly not transformative. In our view, the discrepancy between IGMSY's coverage of around 50 percent of eligible births and the program's direct effects of no more than 6 percentage points is likely a result of its limited outreach and self-selection of mothers into the program who would have sought institutional perinatal care in the scheme's absence. In this perspective, IGMSY/PMMVY can largely be viewed as an income support program for mothers with a high ex-ante demand for government health care. This pattern stands in marked contrast to the earlier JSY program, which has achieved coverage rates of close to 80 percent in some of the prioritized areas by 2007 (Modugu et al. 2012).

Our findings have important policy implications. On the one hand, they show that deliberate policy design can avoid unintended side effects. On the other hand, they make clear that even when the design of a maternal CCT succeeds in aligning incentivized and indirect effects, the overall improvements are not sufficient to significantly improve health outcomes in children and mothers as long as the program fails to reach every other eligible mother. Accordingly, we argue that the outreach of such schemes needs to be massively improved and additional policies are warranted which identify and tackle the causes of health risks and simultaneously raise the quality of health services.

Our study contributes to a rapidly growing literature on the effectiveness of CCTs in the context of child and maternal health (Ranganathan and Lagarde 2012; Bastagli et al. 2019; Glassman et al. 2013; Grépin et al. 2019). Most closely related to our work are the numerous studies of India's JSY program (Powell-Jackson et al. 2015; Lim et al. 2010; De and Timilsina 2020; Glick) and Ghosh and Kochar (2018), who evaluate IGMSY with primary data collected in two districts of Bihar, India's most destitute state. Our principal contributions to this literature are, first, that we evaluate this IGMSY program on a national scale. Second, we consider a wide range of incentivized and indirect effects in a systematic fashion and, third, we comparatively assess the performance of a second-generation maternal CCT relative to its precursor.

The rest of the paper is structured as follows. Section 2.2 introduces the IGMSY/PMMVY program and discusses anticipated effects. Section 2.3 lays out our empirical strategy. Section 2.4 describes the data including balancing tests. Results are in section 2.5 and section 2.6 concludes.

2.2 Background

2.2.1 The IGMSY/PMMVY conditional cash transfer program

India's first mother-and-child-health-related CCT, the JSY, was partially motivated by the fifth millennium development goal (MDG), to improve maternal health, whereby one of two progress indicators is the proportion of births attended by skilled personnel (UN DESA 2015). The program has not been universal: while all births of mothers have been eligible for a transfer of ₹ (rupees) 1,400 in rural and ₹ 1,000 in urban areas of ten comparatively poor states, only the first two births of mothers in below-the-poverty-line households or from disadvantaged social groups have been eligible for ₹ 700 and ₹ 600 in all other states (Lim et al. 2010). However, given the program's unintended effects on fertility and the failure to reach the higher-order policy goal of improving child and maternal health (Powell-Jackson et al. 2015), policymakers recognized that JSY fails to cover the wage loss of mothers (Ministry of Women and Child Development 2011), which may prevent their rest and adequate nutrition during pregnancy and lactation. Furthermore, it does not incentivize mothers to engage in behavioral practices beneficial to both mother and child that go beyond safe delivery, such as adequate nutrition, preventive health care or birth spacing (Ministry of Health and Family Welfare 2011b).

The *Indira Gandhi Matritva Sahyog Yojana* (Indira Gandhi Motherhood Support Scheme) has been an attempt to fill this gap. It aims to improve maternal and infant health through a conditional cash support during the time of pregnancy and lactation. Unlike JSY, this program has been universal in the program's pilot districts, covering the first two live births of all women (with the exception of public sector employees). The restriction to two parities was imposed in order to counter-act adverse fertility incentives (Ministry of Women and Child Development 2011). The IGMSY cash benefit of initially ₹ 4,000 (approximately US$ 65) is equivalent to 7.3 (6.0) times the monthly rural (urban) poverty line or 31 female unskilled agricultural daily wages in 2011. It is funded by the national government via the state and district branches of the Integrated Child Development Scheme (ICDS). Mothers who are eligible for both IGMSY and the continued JSY may receive benefits from both these programs at the same time. During its early years, the IGMSY transfer has been sent to the mother's bank account in three installments: at the last trimester of the pregnancy and three and six months after the delivery, conditional on the program conditions.

All activities related to the conditions of IGMSY take place at local primary health care centers, known as *Anganwadi* centers. Compliance with the program's conditions is monitored there and the *Anganwadi* staff also receives a monetary incentive of ₹ 100 to 200 per completed case. In order to receive the first installment, mothers have to register their pregnancy, receive one antenatal check-up as well as tetanus immunization, collect iron and folic acid (IFA) tablets and participate in one nutrition and health counseling session. After delivery, the child's birth has to be registered to receive the second installment. In addition, for both the second and third installment, the mother needs to attend two counseling sessions on child nutrition and the child's weight is to be recorded twice. IGMSY directly incentivizes the recommended number of doses of three child immunizations: one dose of BCG (Bacillus Calmette-Guérin), which protects primarily against tuberculosis, and three doses of polio and DPT (diphtheria, pertussis and tetanus) vaccine, respectively. The first two doses of polio and DPT are conditions for receiving the second installment, whereas the third dose of polio and DPT has to be completed before the third installment, alongside with exclusive breastfeeding of the child for six months.[5] As IGMSY is meant to complement JSY, it does not incentivize facility deliveries.

With at least one Anganwadi worker in most villages, the public primary health care infrastructure necessary to fulfill the program conditions was already dense at the time of IGMSY's rollout (IIPS 2010b) and, at least in principle, free of charge. In addition, part of the government's budget allocation for IGMSY has been dedicated to hiring of additional staff in Anganwadi centers of pilot districts, to prevent an undersupply of the incentivized services (PIB 10/20/2010).

[5] The conditions and timing of the cash transfers under the original scheme are summarized in Table A2.1 in the appendix.

During a pilot phase, the program was geographically targeted and implemented in only 52 of India's 640 districts (Ministry of Women and Child Development 2019).[6] Unlike the majority of geographically targeted welfare programs in India, the pilot phase of IGMSY has not focused on the country's most destitute areas. Instead, a deliberate attempt was made to identify a set of districts representative of the country as a whole. The following stratified selection procedure was employed (Ghosh and Kochar 2018; Ministry of Women and Child Development 2011): First, an index of six health and development indicators was calculated from the third round of the District Level Household and Facility Survey (DLHS-3), fielded in 2007 and 2008. Second, according to this index, all 640 districts of India were categorized as either low, medium or high-performing. From these three groups, the pilot districts were randomly selected; eleven from each of the high and low-performing categories, and twenty-six from the medium-performance category. The remaining four districts were union territories (UTs).

In October 2010, the program was approved for implementation by the central government with budgetary allocations of ₹ 3.9 and ₹ 6.1 billion for the fiscal years 2010/11 and 2011/12 (PIB 10/20/2010). According to our calculations, the latter amount corresponds to roughly ₹ 4,600 per eligible birth. Program guidelines were agreed upon between the center and the states in April 2011 and training of implementation staff on the ground was to be completed by the end of May 2011 in all states (Ministry of Women and Child Development 2011). However, transfers from the central to the state governments were in some cases delayed until September 2011 (Sinha et al. 2016). For the pilot districts included in our analysis, state expenditures per eligible child on IGMSY were only ₹ 6 in 2010/11, rising to 3,190 in 2011/12 and 3,438 in 2012/13.[7] Only one state (Meghalaya) spent any funds in 2010. Despite identical numbers of target beneficiaries in 2011 and 2012, the amount of funds expended by state governments in 2011 was only half of what was spent in 2012 (Falcao et al. 2015). This suggests that the program was up and running only in the second half of the financial year 2011/12. As a consequence, there were virtually no beneficiaries in the fiscal year 2010/11. In the following two fiscal years, about 28 and 59 percent of the 1.2 million target beneficiaries were reached (Falcao et al. 2015). In 2013/14, the cash transfer increased from ₹ 4,000 to ₹ 6,000 (implying expenditures of ₹ 8,458 per eligible child) and was paid in two rather than three installments.

In 2017, the program was renamed *Pradhan Mantri Matritva Vandana Yojana* (PMMVY), in English the Prime Minister's Reverence for Maternity Scheme, and expanded to all districts of India with a cash transfer of ₹ 5,000 per woman paid in three installments, two during pregnancy and one after. Only the first live birth is eligible under PMMVY. There are two stated objectives of this program. First, to provide "partial compensation for the wage loss [...] so that the woman can take adequate rest before and after delivery" and second, that "the cash incentive provided would lead to improved health seeking behavior amongst the pregnant

[6] Due to the separation of Kundagaon from the pilot district Bastar, this number increased to 53 districts in 2012.

[7] For our definition of program expenditure per eligible case and data sources see appendix A2.1.

women and lactating mothers" (PIB 12/28/2017), in particular, safe delivery and immunization of firstborn children (PIB 12/6/2019).

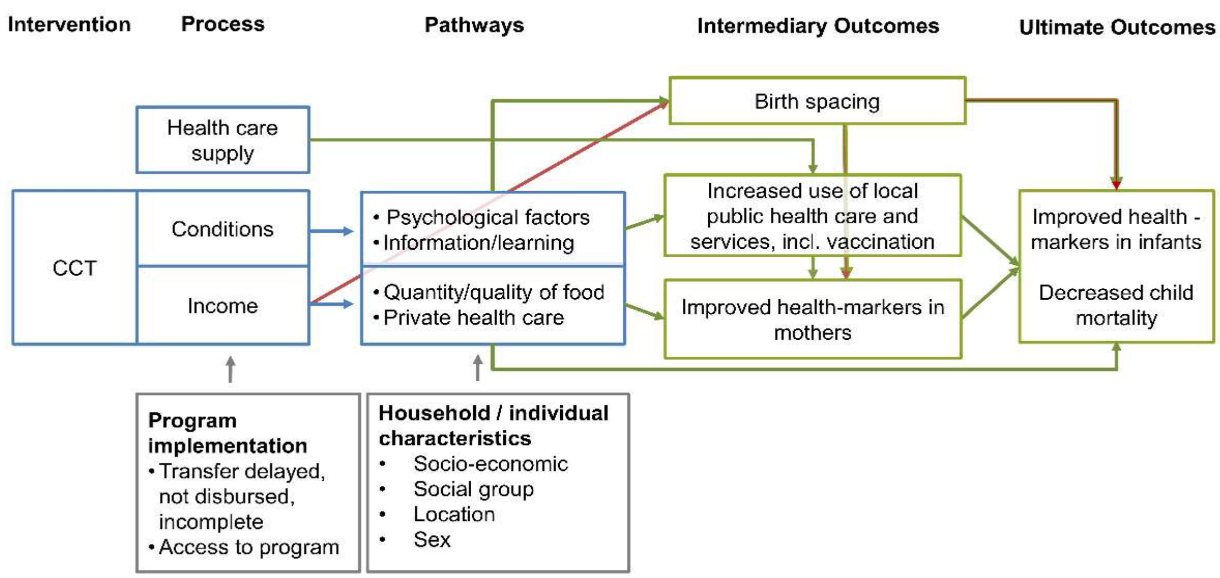
In terms of expenditures, with a budget allocation in 2017/18 of 27 billion rupees (US$ 400 million), PMMVY is India's largest conditional cash transfer program ever, accounting for 12 percent of the Ministry of Women and Child Development's budget or 0.13 percent of government expenditures. In comparison, the older JSY program, which has paid a maximum of ₹ 2,000 for an institutional delivery as of 2015, had a budget allocation of 20 billion rupees in 2017/18. On the other hand, according to administrative data, JSY reached out to 10.0 million women during the financial year 2018/19, while there were not more than 7.5 million beneficiaries under PMMVY. Still, like the IGMSY pilot, PMMVY appears underfunded relative to the program entitlements. A back-of-the-envelope calculation suggests that twice the allocated budget would be needed to adequately fund the roughly 9.5 million first births in India in 2018.

2.2.2 Anticipated effects

Perinatal conditions contribute to persisting deficits in child health as unhealthy mothers are more likely to give birth to unhealthy children (Black et al. 2008). In the following, we discuss how we expect the different elements of IGMSY to affect maternal and child health. We draw on Gaarder et al. (2010), who develop a theory of change for the impact of health-related CCTs, which we adapt to the setting of perinatal health and the Indian social context (see Figure 2.1).

First, conditions for the receipt of cash transfers may directly incentivize healthy behaviors in the form of increased demand for maternal and child health inputs. In the case of IGMSY, the program conditions not only directly mandate prolonged exclusive breastfeeding and vaccinations but also require the repeated interaction with a health worker. This raises the

Figure 2.1: Theory of change



probability of prevention, early detection and treatment of health deficiencies in mothers and children. Second, IGMSY may indirectly promote healthy behaviors, including birth spacing decisions through information conveyed in the incentivized educational sessions (Ministry of Health and Family Welfare 2011a). Third, there can be an income effect due to the cash transfer, whose funds are likely to be spent on additional healthcare and food during the critical phase of pregnancy, childbirth and lactation (Quisumbing and Maluccio 2000). Fourth, the additional income may also contribute to maternal health by reducing the labor supply of beneficiary women during pregnancy. On the other hand, the cash transfer can serve as an unintended incentive for impatient parents to shorten the birth interval between the first and second child (Powell-Jackson et al. 2015). Short birth spacing in turn contributes to adverse health outcomes for children and mothers (Cleland et al. 2012). Finally, the expansion of health service supply, which has been part of IGMSY, may lift access barriers to health care and thus increase health care use regardless of the cash transfer or its conditions.

Accordingly, regarding birth spacing, we anticipate two opposing effects: a positive one from trainings and more frequent interactions with the government health system which strongly promotes family planning, as well as a greater focus of parents on child quality rather than quantity; and, at least for the first birth interval, a negative one from the prospect of extra cash for the next pregnancy.

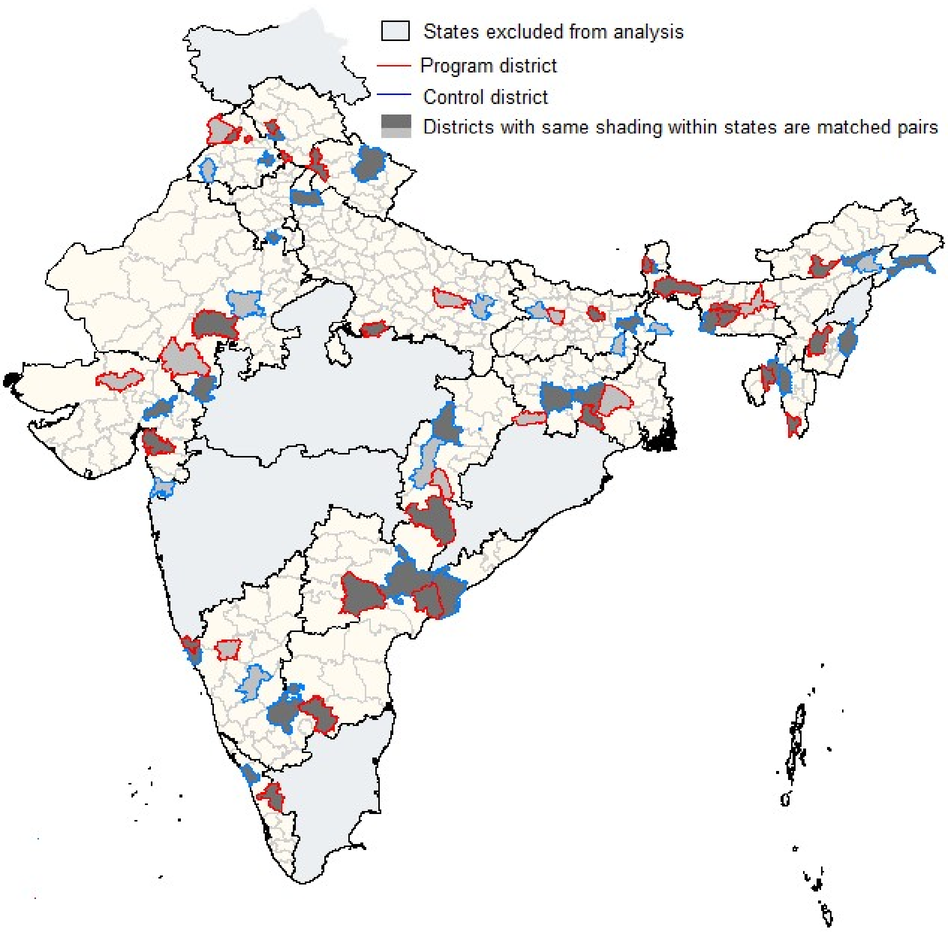
We expect unambiguous positive effects on health service use, in particular vaccinations, on breastfeeding and on maternal health. As the literature suggests large gains in longer-term child health and mortality through these factors (WHO 2010), we also expect improvements in child health markers at the time of the survey interviews, when children are three years old on average. On the other hand, mothers' health markers at the time of the survey interview may or may not improve as the beneficial effects on maternal health around the time of birth will likely dissipate faster than for children.

The program's impacts on different population groups may vary. If, due to poor implementation, the cash transfer does not reach eligible mothers, this will mitigate the program's effect on the children concerned. Similarly, individual and household characteristics likely play a role for the program's effectiveness. We expect traditionally disadvantaged households – those that are poor, from a scheduled caste or tribe (SC/ST) or living in the rural areas (Balarajan et al. 2011) – to profit more from the program (Gaarder et al. 2010), unless these characteristics restrict their access to the program (Powell-Jackson and Hanson 2012). In addition, effects may disproportionately benefit girls as Indian parents tend to invest more in boys' than in girls' health (Asfaw et al. 2010). Furthermore, we hypothesize that benefits are concentrated on first- and second-born children if the eligibility rules are enforced.

2.3 Empirical approach

We approach the pilot phase of the program as a natural experiment. Similar to the methodology of Ghosh and Kochar (2018), we elicit intention-to-treat effects (ITT) of IGMSY through a matched-pairs difference-in-difference approach. We compare the effect of the program on children that were eligible for the program and those that were not along two dimensions: first by taking the difference between eligible and ineligible cohorts within each district and second by comparing pilot to control districts. We identify one control district for each pilot district through a matched-pair design (see figure 2.2). First, we recalculate for all districts the health development index used by the Indian government to select pilot districts, based on data from DLHS-3 reports (IIPS 2010b). Index scores of all districts in our analysis are set out in Table A2.2 of the appendix. Next, we select within each state the nearest neighbor of each pilot district in terms of this index.[8]

Figure 2.2: Matched pairs of pilot and control districts



Ghosh and Kochar (2018) report that pilot districts were chosen randomly. If the selection of pilot districts within each stratum was non-random but driven by political or need considerations, however, this could bias naïve cross-sectional estimates of the program's effect.

[8] For the selection of control districts within states we treat Telangana, which separated from Andhra Pradesh in 2014, as part of Andhra Pradesh.

Our empirical approach addresses this concern in two ways, first by selecting control districts that are most similar to the program districts according to a pre-determined score. Second, in addition to the cross-sectional comparison, we also compare across exposed and not-exposed birth cohorts within each district. Differences between the latter are very unlikely to be affected by administrative selection as both types of cohorts are born subsequent to the program's planning stage in early 2010.

In principle, all women who fulfilled the eligibility criteria in December 2010 and thereafter were eligible to receive IGMSY benefits (Ministry of Women and Child Development 2011). However, given that the training of the ICDS staff was scheduled to be completed no earlier than June 2010, we expect registration of beneficiaries, which is possible until the fourth month of pregnancy, not to have started much before August 2011. As a consequence, the first cohort of children that profited from IGMSY would be born five months later in January 2012. Therefore, our estimator compares the difference in outcomes of children born in 2011 or earlier to children born in 2012 or later within pilot districts to the same difference in control districts of the same state. The somewhat fuzzy onset of the program and our classification of treatment and control cohorts imply that some children in our control cohort might already have had access to the program. Hence, our estimator delivers a lower bound of the actual program effect.

Our key regression equation is

$$H_{sdit} = \alpha_{sd} + \mu_t + (\delta_s \times eligbirth_t) + \beta(P_{sd} \times eligbirth_t) + \gamma X_{sdit} + u_{sdit}, \tag{2.1}$$

where s indexes states, d districts, i children and t birth year; H_{sdti} is an outcome of interest; P_{sd} is a dummy variable indicating whether a child lives in a program (or pilot) district; $eligbirth_t$ specifies whether a child was born in 2012 or later; X_{sdit} is a vector of individual and household-level control variables;[9] α_{sd} and μ_t are district and birth year fixed effects, respectively: the former account for cohort-independent differences between districts, the latter capture average cohort-specific differences between children common to both pilot and control districts. The term $\delta_s \times eligbirth_t$ captures state-specific time trends and u is a stochastic error term. The coefficient β gives the ITT effect of IGMSY. In line with common practice, we cluster standard errors at the level at which program status varies, the district (Abadie et al. 2017). The empirical approach to estimate the effect on mothers' outcomes is analogous, with eligibility defined by a dummy variable that takes on a value of one if the

---

[9] Control variables are listed in Table 1 and derived from Ghosh and Kochar (2018) and De (2017), who particularly stress the importance of maternal education as a driver of child health and vaccinations. In our research design, we are unable to control for father's characteristics (Ghosh and Kochar 2018), as these are only available for a small subsample, or health worker coverage (Anand and Bärnighausen 2007) since sufficiently detailed panel data on it is not available for India at the district level. To ensure that such unobserved factors do not drive our results, we also carry out extensive balancing and placebo tests in the sequel.

mother gave birth to a first or second-born child in 2012 or later (for the detailed regression equation, see appendix A2.1).

We account for the threat of falsely rejecting the null-hypothesis of no effect when estimating several outcome variables with the same sample by adjusting the p-values for the respective null hypotheses for families of outcomes as suggested by Romano and Wolf (2005). This method corrects for multiple inference through bootstrapping by controlling the family-wise error rate.

2.4 Data

2.4.1 The National Family Health Survey

Our data source for health service use as well as health outcomes is the seventh round of the Indian Demographic and Health Survey (IIPS and IFG 2018), commonly known as the fourth National Family and Health Survey (NFHS-4). This survey of more than half a million households was fielded between January 2015 and December 2016. We derive our outcome and control variables from the children's dataset, which contains data on 259,627 children born five years before the survey took place and later. This means our data covers roughly half of the 2010 birth cohort and all children in later cohorts. We use the sampling weights included in the data throughout to make all figures representative for the respective populations.

We use only a subset of the observations in the NFHS children's module. First, to ensure that control districts are free from other similar programs, we eliminate states that operated additional state-specific and state-wide maternity benefit programs with cash transfers in 2011 or 2012.[10] Second, we exclude UTs[11] and Nagaland, which were not surveyed in DLHS-3, as well as Jammu and Kashmir since the latter state's districts cannot be matched unequivocally with districts in the NFHS. Third, in accordance with IGMSY's eligibility rules, we restrict the analysis to children of mothers aged at least 19 years at the time of birth of the child. Fourth, to make the cohorts exposed to the program as comparable as possible to the older cohorts not exposed to IGMSY, we focus on children born in the four years around IGMSY's onset. Hence, we restrict the children's sample to the birth cohorts 2010, 2011, 2012 and 2013. Finally, we exclude observations with a missing value among the control variables.[12] We do not restrict our sample to the officially eligible first and second-born children due to the possibility of imperfect compliance with this rule. The dataset for our empirical analyses contains 13,367

[10] These states (programs) are Madya Pradesh (Mukhyamantri Mazdoor Suraksha Yojana), Maharashtra (Matrutva Anudan Yojana), Odisha (Mamata) and Tamil Nadu (Dr. Muthulakshmi Maternity Assistance Scheme).

[11] Chhandigarh, Dadra and Nagar Haveli, Daman and Diu, Lakshadweep, Puducherry, Andaman and Nicobar Islands.

[12] Detailed observation numbers documenting how we arrive at our estimation sample contains Table A3 of the appendix. Not excluding observations with missing controls yields very similar estimates.

children in 70 districts (35 program and control districts each) of 24 states, which host around 70 percent of India's population.

2.4.2 Outcome variables

The outcomes of interest we consider can be partitioned into two broad categories. First, health inputs which are directly incentivized by conditions for obtaining cash transfers. These comprise self-reported use of public health care, breastfeeding practices as well as completion of basic infant vaccinations. The second set of outcomes are health markers of children and mothers as well as birth-spacing. Health markers of children include low birthweight reported by the mother, as well as anthropometric and hemoglobin measurements taken at the survey interview, and mortality. For maternal health we consider a mother's body mass index (BMI) and hemoglobin level at the time of interview.

As we have laid out in Section 2.2, a central objective of IGMSY/PMMVY is to improve pregnant and lactating women's use of health services. The program's implementation rules mandate that the corresponding directly incentivized activities are provided by Anganwadi centers. Therefore, we include all survey questions in the NFHS children's dataset which literally contain the word "Anganwadi" and are administered either for all children aged five and younger or all women aged 15 to 49. These are first, the incidence of benefits from Anganwadi centers, essentially health services and products received by the mother during pregnancy or lactation, which we take as measures of antenatal and postnatal health care respectively; second, the incidence of Anganwadi benefits among children as well as contacts of their mothers with Anganwadi staff during the months preceding the survey interview, which we take as measures of the program's medium-term effects on access to government health care.[13] For breastfeeding, we use two indicators: initiation of breastfeeding within one hour of birth (not directly incentivized by the program) and duration of exclusive breastfeeding for a minimum of six months (a program condition). Regarding immunization, we include all three child vaccinations which IGMSY/PMMVY directly incentivizes: BCG, DPT-3 and polio-3. In addition, we include measles, the fourth vaccination required for complete vaccination status by the country's own national definition (Ministry of Health and Family Welfare, 2019), which coincides with the WHO definition of full vaccination among one year-olds (WHO 2021). BCG is typically given right after birth, three shots of DPT (DPT-3) and polio (polio-3) within the first four months, while a single dose of measles vaccine should be administered after nine

[13] A referee of this paper has pointed out that IGMSY households might misrepresent health services use to receive benefits. We believe this to be a minor threat for the following reasons. First, the NFHS is administered independently from the public health care infrastructure and not connected administratively to the IGMSY program. Second, IGMSY does not appear on the NFHS questionnaire – unlike JSY. Third, for the cases included in our research design, survey interviews were conducted no earlier than two years after the expiration of IGMSY benefits.

months at the earliest (Ministry of Health and Family Welfare, 2019). Since all these vaccinations should be completed by the age of two at the latest, our sample, in which children are on average three years old, is well suited to assess complete immunization status. As the measles immunization is beyond the program's scope, which ends six months postpartum, we view any effects on measles immunization status as an indirect (or spillover) effect of the program. Our principal focus will be on the composite measure *complete vaccination status*, first, because of its immediate policy interest and, second, because it summarizes four individual effects in a meaningful way while avoiding statistical complications of multiple inference (Christensen and Miguel 2018).

We define child health outcomes following WHO standards. According to these, a child exhibits low birth weight if the birth weight falls short of 2,500 grams (WHO 2014). A child is underweight at the time of the survey if the weight-for-age z-score (WAZ), the number of standard deviations from the WHO reference population's median, is smaller than -2 (WHO 2018). Similarly, a child is stunted if the height-for-age z-score (HAZ) is smaller than -2. We code a child as anemic if the hemoglobin level falls short of 11 g/dl (WHO 2011). We focus on these binary indicators because they capture deprivations, whose overcoming is the ultimate policy goal. We measure mortality through a variable indicating whether a child is no longer alive at the time of interview.

For mother's health, we follow the standard WHO (2011) thresholds. We code a mother as anemic if the hemoglobin falls short of 12g/dl (below 11 g/dl for pregnant women). We define a mother as underweight if her BMI falls short of 18.5. Finally, we measure birth spacing as the number of months between a child's birth and the succeeding birth.

According to the summary statistics set out in Tables A2.5 and A2.6 of the appendix, the sample means for the districts in our research design are close to the all-India averages of the outcome variables (Dhirar et al. 2018). While the NFHS includes a number of measures of activities incentivized by IGMSY/PMMVY in addition to the ones just discussed, we choose to ignore them in our analysis because they are recorded only for the last birth of each woman interviewed.

2.4.3 Balancing test and challenges to internal validity

For an unbiased estimate of the program effect, children in our control and pilot districts should be similar with respect to outcomes and other observable characteristics at baseline (Kahn-Lang and Lang 2020). We assert this in a balancing test set out in Table 2.1. There are no significant differences in outcomes and other observable characteristics between children in pilot and control districts before the start of the program except for the percentage of children with low birth weight, which is significant at the 10% level, and the percentage of children in rural areas. We account for this small difference by including rural/urban location as a control variable in our analysis and conducting subsample analyses for various outcomes (Table A 2.27-A2.30).

Table 2.1: Balancing test

	Control districts		Pilot districts			
	Mean	SD	Mean	SD	Pr |T|>t difference in means	Obs
Outome variables						
Median birth interval first-second born	36.00	13.41	36.00	13.97	1.00	1608
Mother underweight (%)	16.49	37.11	16.81	37.40	0.90	7413
Mother anemic (%)	55.14	49.74	56.81	49.54	0.60	7383
AWC benefits during pregnancy (%)	52.39	49.95	56.71	49.56	0.60	4666
AWC benefits during breastfeeding (%)	48.04	49.97	52.29	49.96	0.62	4657
AWC benefits in last 12 months (child) (%)	49.22	50.01	55.63	49.69	0.33	4482
Mother saw AWW in last 3 months (%)	40.27	49.06	41.48	49.28	0.84	4669
BCG vaccinated (%)	90.29	29.61	91.60	27.74	0.54	4482
DPT-3 vaccinated (%)	79.62	40.29	80.58	39.57	0.83	4482
Polio-3 vaccinated (%)	67.50	46.85	69.52	46.04	0.71	4482
Measles vaccinated (%)	83.83	36.83	85.14	35.57	0.70	4482
Complete vaccination (%)	59.11	49.17	61.21	48.74	0.74	4482
Breastfed within one hour of birth (%)	72.90	44.46	68.26	46.56	0.28	4346
Breastfed exclusively for min. 6 months (%)	73.25	44.28	70.42	45.66	0.54	2766
Low birth weight (%)	9.96	29.95	14.32	35.04	0.06	4482
Underweight (%)	36.64	48.19	39.45	48.89	0.44	4190
Stunted (%)	38.89	48.76	37.96	48.54	0.87	4190
Anemic (%)	47.49	49.95	47.10	49.93	0.92	4203
Mortality (%)	3.75	19.00	3.98	19.54	0.84	4669
Other characteristics						
Child age	3.72	0.45	3.72	0.45	0.96	4482
Female (%)	45.58	49.82	47.09	49.93	0.45	4669
Mother's height (cm)	151.19	6.24	151.44	5.89	0.88	4669
Mother's age	28.96	4.69	28.89	4.74	0.70	4669
Mother formally educated (%)	65.67	47.49	65.50	47.55	0.98	4669
SC/ST (%)	32.33	46.78	35.73	47.93	0.48	4669
Hindu (%)	81.25	39.04	81.64	38.72	0.94	4669
Wealth index quantile						
First (%)	26.24	44.00	28.05	44.93	0.84	4669
Second (%)	22.85	42.00	19.82	39.87	0.40	4669
Third (%)	17.97	38.40	19.33	39.50	0.66	4669
Fourth (%)	17.99	38.42	18.26	38.64	0.95	4669
Fifth (%)	14.94	35.66	14.54	35.26	0.94	4669
Rural (%)	70.89	45.43	76.78	42.23	0.00	4669

Notes: The p-values are corrected for clustering at the district level. Sample: children born in 2010 and 2011 to mothers aged at least 19 at birth in the 70 districts of our research design. AWC (AWW) abbreviates Anganwadi Center (Worker).

Another potential threat to our identification strategy is a selection bias introduced by families moving from control to pilot districts in order to profit from the program. However, we find no effect of IGMSY on years of residence in the current location (see Table A2.8). Similarly, our estimates might be biased if fertility was increased in order to select *into* the program, and not as a result of receiving the program. While De and Timilsina (2020) argue that the uncertainty of government programs in India renders this quite unlikely, they propose to test for fertility selection by testing for program effects on fertility. Applying this test to our data yields a small negative, insignificant effect, which confirms that the IGMSY program has not significantly increased birthrates in treatment districts (Table A2.9).

2.5 Results

2.5.1 Main results

We first turn to the program's impacts on directly incentivized health inputs, health facility use and immunization. The results for the former are set out in Table 2.2. While all estimates have the expected positive sign, by far the greatest effect occurs for the outcome with the shortest recall period, the probability that a mother has met an Anganwadi worker during the three months preceding the survey interview, which increases by 13 percent. This result is significant at the 1% level even when adjusting for multiple inference.

Table 2.3 sets out the ITT effects on immunization. The coefficients all have the expected positive sign, with larger effects for immunizations where vaccination rates at baseline are low. Column 5 shows that children eligible for the program are on average 9 percent more likely to be fully immunized. Despite this significant overall improvement and a significant point estimate for polio-3 vaccination, we find no statistically significant effects for the individual vaccinations (columns 1 - 4) when we account for multiple inference.

Table 2.4 reports the results for breastfeeding. The coefficient for breastfeeding within one hour of birth is close to zero, which is not surprising given that this indicator is not directly incentivized by the program. On the other hand, the coefficient for breastfeeding for at least six months, which is a program condition, is sizable and equals about four percent of the control mean. It suffers, however, from a standard error which is twice as large as in the estimations in Tables 2.2 and 2.3. This is due to a relatively high non-response rate, of about 30 percent, regarding the corresponding survey question.

We now turn to the question whether IGMSY and the above-documented improvements in health inputs also improve health outcomes. Table 2.5 provides estimates of the program's effects on child health markers and mortality. According to column 1, the program significantly decreases the incidence of low birth weight by four percentage points, which equals 40 percent

Table 2.2: Program effect on Anganwadi center contact

	Prenatal benefits mother	Postnatal benefits mother	Mother met AWW last 3 months	Child benefits in last 12 months
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	1.85	0.97	5.56***	0.91
	(1.26)	(1.35)	(1.08)	(1.27)
	{0.31}	{0.71}	{0.00***}	{0.71}
Control mean (percent)	52.4	48.0	40.3	49.2
Observations (children)	13886	13854	13897	13308
Clusters (districts)	70	70	70	70

Notes: The dependent variables are dummies multiplied by 100 indicating that the mother received Anganwadi/ICDS center benefits or services during pregnancy (1) or lactation (2), that the child received such benefits or services during the 12 months preceding the interview (4) and that the mother met an Anganwadi worker (AWW), Accredited social health worker (ASHA) or other community health worker during the 3 months preceding the interview (3). Linear probability models. *Birth year in 2012 or later* is a dummy variable equal to one if the a child is born in 2012 or later. *Program district* is a dummy variable equal to one if a child lives in an IGMSY pilot district. Additional controls included but not reported in the table are birth order of the child and birth year dummies; mother's educational level, age, squared age and height; household's religious affiliation (6 categories) wealth index quintiles, SC/ST and rural/urban status; district fixed effects and cohort-specific state dummies. The sample are children alive at time of interview, born during 2010-2013 to mothers aged at least 19 at the time of birth. Robust standard errors clustered at the district level in parentheses. Romano-Wolf bootstrapped q-values (p-values adjusted for multiple inference in columns (1)-(4)) in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

Table 2.3: Program effect on child immunization

	Vaccinations by type				Complete vacc.
	BCG	DPT-3	Polio-3	Measles	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	0.93	1.53	2.58**	1.11	5.31***
	(1.15)	(1.68)	(1.28)	(1.38)	(1.56)
	{0.60}	{0.60}	{0.12}	{0.60}	
Control mean (percent)	90.3	79.6	67.5	83.8	59.1
Observations (children)	13308	13308	13308	13308	13308
Clusters (districts)	70	70	70	70	70

Notes: The dependent variables are dummies multiplied by 100 indicating BCG, DPT-3, Polio-3 and measles vaccinations, and complete vaccination. All other notes from Table 2.2 apply.

Table 2.4: Program effect on breastfeeding

	Breastfeeding	
	Within one hour of birth	For at least six months
	(1)	(2)
Program district x birth year 2012 or later	0.57	3.32
	(1.31)	(2.67)
	{0.68}	{0.37}
Control mean (percent)	72.9	73.2
Observations (children)	13014	9488
Clusters (districts)	70	70

Notes: The dependent variables are dummies multiplied by 100 indicating that the child was breastfed within one hour of birth (1) and exclusive breastfeeding length was at least six months (2). The smaller number of observations in column 2 arises from missing values, which are due to a "don't know" response option not available to respondents for the dependent variable in column 1 and consistency checks performed by NFHS. All other notes from Table 2.2 apply.

Table 2.5: Program effect on child health outcomes and mortality

	Health outcomes				Mortality
	Low birth weight	Underweight	Stunting	Anemia	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	-4.03***	-4.63***	0.43	3.90**	-1.02*
	(1.17)	(1.66)	(2.10)	(1.90)	(0.56)
	{0.01**}	{0.02**}	{0.83}	{0.12}	{0.13}
Control mean (percent)	10.0	36.6	38.9	47.5	3.7
Observations (children)	13308	12390	12390	12523	13897
Clusters (districts)	70	70	70	70	70

Notes: The dependent variables are dummies multiplied by 100 indicating low birth weight as recalled by the mother; underweight, stunting and anemia at the survey interview; and whether the child has died before the survey interview. All livebirths form the sample in columns 1 and 5. All other notes from Table 2.2 apply.

of the control group's mean. This effect is significant at the five percent level when accounting for multiple inference. Following Kahn-Lang and Lang (2020), we take this result with some caution as the balancing test in section 2.4.3 shows a borderline significant positive difference between pilot and control districts of almost three percentage points. The program also reduces underweight at the time of interview, when children are one to five years old, by 4.6 percentage points (column 2). As for low birthweight, this result is significant at the 5 percent level when accounting for multiple inference. According to column 5, IGMSY also has a beneficial and relatively large effect on child mortality. On the other hand, the point estimates for stunting and anemia are positive (columns 3-4). However, neither of these outcomes attains statistical significance when accounting for multiple inference.

Table 2.6 sets out the program effects on two health markers of mothers, underweight and anemia at the time of the survey interview. While underweight decreases by ten percent according to column 1, the estimated effect is not statistically different from zero due to the relatively large standard error.

Table 2.6: Program effect on mothers' health outcomes

	Underweight	Anemia
	(1)	(2)
Program district x eligible mother	-1.65	-0.21
	(1.19)	(1.48)
	{0.29}	{0.89}
Control mean (percent)	16.55	56.03
Observations (children)	14881	14807
Clusters (districts)	70	70

Notes: The dependent variables are dummy variables multiplied by 100 indicating that the child's mother is underweight (BMI below 18.5) or anemic (hemoglobin level below 12 g/dl, 11 g/dl for pregnant women) at the time of interview. *Eligible mother* equals one if the mother gave birth to a first- or second-born child in 2012 or later. Program district is a dummy variable equal to one if a mother lives in an IGMSY pilot district. Additional controls included but not reported in the table are mother's birth year dummies, mother's educational level (four categories) and marital status; household's religious affiliation, wealth index quintiles, SC/ST and rural/urban status; district fixed effects and child-cohort-specific state dummies. The sample are mothers of a first or second child born between 2010 and 2013 aged at least 19 at the time of birth. Other notes from Table 2.2 apply.

Table 2.7: Program effect on birth spacing

Sample split by:		Birth order		
	Full sample	First born	Second born	Third or higher
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	0.89**	0.83***	1.01	1.10
	(0.05)	(0.06)	(0.10)	(0.11)
		{0.02**}	{0.93}	{0.63}
Control median birth interval (months)	59	35	censored	censored
Observations (children)	13897	4880	4502	4515
Clusters (districts)	70	70	70	70

Notes: Cox proportional hazards models with separate baseline hazards by sex (columns (1)-(4)) and birth order (columns (1) and (4)). The table reports birth hazard ratios (one indicates no effect). Significance stars are for a test of the hypothesis that the coefficient equals one. All other notes from Table 2.2 apply.

Table 2.7 contains results regarding the program's impact on birth spacing. For first-born children, column 2 documents a 17 percent decrease in the hazard rate for the occurrence of the second birth, which roughly corresponds to a stretching of the median expected birth interval between first and second births by 5.2 months. While there is also a significant 11 percent decrease in the birth hazard rate in the whole sample (an increase by 7.3 months of the median interval), the estimated effects for birth intervals subsequent to the second and higher-order births are not significant.

2.5.2 Internal validity and robustness

To assess whether the key assumption underlying our matched-pair difference-in-difference approach - parallel trends in outcomes absent IGMSY - is internally valid, we conduct a parallel trend placebo test (Kahn-Lang and Lang 2020). To this end, we repeat our main analysis with data from India's third District Level Household Survey, DLHS-3 (IIPS 2010a), which was fielded in 2007 and 2008, well before the onset of IGMSY. We choose this survey for the following reasons. Using only the years 2010 and 2011 in the NFHS-4 sample leads to a grossly unbalanced sample, with no observations for 10 states in 2010. The preceding demographic and health survey NFHS-3, fielded in 2005 and 2006, contains only state but no district identifiers. On the other hand, the latest DLHS, the DLHS-4 from 2012, does not cover nine major Indian states which feature prominently in our main estimation sample. The complementary AHS (Annual Health Survey), which covers those nine states, does not allow the combination of the household, woman and child modules. We therefore select the DLHS-3. While its household sample is twenty percent larger than in NFHS-4, it contains no information on child anthropometrics and no comparable information about Anganwadi service use, and the remaining variables are recorded for children only up to three years of age instead of five as in NFHS-4. More precisely, DLHS-3 contains only the birth cohorts 2004, 2005 and 2006, while the cohorts 2010-2016 are featured in NFHS-4. In parallel to our main estimations, where the oldest two cohorts featured by the survey constitute the ineligible children in treated districts, we use the oldest two cohorts in DLHS-3 for this purpose – which leaves us with only the 2006 cohort as placebo-eligible children. Descriptive statistics for the placebo sample are set out in Table A2.7 in the appendix and the estimation results in Table 2.8. Consistent with the relatively large sample size in DLHS-3 and a higher fertility rate in the 2000s compared to the 2010s, there are around 25 percent more observations in the placebo than in our main sample. Among the 12 estimates there is a single one, for DPT-3, which is negative and borderline significant. However, the p-value equals merely 0.22 when accounting for multiple inference. We conclude that our research design passes muster in terms of this placebo test.

As discussed in section 2.3, some children born between June and December 2011 may have profited early from the program, leading to an underestimation of the program effect. To test whether the definition of treated cohorts influences our results, we exclude children that could have potentially benefited from the program, those born June to December 2011, from the control cohort. While this reduces contamination of the control group, it renders the treatment and control group in our research design slightly less comparable, e.g. regarding age. As expected, the magnitude of the point estimates in Tables A2.10 to A2.15 overall increases for

Table 2.8: Placebo test

	Birth spacing				Vaccinations by type				Complete vacc.	Breastfeeding		Mortality
	Full sample	Birth order			BCG	DPT-3	Polio-3	Measles		Within	For	
		First	Second	Third or higher						1 hour of birth	min. 6 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Program district x birth	0.94	0.93	0.94	1.11	0.15	-2.59*	-1.39	-0.13	-0.42	-0.01	0.01	-0.05
year 2005 or later	(0.06)	(0.08)	(0.09)	(0.09)	(1.17)	(1.47)	(1.33)	(1.26)	(1.44)	0.01	0.01	(0.58)
		{0.40}	{0.50}	{0.20}	{0.99}	{0.22}	{0.59}	{0.99}		{0.83}	{0.63}	
Control median (months)	55	39	censored	censored								
Control mean (percent)					80.7	62.9	57.7	73.3	50.2	0.50	0.70	3.4
Observations	27564	8484	7843	11237	17990	17990	17990	17990	17990	18128	17668	18676
Clusters	70	70	70	70	70	70	70	70	70	70	70	70

Notes: The sample are children born between 2004 and 2006 to mothers aged at least 19 at birth. Notes from tables 7, 3, 4 and 5 apply to the respective columns.

most outcomes. In particular, IGMSY's effect on child mortality now becomes significant at the one percent level even when accounting for multiple inference. On the other hand, the coefficient for underweight drops to almost half of its previous value and becomes insignificant at conventional levels. Due to this sensitivity, we judge our previous finding regarding underweight as less robust.

Finally, we conduct two robustness checks to assess whether our results are sensitive to the functional form of the estimating equation and the inclusion of control variables. First, we estimate equation (2.1) without control variables and, second, we employ a logit instead of a linear probability model.[14] The results are set out in Tables A2.16 to A2.21 and A2.22 to A2.26, respectively. They are all very similar to the ones reported in Tables 2.2 through 2.7.

2.5.3 Magnitude of effects and cost-effectiveness

Regarding immunization, both directly incentivized and not directly incentivized vaccinations increase. However, all estimated effects regarding individual vaccinations are insignificant when adjusting for multiple inference, while the full vaccination rate increases by a sizable nine percent. From the available evidence, the following two factors are mainly responsible for this. First, more frequent contacts with the health system orchestrate the vaccinations: children who completed some, but not all recommended vaccinations now complete the vaccination schedule in full instead of only partly. Second, the low polio vaccination rates at baseline of 67.5 percent demonstrate that polio immunizations constituted a bottleneck which the program has successfully mitigated – IGMSY's effect on polio-3 of 2.8 percentage points is by far the largest among the four individual immunizations. Overall, given a recommended 95 percent coverage goal and the vaccination rates in our sample before the onset of IGMSY, the program has the potential to close India's immunization gap by a moderate seven percent.

Perhaps our most remarkable finding is a positive side effect: a 14 percent increase in interactions with local public health centers three to five years subsequent to delivery, which suggests that a perinatal CCT can be long-term effective regarding access to maternal and child health care. Taken together, our results regarding Anganwadi interactions and immunization suggest that health service use significantly expanded through IGMSY, consistent with findings reported in the meta-analysis of Ranganathan and Lagarde (2012) for other geographical contexts. Interestingly, the size of the effects found on vaccination and health service use correspond roughly to the magnitude of effects found for the JSY program, where vaccination and health service use were not directly incentivized (Powell-Jackson et al. 2015; Glick). While we cannot isolate the channels by which these improvements occur, they

---

[14] Kahn-Lang and Lang (2020) show that a double-difference logit model can be justified if it captures the correct functional form of the data-generating process.

are consistent with an interaction of increased health-services demand triggered by the program conditions and expanded health service supply. Further research yet has to determine whether these effects of IGMSY are additional to those of JSY or whether they crowd out the positive side effects of JSY.

Considering our results for child health markers, we cannot conclude that IGMSY has had transformative impacts. On the one hand, while there are some indications of improvements in children's weight-related outcomes, the respective estimates suffer from a lack of robustness. And we find no improvements for two other important markers recorded three to five years after program benefits have elapsed, anemia and stunting. These patterns are consistent with Rivera et al.'s (2004) evaluation of Mexico's Progresa program as well as a meta-study of CCTs in other country contexts by Manley et al. (2013). In our view, this is little surprising given the rather small improvements in health inputs during infancy of no more than five percentage points.

Regarding maternal health, the relatively large beneficial but statistically insignificant effect on mothers' underweight is consistent with our theory of change. A limitation of our analysis in this context, however, is that any moderately-sized instantaneous effects of a perinatal CCT program on maternal health may be too short-lived to be detectable at the time of the survey interview, three to five years after program participation.

According to our theory of change, a maternal CCT's effect on birth spacing is ambiguous. While the fertility rate remains unchanged, we document a sizable decrease in birth hazard rates for second-born children. This finding is in accordance with Ghosh and Kochar's (2018) evaluation of IGMSY in two districts of Bihar. In our view, the most likely explanation for these results is the restriction of benefits to only the first two children, and, for birth spacing, increased interaction of mothers with Anganwadi workers, whose mandate it is to also conduct family-planning counseling. One indication for this is that our findings contrast the fertility increase found for the JSY program, which does not incentivize such trainings or interactions with Anganwadi workers and is available for all birth orders (Powell-Jackson et al. 2015). Taken together, these findings suggest that improved access to counseling combined with aligned indirect incentives is effective for increasing birth intervals and perhaps fertility in the longer run.

While a full-fledged cost-benefit analysis is beyond the scope of this paper, we present some simple back-of-the-envelope calculations of the program's cost-effectiveness with respect to two outcomes, immunization and underweight. We put these estimates into perspective by comparing them to the cost-effectiveness of a vaccination intervention in India (Banerjee et al. 2010) and another CCT program, the Nicaraguan *Red de Protecciòn Social*, evaluated by Maluccio and Flores (2005). The estimated cost per additional fully immunized child in

IGMSY/PMMVY amounts to ₹ 49,906 (659.88 US$).[15] These costs are excessive compared to Banerjee et al. (2010), who estimate the costs for non-cash incentives and recruiting per fully immunized child in their study at ₹ 2,011. The discrepancy is not surprising, given that a policy tailored to improve a particular indicator should be more cost-effective than a cash transfer, which can be spent in a number of ways and may improve a range of indicators, albeit each one only to a limited extent (Banerjee et al. 2019). Moreover, average treatment effects as estimated by Banerjee et al. (2010) usually surpass intent-to-treat effects such as ours. A rough estimate based on Maluccio and Flores' (2005) findings regarding the Nicaraguan CCT arrives at $ 6,161.29 per child lifted out of underweight. This estimate is three times larger than our estimated cost of ₹ 71,571.49 ($ 2,282.47) per child prevented from being underweight by IGMSY, suggesting a favorable performance of India's program regarding this outcome.

2.5.4 A closer look at immunization

In our exploration of differences in program effects for different population sub-groups, we estimate heterogeneous program effects by child and household characteristics as well as the intensity of program implementation. The p-values of the differences in the estimated treatment effects across subsamples are adjusted for multiple inference with Benjamini and Hochberg's method (Benjamini and Hochberg 1995). Tables A2.27 to A2.30 in the appendix contain IGMSY's effects on child outcomes for various sample splits and the differences between the corresponding subsample. Here, we focus on one of the two outcomes for which we find the most robust effects, full immunization (Table A2.27).

While the coefficients between wealthy and poorer households differ only slightly, the coefficients for children of the first two birth orders, girls and children living in SC/ST and rural households are much larger in magnitude than those for the respective complementary group. The same applies to states with a high program intensity, which we define by an above-median (Rs 1,578) expenditure per eligible case in 2012 and 2013 (see appendix A2.1 for details). In line with expectations, these results provide suggestive evidence that disadvantaged groups and directly targeted children benefit disproportionately from the program. The latter finding is also consistent with the effects on birth spacing between first and second-born children documented above. However, the associated difference between the two sub-groups is significant at conventional levels only for child's sex. For considerations of space, we do not discuss heterogeneous effects for interactions with health-center staff here, which are very similar.

---

[15] See appendix A2.1 for the cost-effectiveness calculation.

2.6 Conclusions

In this paper we have investigated the effects of the five-year long pilot phase of India's thus far largest conditional cash transfer program. Overall, we find that the design of IGMSY effectively improved several health inputs such as access to formal health care, vaccination, and birth spacing. However, these changes were not always achieved through direct conditions but in some cases through indirect side effects of related conditions or program features. For instance, we find important effects on birth intervals at low parities, in particular after a woman's first birth. In contrast to Powell-Jackson et al. (2015), who document fertility increases as a consequence of a CCT for institutional deliveries, the fertility effects of IGMSY support the program's broader objectives, to improve maternal and child health. These seemingly contradictory findings highlight the importance of implementation details of a CCT: while IGMSY directives explicitly mention extended birth spacing as an element of monetarily incentivized educational sessions, JSY's guidelines do not. Moreover, as the rule awarding the cash transfer for each livebirth in JSY was seen as the main incentive for decreasing birth intervals, IGMSY restricted eligibility to only the first two live births. Notwithstanding this, while our findings on health outcomes suggest improvements in children's weight-related anthropometric markers as well as mortality, the limited geographical scope of the program's pilot phase puts limitations on the precision of these estimates.

The most convincing explanation for the only small to moderate magnitude of the effects of this large program is its failure to reach out to eligible populations at large and selection into the program by mothers who would have sought ante- and postnatal care anyway. While we cannot rigorously trace such a pattern with our data, Falcao et al. (2015) document that IGMSY has reached no more than 50 percent of target populations during its early years. While these administrative figures are certainly upper bounds for the program's actual outreach, they equal precisely the order of magnitude of key outcome variables prior to the program's onset as well as the fraction of women in our sample reporting to hold a bank account – which is a prerequisite for receiving IGMSY cash transfers. Hence, in an extreme case, there could be no effects at all if only mothers enrolled in the program who would have sought government healthcare in the absence of the incentive. Such a pattern is more than likely, given that such women face the lowest opportunity costs of enrolling and meeting the program conditions. This highlights the importance of a health-related CCT's outreach at the intensive margin: while we expect the expansion of IGMSY/PMMVY to all districts of India in 2017 to yield moderate improvements similar to the ones reported here, expanding the outreach within program districts well above 50 percent, e.g. by more active recruitment and by lowering entry hurdles such as the bank-account requirement, promises much greater improvements regarding both health inputs and outcomes. Of course, such an expansion would inflate the program's price tag, which will likely face political pushbacks.

Which lessons can be learned from our results regarding the IGMSY pilot for its successor PMMVY, which has been universal since 2017? The only major difference between the two is PMMVY's restriction of eligibility to a woman's first birth, while IGMSY has covered the first two parities. While our data does not allow us to quantify spillover effects to younger, ineligible siblings, there exists evidence from two studies of parent-focused medical care and counseling interventions covering exclusively first-born infants in vulnerable US households. These document spillover effects to later-born siblings of a magnitude similar to the direct effect on eligible children (Ruggiero et al. 2020; Seitz and Apfel 1994). This evidence suggests that the effect of PMMVY on mothers' health-system contacts, birth spacing and immunization among children of all parities will be similar to the effects we have documented for IGMSY. On the other hand, PMMVY's budget allocation, which allows to fund no more than roughly half of all first births in India, makes effects larger than the ones reported here unlikely, and the main lessons learned from IGMSY apply to PMMVY.

Overall, our impression of India's maternal conditional cash transfer program IGMSY/PMMVY is more optimistic than Jackson-Powell et al.'s (2015) assessment of JSY. Our evidence suggests that Indian policy makers have learned from the experiences with this earlier program and designed IGMSY/PMMVY more carefully. According to our findings, maternal CCTs can be an important, albeit not transformative element for improving access to health care and health outcomes in low- and middle-income countries. However, to increase their effectiveness, the problem of outreach within the target groups deserves more attention.

A.2 Appendix

A2.1 Further methodology

Empirical approach for mothers' outcomes

Whether a mother is eligible is measured via a dummy variable that takes on one if the mother gave birth to a first or second born child after the year 2011.

$$H_{sdti} = \alpha_{sd} + \mu_t + (\delta_s \times elig\_w_t) + \beta(P_{sd} \times elig\_w_t) + \gamma X_{sdit} + u_{sdit} \qquad (2.2)$$

where subscript i indicates mother, d district, s state, and t birth year of mother

H: health outcome (underweight, anemia)

$\boldsymbol{\beta}$: ITT effect

P: dummy for pilot program district

$elig\_w_t$: eligibility of the mother

X_{sdit}: control variables

α_{sd} : district fixed effects

δ_s : state fixed effects

μ_t: mother's birth year fixed effects

u_{sdit}: robust standard errors clustered at the district level

Sample: mothers aged at least 19 who had a first or second child between 2010 and 2013 (in order to ensure a relatively even sample split and comparability)

Measure of program implementation

In order to detect heterogeneous program effects by intensity of implementation, we use each state's disbursements per potentially eligible woman between 2011 and 2014 under IGMSY. For state-wise program expenditures we draw on data from Falcao et al. (2015). We define the number of eligible cases in each state by calculating the state-wise share of first and second births of mothers aged at least nineteen years old in pilot districts (years 2012 and 2013) in the overall number of births for the same period using NFHS – 4 data and then multiplying it with the Indian population (in thousands) and the Indian birthrate (sourced from World Bank (2019)). The share of eligible children in all births in pilot districts is roughly 65 percent. A state is considered a high implementation state if it has an above-median (₹ 1,578) expenditure per eligible case in 2012 and 2013. We choose this measure instead of the state-wise number of beneficiaries reported by the government since our measure is more strongly correlated with survey measures of program coverage from Niti Aayog and DMEO (2017). Moreover, it corresponds more closely to the ITT effects we are estimating.

Cost effectiveness calculation

The cost effectiveness of the intervention was estimated as follows:

$$\frac{\textit{Average program expenditure per eligible case } 2012 - \ 2013/14}{\textit{Program effect on first and second born children } (\textit{share})}$$

For a description of how we arrive at the yearly expenditure per eligible case, see A2.
Maluccio and Flores (2005) find a 6.2 percentage point reduction in underweight (ITT-effect) and their reported average value of received direct cash transfer and health related in-kind benefits per year amount in total to US$ 382 per child. Note that this cost estimate does not include administrative costs for running the program, as does our estimated cost for IGMSY.

A2.2 Additional tables

Table A2.1: Timing of conditions and cash transfer disbursement in IGMSY

Install-ment	Timing of disbursement	Amount	Conditions
			2011-2013
1	At the end of six months of pregnancy	1500	Pregnancy registered within four months at the Anganwadi Center (AWC) or Health Center Mother participated in min. one antenatal check-up Mother picked up IFA tablets Mother received at least one tetanus vaccination Mother attended a nutrition and health counseling at least once
2	At the end of three months after delivery	1500	Child birth is registered Child has received Polio 0 and BCG vaccination Child has received Polio-1 and DPT-1 vaccination Child has received Polio-2 and DPT-2 vaccination Child has been weighed at least twice since birth After delivery, mother participated in at least two nfant and young child feeding (IYCF) counseling meetings
3	At the end of six months after delivery	1000	Child has been exclusively breastfed for first six months, unless advised otherwise by a medical doctor

			After six months, the child has been started to be fed complementary foods Child has received Polio-3 and DPT-3 vaccination Child has been weighed at least twice between three and six months Between three and six months after birth the mother participated in at least two infant and young child feeding (IYCF) counseling meetings
			With increase of transfer amount in 2013
1	At the end of six months of pregnancy	3000	Pregnancy registered Mother participated in at least two antenatal care visits where she received iron and folic acid tablets and tetanus vaccination
2	At the end of six months after delivery	3000	Child birth is registered Child is immunized against BCG, Polio 1-3 and DPT 1-3 In the first three months after delivery, mother participates in at least three IYCF meetings and had the child's growth measured at least three times Mother exclusively breastfeeds for six months, afterwards child is introduced to complimentary food

Notes*:* Sources are Ministry of Women and Child Development (2011) and Niti Aayog and DMEO (2017).

Table A2.2: Ranking of index scores of pilot and matched control districts

District	State	Maternity Index	Performance Group	Pilot / Control	District	State	Maternity Index	Performance Group	Pilot / Control
Katihar	Bihar	139.2	low	C	Dehradun	Uttarakhand	348.2	middle	P
Saharsa	Bihar	139.3	low	P	Purbi Singhbhum	Jharkhand	349.2	middle	P
Godda	Jharkhand	165.2	low	C	West District	Sikkim	352.7	middle	P
Simdega	Jharkhand	165.9	low	P	South	Delhi	356.8	middle	C
Tamenglong	Manipur	179.7	low	P	North West	Delhi	359.5	middle	P
Vaishali	Bihar	191.8	low	P	Dhamtari	Chhattisgarh	359.6	middle	P
Saran	Bihar	192.6	low	C	South District	Sikkim	362.7	middle	C
Muzaffarnagar	Uttar Pradesh	192.9	low	C	Rewari	Haryana	364.8	middle	C
Mahoba	Uttar Pradesh	194.1	low	P	Panchkula	Haryana	368.3	middle	P
Sultanpur	Uttar Pradesh	203.6	low	P	Patan	Gujarat	372.4	middle	P
Azamgarh	Uttar Pradesh	207.3	low	C	Fatehgarh Sahib	Punjab	373.7	middle	C
East Garo Hills	Meghalaya	210.8	low	P	Kapurthala	Punjab	374.0	middle	P
West Garo Hills	Meghalaya	234.8	low	C	Valsad	Gujarat	374.2	middle	C
Banswara	Rajasthan	246.1	low	C	Kamrup	Assam	376.8	middle	P
Ukhrul	Manipur	248.9	low	C	Dibrugarh	Assam	379.5	middle	C
Bilaspur	Chhattisgarh	251.0	low	C	Kheda	Gujarat	379.7	middle	C
Bhilwara	Rajasthan	251.7	low	P	West	Delhi	382.6	middle	P
Udaipur	Rajasthan	252.7	low	P	Bharuch	Gujarat	382.7	middle	P
Dhalai	Tripura	252.8	low	P	East	Delhi	383.6	middle	C
Bastar	Chhattisgarh	257.5	middle	P	Nalgonda	Telangana	391.2	middle	P
Tonk	Rajasthan	257.7	middle	C	Y.S.R.	Andhra Pradesh	392.4	middle	C
Dhemaji	Assam	263.7	middle	C	Davanagere	Karnataka	403.4	high	C
Goalpara	Assam	266.5	middle	P	Bilaspur	Himachal Pradesh	413.7	high	C
Ranchi	Jharkhand	276.1	middle	C	Dharwad	Karnataka	421.0	high	P
North Tripura	Tripura	280.2	middle	C	Muktsar	Punjab	421.6	high	C
Changlang	Arunachal Pradesh	306.7	middle	C	Amritsar	Punjab	426.5	high	P
Lawngtlai	Mizoram	307.8	middle	P	Kolar	Karnataka	428.0	high	P
Chamoli	Uttarakhand	313.4	middle	C	Rangareddy	Telangana	433.6	high	C
Puruliya	West Bengal	320.5	middle	C	West Godavari	Andhra Pradesh	435.8	high	P
Jalpaiguri	West Bengal	328.5	middle	P	Tumkur	Karnataka	442.2	high	C
Durg	Chhattisgarh	329.0	middle	C	Hamirpur	Himachal Pradesh	470.1	high	P
Bankura	West Bengal	343.5	middle	P	Palakkad	Kerala	488.1	high	P
Mamit	Mizoram	345.3	middle	C	Kozhikode	Kerala	501.1	high	C
Dakshin Dinajpur	West Bengal	346.2	middle	C	North Goa	Goa	502.3	high	P
Papumpare	Arunachal Pradesh	347.2	middle	P	South Goa	Goa	517.8	high	C

Notes: Index calculated from District Level Household Survey 2007-08. Components: (i) % literate female population (age 7 +), (ii) % mothers registered their pregnancy in the 1st trimester , (iii) % mothers who had at least 3 antenatal care visits during their last pregnancy, (iv) % institutional births, (v) % children (12-23 months) fully immunized (BCG, 3 DPTs, Polio and measles), and (vi) % children breastfed within one hour of birth. Pilot/Control indicates pilot and control districts.

Table A2.3: Data restrictions and number of observations

Restriction	*N children*
Original dataset (NFHS-4, children schedule)	259,627
Excluding UTs	252,064
Restricted to program and control districts	32,741
Excluding states with other maternity programs	26,573
Excluding Jammu and Kashmir	25,170
Restricted to children of mothers who were at least 19 at the birth of the respective child	23,762
Restricted to children born between 2010 and 2013	14,721
Restricted to observations for which data is available for at least one main outcome	14,721
Restricted to observations for which data is available for all controls	13,897

Table A2.4: Description of variables

Unit of observation	**Variable**	**Description**	**Source**
		Outcome variables	
Child	Anemia	Dummy variable, equals one if the child has mild, severe or moderate anemia (hemoglobin level below 11 g/dl)	Generated from NFHS-4
Child	Birth spacing	Succeeding birth interval in months (if birth interval missing but a birth took place after the respective child: months since last birth of mother)	Generated from NFHS-4
Child	Breastfed within one hour of birth	Dummy variable, equals one if the child has been breast-fed within one hour of birth	NFHS-4
Child	Breastfed for at least six months	Dummy variable, equals one if the child has been exclusively breast-fed for at least six months (the child is at least six months old).	NFHS-4
Child	BCG	Dummy variable, equals one if a child has been administered the Bacillus-Calmette-Guèrin-vaccination. Following standard DHS procedure, “Don’t know” is recoded to “No” for all vaccinations	NFHS-4
Child	Complete vaccination	Dummy variable, equals one if the child has been administered one BCG, one measles, three DPT and three polio doses	

Child	DPT-3	Dummy variable, equals one if child has been administered the last combined diphtheria, pertussis and tetanus vaccination dose	NFHS-4
Child	Low birth weight	Dummy variable, equals one if weight at birth lies below 2.5 kg	Calculated from NFHS-4
Child	Measles	Dummy variable, equals one if the child has been administered one measles dose	NFHS-4
Child	Mortality	Dummy variable, equals one if the child has perished	Calculated from NFHS-4
Child	Polio-3	Dummy variable, equals one if child has been administered the last polio vaccination dose	
Child	Stunted	Dummy variable, equals one if the height for age z-score (using the WHO reference population) (HAZ) lies below -2. The HAZ is equal to the number of standard deviations below or above the reference median and calculated as follows: (observed height/age) – (median height/age of the reference population) / standard deviation of the reference population	NFHS-4
Child	Underweight	Dummy variable, equals one if weight for age z-score (using the WHO reference population) (WAZ) below -2	NFHS-4
Mother	Anemia	Dummy variable, equals one if the mother has mild, severe or moderate anemia (hemoglobin level below 11 g/dl for pregnant women and below 12 g/dl for all other adult women)	Generated from NFHS-4
Mother	Underweight	Dummy variable, equals one if Body Mass Index lies below 18.5	NFHS-4
Control variables and variables employed for heterogeneous effects estimation			
Child	Birth order	Birth order of the child	NFHS-4
Child	Sex	Dummy variable, equals one if the child is female	NFHS-4
Mother	Age	Age in years	NFHS-4
Mother	Educational level	Woman's highest educational level. Consists of the following categories: no education, primary education, secondary education, higher education	NFHS-4
Mother	Height	Height in cm	NFHS-4
Mother	Marital status	Marital status of the mother, consists of the following categories: never in union, married, widowed/separated	Generated from NFHS-4

Mother	Squared age	Squared age in years	Calculated from NFHS-4
Household	Poor	Dummy variable, equals one if a household belongs to the poorest 40% in the NFHS-4 sample in terms of the wealth index	NFHS-4
Household	Rural	Dummy variable, equals one if the place of residence lies in a rural area	NFHS-4
Household	Religion	Religion of the household. Consists of the following categories: Hindu, Muslim, Christian, Sikh, Buddhist, no or other religion	Generated fom NFHS-4
Household	SC/ST	Dummy variable, equals one if household belongs to a scheduled caste or tribe	Generated from NFHS-4
Household	Wealth	Quintiles of a continuous measure of relative wealth of a household equal to the factor score of an index of owned assets (range of index: -2.25822 to 2.86687)	NFHS-4
State	State implementation	Dummy variable, equals one for states with average IGMSY expenditure between 2011-2014 per eligible case above the median (eligible are first and second born children of mothers at least 19 at birth in pilot districts)	Generated from expenditure (Falcao et al. (2015)) and population data (World Bank (2019)), NFHS-4
		Treatment and eligibility variables	
Child	Birth year 2012 or later (Elig_birth)	Dummy variable, equals one if child was born in 2012 or later	Generated from NFHS-4
District	Program district	Dummy variable, equals one for districts in which IGMSY was implemented in 2011. The variable equals zero if the district is a control district (district which is nearest neighbor in terms of the maternity and child health index score used for selection of pilot districts, in the same state)	Pilot districts (Ministry of Women and Child Development 2019), control districts matched by authors

Table A2.5: Summary statistics NFHS-4 (Children, full sample)

	Mean	SD	Obs
Outcome variables (child)			
Succeeding birth interval (median)	33.00	12.42	13897
Breastfeeding			
Within one hour of birth (%)	70.70	45.51	13014
For at least six months (%)	78.17	41.31	9488
Anganwadi contact			
AWC benefits during pregnancy (%)	56.82	49.53	13886
AWC benefits during breastfeeding (%)	51.73	49.97	13854
AWC benefits in last 12 months (child) (%)	56.32	49.60	13308
Mother saw AWW in last 3 months (%)	44.44	49.69	13897
Vaccination			
BCG vaccinated (%)	91.30	28.18	13308
DPT-3 vaccinated (%)	81.25	39.03	13308
Polio-3 vaccinated (%)	72.29	44.76	13308
Measles vaccinated (%)	84.60	36.10	13308
Complete vaccination (%)	63.76	48.07	13308
Health			
Low birth weight (%)	13.60	34.28	13308
Underweight (%)	37.95	48.53	12390
Stunted (%)	40.80	49.15	12390
Anemic (%)	55.68	49.68	12523
Mortality (%)	4.41	20.53	13897
Child characteristics			
Female (%)	47.54	49.94	13897
Age (months)	40.17	11.61	13125
Mother characteristics			
Height (cm)	151.40	6.13	13897
Age (years)	28.05	4.79	13897
Education			
None (%)	32.96	47.01	13897
Primary (%)	14.28	34.99	13897
Secondary (%)	42.24	49.40	13897
Higher (%)	10.52	30.68	13897

Table A2.5, continued

Household characteristics			
Rural (%)	72.65	44.57	13897
SC/ST (%)	34.35	47.49	13897
Religion			
Hindu (%)	80.40	39.70	13897
Muslim (%)	13.91	34.61	13897
Christian (%)	2.78	16.43	13897
Sikh (%)	2.05	14.18	13897
Buddhist (%)	0.22	4.72	13897
None or other religion (%)	0.63	7.92	13897
Wealth quintile			
First (%)	26.51	44.14	13897
Second (%)	20.96	40.71	13897
Third (%)	18.74	39.03	13897
Fourth (%)	18.05	38.46	13897
Fifth (%)	15.74	36.42	13897
Eligibility variable			
Born 2012 or later (%)	65	48	13897

Notes: AWC benefit: received goods or services from Anganwadi center. DPT-3: child completed the third diphteria, pertussis and tetanus vaccination. BCG: child completed the Bacillus Calmette-Guèrin vaccination (primarily employed against tuberculosis). Polio-3: child completed the third polio vaccination. Wealth index quintiles included in NFHS-4 (derived from the factor score of principal component analysis of a household asset index). SC/ST indicates whether the child's household belongs to a Scheduled Caste or Tribe. Summary statistics are based on data from NFHS-4 and constructed using state mother/child sampling weights provided by NFHS-4. Sample: Children born 2010-2013 to mothers aged at least 19 at birth of child, 70 districts.

Table A2.6: Summary statistics NFHS-4 (Mothers, full sample)

	Mean	SD	Obs
Outcome variables (mother)			
Underweight (%)	20.23	40.18	14881
Anemia (%)	57.89	49.37	14807
Mother characteristics			
Age (years)	28.64	6.13	14807
Education			
None (%)	23.76	42.56	14807
Primary (%)	13.13	33.77	14807
Secondary (%)	48.69	49.98	14807
Higher (%)	14.43	35.14	14807
Marital status			
Married (%)	95.66	20.37	14807
Never in union (%)	0.11	3.36	14807
Widowed or separated (%)	4.23	20.12	14807
Household characteristics			
Rural (%)	65.41	47.57	14807
SC/ST (%)	32.04	46.66	14807
Religion			
Hindu (%)	82.80	37.74	14807
Muslim (%)	10.42	30.55	14807
Christian (%)	2.82	16.57	14807
Sikh (%)	2.82	16.55	14807
Buddhist (%)	0.29	5.38	14807
None or other (%)	0.85	9.19	14807
Wealth quintile			
First (%)	18.12	38.52	14807
Second (%)	18.22	38.60	14807
Third (%)	19.25	39.43	14807
Fourth (%)	21.48	41.07	14807
Fifth (%)	22.93	42.04	14807
Eligibility variable			
Gave birth to first or second child in 2012 or later (%)	52	50	14807

Notes: Underweight: BMI below 18.5. Anemic: hemoglobin level below 12 g/dl, below 11 g/dl for pregnant women. Wealth index quintiles included in NFHS-4 (derived from the factor score of principal component analysis of a household asset index). SC/ST indicates whether the child's household belongs to a Scheduled Caste or Tribe. Summary statistics are based on data from NFHS-4 and constructed using state mother/child sampling weights provided by NFHS-4. Sample: mothers aged at least 19 with a first or second child born between 2010 and 2013, 70 districts.

Table A2.7: Summary statistics DLHS-3

	Mean	SD	Obs
Outcome variables (child)			
Birth Interval (median, months)	24.00	10.41	6140
Vaccination			
BCG vaccinated (%)	81.59	38.76	19641
DPT-3 vaccinated (%)	63.81	48.06	19641
Polio-3 vaccinated (%)	58.66	49.25	19641
Measles vaccinated (%)	74.06	43.83	19641
Complete vaccination (%)	51.25	49.99	19641
Breastfeeding			
Within one hour of birth (%)	0.47	0.50	19807
For min. 6 months (%)	0.74	0.44	19291
Health			
Mortality (%)	3.79	19.09	20432
Child characteristics			
Female (%)	47.44	49.94	20432
Mother characteristics			
Age (years)	26.88	5.15	20432
Education			
No education (%)	39.74	48.94	20432
Primary education (%)	31.97	46.64	20432
Secondary education (%)	15.07	35.78	20432
Higher education (%)	13.22	33.88	20432
Household characteristics			
Rural (%)	70.78	45.48	20432
SC/ST (%)	42.66	49.46	20432
Religion			
Hindu (%)	69.14	46.19	20432
Muslim (%)	11.90	32.38	20432
Christian (%)	11.30	31.66	20432
Sikh (%)	4.02	19.63	20432
Buddhist (%)	1.85	13.48	20432
None or other religion (%)	1.79	13.27	20432
Wealth quintile			
First	24.27	42.87	20432
Second	19.87	39.91	20432
Third	19.40	39.55	20432
Fourth	18.34	38.70	20432
Fifth	18.11	38.51	20432
Eligibility variable			
Born 2005 or later (%)	72	45	20432

Notes: DPT-3: child completed the third diphteria, pertussis and tetanus vaccination. BCG: child completed the Bacillus Calmette-Guèrin vaccination (primarily employed against tuberculosis). Polio-3: child completed the third polio vaccination. Wealth index quintiles derived from the factor score of principal component analysis of the DLHS-3 household asset index. SC/ST indicates whether the child's household belongs to a Scheduled Caste or Tribe. Summary statistics are based on data from DLHS-3 and constructed using state mother/child sampling weights provided by DLHS-3. Sample: Children born 2004-2006 to mothers aged at least 19 at birth of child, 70 districts.

Table A2.8: Program effect on years of residence

	Years of residence
	(1)
Program district x eligible mother	1.002
	(1.134)
Control mean (years)	13.56
Observations	14,917
Clusters	70

Notes: The dependent variable are years of residence. Eligible mother equals one if the mother gave birth to a first or second born child in 2012 or later. Program district is a dummy variable equal to 1 when a woman lives in a district which was IGMSY pilot district. Additional controls included but not reported in the table are educational level, marriage status, religion, wealth index factor score of the household, and whether the household belongs to a Scheduled Caste or Tribe and is situated in a rural area respectively. Sample: mothers aged at least 19 with first or second children born between 2010 and 2013. All estimates are computed using sampling weights, district fixed effects, mother's birth year fixed effects and child-cohort specific state fixed effects. Robust standard errors clustered at the district level in parentheses.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A2.9: Program effect on fertility

	Fertility
	(1)
Program District x 2012 or later	-0.09
	(0.29)
Control mean (years)	8.2
Observations	238896
Clusters	70

Notes: Linear probability models. The dependent variable is a dummy variable equal to one if a woman gives birth in a given calendar year. *2012 or later* is a dummy variable equal to 1 for the years 2012 and 2013, and 0 for the years 2010 and 2011. *Program district* is a dummy variable equal to 1 if a woman lives in an IGMSY pilot district. Coefficients in percentage points. Additional controls included but not reported in the table are educational level, marital status, household religion and wealth index factor score, and dummy variables equal to 1 if the household belongs to a Scheduled Caste or Tribe and is situated in a rural area respectively. Sample: Women at least 19 years or age, 70 districts. All estimates are computed using sampling weights, birth year and district fixed effects as well as cohort-specific state fixed effects. Robust standard errors clustered at the district level in parentheses.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.10: Program effect on Anganwadi service use - Excluding children born June-December 2011

	Prenatal benefits mother	Postnatal benefits mother	Mother met AWW last 3 months	Child benefits in last 12 months
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	1.85	0.97	5.56***	0.91
	(1.26)	(1.35)	(1.08)	(1.27)
	{0.31}	{0.71}	{0.00***}	{0.71}
Control mean (percent)	52.4	48.0	40.3	49.2
Observations	13886	13854	13897	13308
Clusters	70	70	70	70

Notes: See Table 2.2. Sample: children born 2010-2013 (excluding children born June-December 2011) to mothers aged at least 19 at birth of child.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.11: Program effect on vaccination (excluding children born June-December 2011)

	Vaccinations by type				Complete vacc.
	BCG	DPT-3	Polio-3	Measles	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	0.93	1.53	2.58**	1.11	5.31***
	(1.15)	(1.68)	(1.28)	(1.38)	(1.56)
	{0.60}	{0.60}	{0.12}	{0.60}	
Control mean (percent)	90.3	79.6	67.5	83.8	59.1
Observations	13308	13308	13308	13308	13308
Clusters	70	70	70	70	70

Notes: See Table 2.3. Sample: children born 2010-2013 (excluding children born June-December 2011) to mothers aged at least 19 at birth of child.
 * p<0.10 ** p<0.05 *** p<0.01

Table A2.12: Program effect on breastfeeding (excluding children born June-December 2011)

	Breastfeeding	
	Within one hour of birth	For at least six months
	(1)	(2)
Program district x birth year 2012 or later	-0.04**	-0.01
	(0.02)	(0.02)
	{0.08*}	{0.78}
Control mean (percent)	0.3	0.9
Observations	10686	8023
Clusters	70	70

Notes: See Table 2.4. Sample: children born 2010-2013 (excluding children born June-December 2011) to mothers aged at least 19 at birth of child.
 * p<0.10 ** p<0.05 *** p<0.01

Table A2.13: Program effect on child health and mortality (excluding children born June-December 2011)

	Health outcomes				Mortality
	Low birth weight	Underweight	Stunting	Anemia	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	-4.78**	-2.82	2.46	3.81	-2.05***
	(1.90)	(2.07)	(3.39)	(2.46)	(0.63)
	{0.05*}	{0.29}	{0.50}	{0.29}	{0.01**}
Control mean (percent)	10.0	38.6	41.1	46.8	3.0
Observations	10920	10146	10146	10281	11407
Clusters	70	70	70	70	70

Notes: See Table 2.5. Sample: children born 2010-2013 (excluding children born June-December 2011) to mothers aged at least 19 at birth of child.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.14 Program effect on mothers' health (excluding children born June-December 2011)

	Underweight	Anemia
	(1)	(2)
Program district x eligible mother	-1.90	-0.10
	(1.28)	(1.57)
	{0.25}	{0.95}
Control mean (percent)	17	56
Observations	14489	14414
Clusters	70	70

Notes: See Table 2.6. Sample: mothers aged at least 19 at birth of a first or second child born between 2010 and 2013 in the 70 districts of our research design (excluding mothers with first and (if applicable) second child born June-December 2011).

* p<0.10 ** p<0.05 *** p<0.01

Table A2.15 Program effect on birth spacing (excluding children born June-December 2011)

Sample split by:		Birth order		
	Full sample	First born	Second born	Third or higher
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	0.89**	0.86	0.79*	1.08
	(0.05)	(0.08)	(0.10)	(0.15)
		{0.24}	{0.16}	{0.59}
Control median birth interval (months)	42.2	36.0	44.6	46.2
Observations	11407	3987	3666	3754
Clusters	70	70	70	70

Notes: See Table 2.7. Sample: children born 2010-2013 (excluding children born June-December 2011) to mothers aged at least 19 at birth of child.
* p<0.10 ** p<0.05 *** p<0.01

Table A2.16: Program effect on Anganwadi service use (without additional controls)

	Prenatal benefits mother	Postnatal benefits mother	Mother met AWW last 3 months	Child benefits in last 12 months
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	1.69	0.76	5.25***	0.82
	(1.37)	(1.44)	(1.10)	(1.30)
	{0.44}	{0.76}	{0.00***}	{0.76}
Control mean (percent)	52.4	48.0	40.3	49.2
Observations	13886	13854	13897	13308
Clusters	70	70	70	70

Notes: See Table 2.2. Without additional control variables.
* p<0.10 ** p<0.05 *** p<0.01

Table A2.17: Program effect on vaccination (without additional controls)

	Vaccinations by type				Complete vacc.
	BCG	DPT-3	Polio-3	Measles	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	0.90	1.51	2.49*	1.08	5.22***
	(1.08)	(1.57)	(1.33)	(1.24)	(1.57)
	{0.58}	{0.58}	{0.17}	{0.58}	
Control mean (percent)	90.3	79.6	67.5	83.8	59.1
Observations	13308	13308	13308	13308	13308
Clusters	70	70	70	70	70

Notes: See Table 2.3. Without additional control variables.
* p<0.10 ** p<0.05 *** p<0.01

Table A2.18: Program effect on breast-feeding (without additional controls)

	Breastfeeding	
	Within one hour of birth	For at least six months
	(1)	(2)
Program district x birth year 2012 or later	0.44	3.18
	(1.30)	(2.74)
	{0.74}	{0.42}
Control mean (percent)	72.9	73.2
Observations	13014	9488
Clusters	70	70

Notes: See Table 2.4. Without additional control variables.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.19: Program effect on child health and mortality (without additional controls)

	Health outcomes				Mortality
	Low birth weight	Underweight	Stunting	Anemia	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	-3.96***	-4.86***	0.17	3.61*	-1.04*
	(1.20)	(1.62)	(2.06)	(1.89)	(0.53)
	{0.01**}	{0.01**}	{0.93}	{0.14}	{0.14}
Control mean (percent)	10.0	36.6	38.9	47.5	3.7
Observations	13308	12390	12390	12523	13897
Clusters	70	70	70	70	70

Notes: See Table 2.5. Without additional control variables.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.20: Program effect on mothers' health (without additional controls)

	Underweight	Anemia
	(1)	(2)
Program district x eligible mother	-1.70	-0.23
	(1.19)	(1.51)
	{0.27}	{0.87}
Control mean (percent)	17	56
Observations	14881	14807
Clusters	70	70

Notes: See Table 2.6. Without additional control variables.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.21: Program effect on birth spacing (without additional controls)

Sample split by:		Birth order		
	Full sample	First born	Second born	Third or higher
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	0.97	0.82***	1.00	1.08
	(0.04)	(0.05)	(0.09)	(0.11)
		{0.01***}	{0.97}	{0.85}
Control median birth interval (months)	40.6	35.8	42.3	43.7
Observations	13897	4880	4502	4515
Clusters	70	70	70	70

Notes: See Table 2.7. Without additional control variables.
* p<0.10 ** p<0.05 *** p<0.01

Table A2.22: Program effect on Anganwadi service use (Logit model)

	Prenatal benefits mother	Postnatal benefits mother	Mother met AWW last 3 months	Child benefits in last 12 months
	(1)	(2)	(3)	(4)
Program district x birth year 2012 or later	1.95*	1.46	4.50***	0.91
	(1.12)	(1.27)	(1.02)	(1.37)
Control mean (percent)	13886	13854	13897	13308
Observations	70	70	70	70
Clusters	13886	13854	13897	13308

Notes: See Table 2.2. Average marginal effects of Logit model. Coefficients multiplied by 100 for comparability with linear probability model (LPM). Differences in observations to LPM arise from some instances of perfect prediction. Computation of multiple inference adjusted p-values in logit model not possible due to large number of fixed effects.
* p<0.10 ** p<0.05 *** p<0.01

Table A2.23: Program effect on vaccinations (Logit model)

	Vaccinations by Type				Complete Vacc.
	BCG	DPT-3	Polio-3	Measles	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	0.73	1.09	2.73**	0.96	5.22***
	(1.22)	(1.55)	(1.23)	(1.35)	(1.52)
Control mean (percent)	90.3	79.6	67.5	83.8	59.1
Observations	12825	13308	13308	13308	13308
Clusters	66	70	70	70	70

Notes: See Table 2.3. Average marginal effects of Logit model. Coefficients multiplied by 100 for comparability with linear probability model (LPM). Differences in observations to LPM arise from some instances of perfect prediction. Computation of multiple inference adjusted p-values in logit model not possible due to large number of fixed effects.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A2.24: Program effect on breastfeeding (Logit model)

	Breastfeeding	
	Within one hour of birth	For at least six months
	(1)	(2)
Program district x birth year 2012 or later	-0.66	0.25
	(1.56)	(1.91)
Control mean (percent)	72.9	73.2
Observations	13014	9522
Clusters	70	70

Notes: See Table 2.4. Average marginal effects of Logit model. Coefficients multiplied by 100 for comparability with linear probability model (LPM). Differences in observations to LPM arise from some instances of perfect prediction. Computation of multiple inference adjusted p-values in logit model not possible due to large number of fixed effects.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A2.25: Program effect on child health and mortality (Logit model)

	Health outcomes				Mortality
	Low birth	Underweight	Stunting	Anemia	
	(1)	(2)	(3)	(4)	(5)
Program district x birth year 2012 or later	-4.45***	-4.52***	0.64	4.06**	-0.88
	(1.13)	(1.66)	(2.12)	(1.88)	(0.66)
Control mean (percent)	10.0	36.6	38.9	47.5	3.7
Observations	13308	12390	12390	12523	13799
Clusters	70	70	70	70	69

Notes: See Table 2.5. Average marginal effects of Logit model. Coefficients multiplied by 100 for comparability with linear probability model (LPM). Differences in observations to LPM arise from some instances of perfect prediction. Computation of multiple inference adjusted p-values in logit model not possible due to large number of fixed effects.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.26: Program effect on mothers' health (Logit model)

	Underweight	Anemia
	(1)	(2)
Program district x eligible mother	-1.91	-0.25
	(1.30)	(1.48)
Control mean (percent)	16.73	56.03
Observations	14772	14807
Clusters	69	70

Notes: See Table 2.6. Average marginal effects of Logit model. Coefficients multiplied by 100 for comparability with linear probability model (LPM). Differences in observations to LPM arise from some instances of perfect prediction. Computation of multiple inference adjusted p-values in logit model not possible due to large number of fixed effects.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.27: Heterogeneous program effects on complete vaccination

Sample split by:	Individual and household characteristics										Intensity of program implementation (by state)	
	Birth order		Sex		Wealth		Social group		Residence			
	First/second	Third/higher	Female	Male	Bottom 40%	Upper 60%	SC/ST	Other	Rural	Urban	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Program district x birth	6.64***	1.84	7.98***	2.47	4.05**	5.51**	8.13**	4.02**	6.00***	-0.97	2.30	5.61***
year 2012 or later	(1.91)	(3.12)	(2.02)	(2.31)	(1.82)	(2.44)	(3.22)	(1.94)	(1.77)	(4.14)	(3.45)	(2.01)
Difference	4.73		5.49		-1.46		3.95		6.92		3.24	
	[0.21]		[0.07]		[0.62]		[0.33]		[0.13]		[0.42]	
	{0.63}		{0.35}		{0.63}		{0.63}		{0.51}			
Control mean (Percent)	63.2	53.0	59.3	60.8	54.3	65.4	61.0	59.6	59.0	63.1	63.3	61.0
Observations	9016	4292	6476	6832	6536	6772	5888	7420	10257	3051	6515	6510
Clusters	70	70	70	70	69	70	70	68	67	70	34	34

Notes: Linear probability model. The dependent variable is complete vaccination. *Birth year in 2012 or later* is a dummy variable equal to 1 when a child is born in 2012 or later. *Program district* is a dummy variable equal to 1 when a child lives in a district which was IGMSY pilot district. Additional controls included but not reported in the table are birth order of the child, mother's educational level, age, squared age and height, household religion and wealth index factor score, and dummy variables equal to 1 if the household belongs to a Scheduled Caste or Tribe and is situated in a rural area respectively. Subsamples: Wealth is based on the asset index included in the survey. Low (high) intensity of program implementation is based on total expenditures 2012 and 2013 per eligible case (first and second births of mothers aged at least nineteen years old in pilot districts) below (equal to or above) the median. Sample: children alive at time of interview, born 2010-2013 to mothers aged at least 19 at birth of child, 70 districts. All estimates are computed using sampling weights, birth year and district fixed effects as well as cohort-specific state fixed effects. Robust standard errors clustered at the district level in parentheses; p-values for differences between two complementary groups are in square brackets; Benjamini-Hochberg q-values (p-values for subgroup differences adjusted for multiple inference across the five differences reported for columns 1 through 10 are in curly brackets).

* p<0.10 ** p<0.05 *** p<0.01

Table A2.28: Heterogeneous program effects on Anganwadi service use

Sample split by:	Individual and household characteristics										Intensity of program implementation (by state)	
	Birth order		Sex		Wealth		Social group		Residence			
	First/second	Third/higher	Female	Male	Bottom 40%	Upper 60%	SC/ST	Other	Rural	Urban	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Program district x birth year 2012 or later	6.00***	2.39	6.22***	5.01**	4.33**	5.99***	5.40***	5.26***	4.98***	-1.02	7.72***	3.90**
	(1.67)	(2.77)	(2.02)	(2.08)	(1.88)	(2.00)	(1.86)	(1.78)	(1.51)	(4.26)	(1.82)	(1.48)
Difference	3.85		1.21		-1.66		0.29		6.06		-3.80	
	[0.31]		[0.73]		[0.60]		[0.92]		[0.21]		[0.10]	
	{0.92}		{0.92}		{0.92}		{0.92}		{0.92}			
Control mean (percent)	41.1	40.3	44.8	37.4	47.4	34.6	44.8	38.8	45.8	27.0	44.8	42.3
Observations	9382	4515	6734	7163	6891	7006	6173	7724	10753	3144	6812	6799
Clusters	70	70	70	70	69	70	70	68	67	70	34	34

Notes: See Table A2.27. Outcome is a dummy variable equal to 1 if the mother met an Anganwadi worker in the three months before the survey interview.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.29: Heterogeneous program effects on low birth weight

Sample split by:	Individual and household characteristics										Intensity of program implementation (by state)	
	Birth order		Sex		Wealth		Social group		Residence			
	First/second	Third/higher	Female	Male	Bottom 40%	Upper 60%	SC/ST	Other	Rural	Urban	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Program district x birth year 2012 or later	-5.91***	0.21	-3.17**	-5.03**	-1.03	-6.35***	-0.94	-5.07***	-3.69***	-5.26	-5.68**	-4.06**
	(1.54)	(1.83)	(1.48)	(2.17)	(1.22)	(1.79)	(1.85)	(1.47)	(1.10)	(3.61)	(2.15)	(1.56)
Difference	-5.86		1.85		5.33		3.94		1.42		1.53	
	[0.02]		[0.52]		[0.02]		[0.12]		[0.68]		[0.56]	
	{0.06*}		{0.68}		{0.06*}		{0.35}		{0.68}			
Control mean (percent)	13.8	7.9	12.0	12.0	9.1	14.6	11.5	12.3	10.9	15.0	10.7	12.8
Observations	9016	4292	6476	6832	6536	6772	5888	7420	10257	3051	6515	6510
Clusters	70	70	70	70	69	70	70	68	67	70	34	34

Notes: See Table A2.27. Outcome is a dummy variable equal to 1 if a child is born with a weight of less than 2.5 kg.

* p<0.10 ** p<0.05 *** p<0.01

Table A2.30: Heterogeneous program effects on underweight

Sample split by:	Individual and household characteristics										Intensity of program implementation (by state)	
	Birth order		Sex		Wealth		Social group		Residence			
	First/second	Third/higher	Female	Male	Bottom 40%	Upper 60%	SC/ST	Other	Rural	Urban	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Program district x birth	-5.49***	-2.70	-2.76	-5.95**	-7.41**	-2.25	-9.05***	-1.91	-4.82**	-5.55	-7.00*	-2.68
year 2012 or later	(1.68)	(3.22)	(2.95)	(2.35)	(3.00)	(2.30)	(3.02)	(1.80)	(2.11)	(4.07)	(3.45)	(2.12)
Difference	-3.01		3.28		-5.15		-6.89		0.68		4.16	
	[0.35]		[0.42]		[0.16]		[0.04]		[0.89]		[0.29]	
	{0.84}		{0.84}		{0.65}		{0.21}		{0.89}			
Control mean (percent)	35.1	44.2	41.7	34.7	47.2	29.3	41.4	36.2	40.7	30.4	34.6	41.6
Observations	8381	4009	6013	6377	6096	6294	5473	6917	9566	2824	6100	6029
Clusters	70	70	70	70	69	70	70	68	67	70	34	34

Notes: See Table A2.27. Outcome is a dummy variable equal to 1 if a child is underweight (WAZ below -2).

* p<0.10 ** p<0.05 *** p<0.01

3 External social audits: the value of experience

Combining external and community monitoring for improved public program delivery

Abstract

Public programs provide important services and goods which the most vulnerable segments of society often particularly rely on. These benefits can be curtailed by corruption and mismanagement in public programs. Social audits, a form of community monitoring involving the cross-checking and public discussion of information on program performance, may help to detect inefficiencies and corruption and thus improve service delivery. The predominant social audit design, in which beneficiaries collect the information for verification themselves, has been criticized as costly, time-consuming and for insufficient audit quality. As a potential solution, 'external' social audits have been put forward, in which information is assembled by external auditors. This paper uses a panel data fixed effects methodology to evaluate the effects of an external social audit on public program outcomes such as infrastructure asset creation and service provision and use in a large public work program, the Indian National Employment Guarantee Act. I find that additional external social audits decrease inefficiencies and corruption in the creation of infrastructure assets. Moreover, this effect is likely sustainable, as auditors in the long run successfully adapt the audit process to new ways to hide corruption. However, external social audits have no or even negative effects on labor expenditure and service provision and use, measured by employment and enrollment in the program. Furthermore, I argue that external social audits are a cost-effective way to reduce embezzlement.

3.1 Introduction

Public programs provide important services and public goods which are often particularly relied on by the most vulnerable segments of society (Besley and Ghatak 2006). These benefits of public programs can be curtailed by mismanagement and corruption: both are associated with loss of funds and efficiency losses in public provision, but also with poorer quality or quantity of public services provided, or exclusion of some targeted population groups (Chêne 2014; Andrews et al. 2006; Olken 2006, 2007, 2009; Ferraz et al. 2012). For these reasons, various measures have been suggested to reduce corruption and improve the quality of public program implementation through community monitoring tools (Molina et al. 2016). One such tool are social audits, a type of community monitoring which allows "citizens who receive a service to review and cross-check information reported by the service provider against information collected from users of the service" (Ringold et al. 2011). The collected information is shared and discussed publicly in the presence of all stakeholders, including beneficiaries, service providers and politicians (Grandvoinnet et al. 2015; Molina et al. 2016). Through the generation and dissemination of new information concerning beneficiary rights and program implementation, reduction of transaction costs by bringing all stakeholders together, empowerment of beneficiaries, public accountability and pressure, social audits are seen as a means to curb corruption and decrease inefficiencies in program implementation and thus improve public service delivery (Ringold et al. 2011; Mansuri and Rao 2013; Molina et al. 2016).

In practice, social audits differ in the way the information for verification is collected. In the predominant design, here referred to as *user-conducted* social audit, beneficiaries collect the information for verification themselves, and thus act as auditors. In contrast, an alternative approach, here called *external* social audit, uses independent external auditors in lieu of beneficiaries to gather the relevant information. The first, user-conducted social audit approach has been criticized for several reasons: two major concerns raised are the large fluctuation and low education of beneficiaries (Tambe et al. 2016a). These result in high and repeated selection and training costs (Tambe et al. 2016a). Moreover, the lack of experience and education impair social audit quality (Tambe et al. 2016a). A third concern is that beneficiaries from disadvantaged backgrounds face repercussions or are excluded from the audit processes by the local elite (Olken 2007; Lakha 2011; Afridi and Iversen 2014). Finally, beneficiaries have an incentive to freeride on the monitoring effort of others if monitoring yields only low individual benefits (Olken 2006, 2007; Khwaja 2004; Besley and Ghatak 2006). Intimidation, exclusion, and freeriding may all impact audit quality negatively.

This study evaluates the alternative 'external' social audit design, which has been suggested as a solution in response to the above concerns (Tambe et al. 2016a), but whose effects on the provision and use of public goods and services remain an open question. The external social audit's permanently employed external auditors are envisaged to conduct audits of better quality due to their independence, accumulated experience, and education. They also have to be selected and trained only once, resulting in lower audit costs (Tambe et al. 2016a). The continued verification of audit results by beneficiaries in the public meeting prevents collusion between auditors and implementers. In this way, external social audits leverage the advantages of external auditors while maintaining the core strengths of social audits: insider information from beneficiaries and public accountability. This renders them potentially particularly useful for public programs which aim to improve both outcomes in which beneficiaries have an information deficit compared to external auditors and high incentive to free-ride (public goods), as well as outcomes with individual benefits that beneficiaries are highly knowledgeable about (e.g. individual goods or services). However, the external approach might also be accompanied by reduced information flow between beneficiaries and social audit personnel due to lower trust towards outsiders, which may also lower public meeting attendance (Varghese et al. 2019).

Despite their potential, evidence on the impact of external social audits on public program outcomes is virtually absent. There exists an abundant literature on the effects of other community monitoring approaches such as scorecards, grievance redressal and information interventions (Molina et al. (2017) and Waddington et al. (2019) provide systematic reviews), as well as external government audits without beneficiary involvement (Avis et al. 2018; Zamboni and Litschig 2018). However, only five studies evaluate the effects of social audits, four of which focus exclusively on user-conducted social audits[16]. In an unpublished matching analysis, user-conducted social audits increased beneficiaries' perceived adequacy of resource management in Colombian public subcontract projects (Molina 2013). Afridi and Iversen (2014; 2017) evaluate user-conducted social audits in the National Rural Employment Guarantee Act (NREGA) in the Indian state Andhra Pradesh. They show that detected irregularities increased (indicating effective monitoring but not deterrence), while program expenditure and employment increased insignificantly. User-conducted social audits in the Indian National Food Security program diversified nutrition and had a positive, and over time intensifying, impact on program awareness and civic norms. Moreover, long-term citizen action improved (Gordon et al. 2019).

Only one study assesses the effects of an external social audit intervention. The randomized controlled trial (RCT), conducted by Olken (2007) in an Indonesian road building program, evaluates three sub-elements of external social audits[17]. External audits of documents and work projects decreased leakage of total program funds, while raising participation in public meetings

[16] Two other studies come relatively close to a user-driven social audit design: first, a study by Alhassan et al. (2019), which evaluates the effects of score cards on health care provision filled in by social village groups. Results are discussed at a meeting open to all stakeholders. The intervention increased usage of maternal and child health services. However, the intervention is not open to all beneficiaries due to its focus on social group members and the distant location of the meetings at the regional capital. Second, a study by Berman et al. (2017) finds positive effects of villagers volunteering as auditors on the quality of public roads in Afghanistan. However, their intervention does not include systematic public hearings open to all beneficiaries and thus lacks an important element of social audits.

[17] In a first experiment, Olken randomly increased the likelihood of receiving a government audit (an inspection of infrastructure and financial documents but no questioning of beneficiaries) from 4 to 100 percent. Audit results were presented at a barely attended public meeting. In a second experiment, Olken raised meeting participation through invitations, coupled in a third intervention with complaint forms.

and gathering information from beneficiaries did not. On closer inspection, participation decreased labor-related but not material-related leakage. Olken's study focuses exclusively on the process indicator leakage of funds and does not examine program outcomes such as public service provision and use. Indeed, the setting is not suitable to evaluate service use, as the roads are freely accessible public goods. In contrast, many other public programs, for example health or employment programs, suffer from access constraints such as exclusion errors, discrimination in the allocation of goods and services or lack of information. For these types of programs, it is important to know whether external social audits can also improve service use. Moreover, Olken evaluates the effects of a one-time intervention, while long-term or average effects of social audits may differ from effects at introduction due to learning of transgressors and auditors (Afridi and Iversen 2014; Gordon et al. 2019).

The present paper aims to close these gaps. In addition to expanding the scarce evidence on the effects of social audits, my contributions are twofold: (i) I evaluate whether external social audits are an effective way to improve public asset and service provision and use in a public employment program, (ii) instead of studying the effects of an isolated intervention, I assess the average effects of repeated external social audits over an extended period of time.

I study external social audits in the context of the Indian National Rural Employment Guarantee Act (NREGA), a public work program which aims to serve as a social security net and improve rural livelihoods through employment and the construction of infrastructure assets (Samuel and Srinivasan 2016). NREGA is both in terms of absolute expenditure and outreach the worldwide largest public work program (UNDP 2015). In the financial year 2013/14, roughly a quarter of rural Indian households and 30 percent of poor rural households worked in NREGA (Desai et al. 2015). I focus on NREGA in the state of Sikkim, the first state in India to officially introduce an external social audit design in NREGA. The effects of this design have so far not been evaluated. I exploit the phase-in of external social audits in Sikkim over four years to estimate the effect of an additional external social audit on the number of detected irregularities, and on NREGA program outcomes in the following year. The analysis of NREGA outcomes, a fixed effects panel data study, is based on a large balanced year-community panel in which I combine information on social audit occurrence with official NREGA data on program costs, employment and project completion status in the years 2011/12 to 2018/19. I complement this by an analysis of irregularities detected during social audits, which I extract from social audit reports for the financial years 2011/12 to 2016/17.

I find that an additional external social audit significantly decreases material and embezzlement irregularities found by the audits. In addition, external social audits have overall beneficial effects on infrastructure-related public provision: additional external social audits on average increase suspended work projects and decrease completed works in the following year significantly (accounting for multiple inference) by 80 and 30 percent respectively. While this partly reflects attempts of bureaucrats to prevent works from being audited, a large part of the reduction in completed works is efficiency enhancing, for instance by promoting the exclusion of inefficient or non-existing works. Moreover, auditors successfully adapt their auditing policy to prevent hiding of irregularities, after which completed works increase and no further works are suspended. Slightly weaker evidence, significant at conventional levels indicates that NREGA expenditure on material increases by 81 rupees (₹) per community inhabitant, while no corresponding improvement occurs in total expenditure, labor expenditure and employment outcomes (enrollment and work days). Notably, the demand for enrollment in the program decreases significantly. While I can rule out that this decrease is due to the removal of fake applications, there remains the possibility that the program looses attraction if collusion between supervisors and workers becomes more difficult. Moreover, I find suggestive evidence that disadvantaged groups reduce their labor supply more strongly. In accordance with a reduction in labor supply, the labor-material ratio declines, albeit not significantly. These results indicate that external social audits are currently more effective at improving technical outcomes than labor outcomes, which require greater informational input and public pressure for action from beneficiaries. While beneficiaries may currently not be sufficiently integrated in the external approach to improve labor-related outcomes, I provide additional, non-causal evidence which suggests that the latter could be ameliorated to a limited extent within the external social audit framework by expanding participation in the public meetings: raising participation in the meetings by one percentage point is associated with a small but significant increase in labor expenditure by ₹20 per inhabitant (a 2.7 percent increase) and an insignificant increase in person days worked, while program enrollment remains unchanged. A notable benefit of the evaluated external social audits is their social cost-effectiveness, which, depending on the time spent by beneficiaries at public meetings, ranges from close to zero to 296 percent of social costs.

My results are relevant to policy-makers, as they show that external social audits are an effective tool to improve provision of asset-related outputs, but do not increase service use. They also suggest that beneficiary participation in public meetings may improve labor expenditure and possibly other outcomes, and that auditing methods have to be constantly adapted to counter adverse learning of bureaucrats.

The rest of the paper is structured as follows. Section 3.2 introduces the NREGA program and social audits in Sikkim and discusses anticipated effects. Section 3.3 describes the data. Section 3.4 sets out the empirical strategy. Results are in section 3.5 and section 3.6 concludes.

3.2 NREGA and Social audits

3.2.1 The National Rural Employment Guarantee Act (NREGA)

The National Rural Employment Guarantee Act (NREGA) is a public work program whose primary goal is to serve as a safety net for India's poor through employment generation. The program guarantees each rural Indian household a maximum of a hundred days of employment in unskilled manual labor at the national minimum wage. The last two conditions, as well as a 33 percent women quota in the work distribution, ensure self-selection of the poor and of disadvantaged groups who might suffer from discrimination in the non-public labor market. The program's restriction to rural areas acknowledges both that a large part of the poor population in India lives in rural areas, as well as the strong seasonal fluctuation of employment and income in rural areas. The households are entitled to an allowance in case of work injuries or death, as well as to an unemployment or travel allowance if NREGA does not provide work promptly and close to the beneficiaries' home. An important secondary goal of NREGA is to improve rural infrastructure and livelihoods through the works constructed by NREGA laborers (Samuel and Srinivasan 2016). However, to ensure that employment generation is prioritized, the guidelines stipulate that at least 60 percent of total (labor and material) expenditure should be spent on labor.

The act was passed in 2005 and rolled out in three phases between 2006 and 2009. 25 percent of the program is funded by the national government and 75 percent by the respective state government. The local communities (Gram Panchayat Units (GPUs)) choose the public work projects, perform registrations for and issue job cards which entitle a household to participate in NREGA, grant employment to the job card holders, and monitor the implementation of the projects. Workers can participate in the selection and prioritization of these projects through the village council (Gram Sabha).

Corruption and inefficiencies in NREGA can occur in numerous ways: embezzlement can occur either by setting up accounts for non-existent ghost workers or ghost projects, by paying existing workers a lower wage or for less working hours, or by billing more hours than actually worked to the project accounts. Particularly difficult to detect are instances where workers and supervisors collude by paying and sharing wages for work that was never executed (Tambe et al. 2016b). Furthermore, corrupt bureaucrats can collude with material suppliers to charge overly high prices, procure excessive quantities or substitute high quality with low quality material and then split the difference in price. Moreover, bureaucrats can ask for bribes in exchange for access to the program (Aiyar and Mehta 2015). Finally, nepotism can occur in the selection of work projects, for instance the creation of private goods for non-poor villagers (Tambe et al. 2016b). In addition to downright corruption, effectiveness and efficiency of the program will be impaired if quality and scope of the program are less than envisaged, for instance if eligible citizens are discouraged from enrolling in the project or denied other rights related to the project (such as pensions or decent working conditions), or if the constructed public assets are of low quality or unusable. The program's social audits are designed to counteract both embezzlement and efficiency irregularities.

3.2.2 Social audits in India and Sikkim

Because of widespread corruption and mismanagement in NREGA (Niehaus and Sukhtankar 2013; Muralidharan et al. 2016), social audits have been conducted in NREGA on a voluntary basis in some states since its implementation and have become mandatory in 2011. The program currently allows for up to 0.5 percent of NREGA administrative funds to be spent on the organization of such social audits. A common design element of social audits in all Indian states is that official documents on program performance are cross-checked against information gained directly from beneficiaries. NREGA officials and beneficiaries are then confronted and heard in a social audit municipality meeting (*Jan Sunwai*). Notwithstanding these similarities and attempts by the central government to harmonize the social audit implementation (MRD and GOI 2011), there remain large variations in the design of audits across states, from more participatory user-conducted to external approaches. In Rajasthan, for instance, social audits were first introduced in the 1990s by the NGO Mazdoor Kisan Shakti Sangathan (MKSS). MKSS is a grassroots organization unaffiliated with the state government whose members plan and carry out user-conducted social audits. The state of Andhra Pradesh/Telangana was the first to institutionalize social audits in 2009 by setting up an independent social audit unit (SAU) within the state government. Andhra Pradesh's SAU plans social audits centrally but partly adopted Rajasthan's approach in that it relies heavily on volunteers in the communities to organize the audits locally and gather information. The state of Sikkim was for many years the only other state with an independent SAU (Tambe et al. 2016a), but implemented an external social audit design, in which external auditors gather information for verification. As of today, the majority of states have set up an SAU, but only few of these are fully independent and functional (Karuna et al. 2019). The long existence and independence of Sikkim's SAU, as well as its unique external social audit design make it the ideal case for evaluating the effectiveness of said design.

Social audits in Sikkim had been conducted from time to time by different untrained NGOs since 2008 but the audits were not standardized, the financial records were not read out publicly and consequently close to no irregularities were found during audits (Tambe et al. 2016a). In 2010, Sikkim invited experts on social audits from the National Institute on Rural Development, and from bodies which had been conducting social audits in Rajasthan and Andhra Pradesh, in order to train pre-existing NGOs in Sikkim to conduct social audits. Following this, in 2011/12, the best-performing NGO was chosen as SAU by the Sikkimese governments' Rural Management and Development Department (Tambe et al. 2016a). This NGO, the Voluntary Health Association of Sikkim (VHAS), initially focused on health and development, but as SAU now took on the additional responsibility to conduct social audits. It has an independent bank account and is not affiliated to the ministry implementing NREGA. Sikkim adopted the concept of a central SAU, as well as most of the process (e.g. structured questionnaires) from Andhra Pradesh, with one important difference: contrary to Andhra Pradesh and Rajasthan, Sikkim implemented a social audit design which does not rely on local volunteer beneficiaries to organize the audits in the communities (GPUs). Instead, social audits are organized in each GPU by permanently employed district level SAU personnel (Subba et al. 2018). Sikkim's decision to pursue the external social audit approach was motivated by observed lower fluctuation and better education of external auditors compared to beneficiary volunteers during the training phase, which decreased audit costs and improved audit quality (Tambe et al. 2016a).

The new, statewide social audit design under the SAU was implemented in a staggered manner: the first four GPUs were audited in the financial year 2011/12, three in 2012/13, 89 in 2013/14, 92 in 2014/15, and from 2015/16 on, all 176 GPUs in Sikkim were audited yearly. During the phase-in, GPUs were randomly selected to be audited in a specific year (personal communication with SAU Sikkim). The role of social auditors is not one of prosecutors but to prepare impartial information for the public hearing (MRD and GOI 2013). This is done according to the following procedure (Karuna et al. 2019): one month before the social audits start, auditors hold a meeting with GPU and block level bureaucrats and worksite supervisors, during which functionaries are given a checklist with all documents they will have to lay open to the auditors. A month later, a team of auditors visits each GPU separately and checks the quality and existence of works and facilities by inspecting them in person, measuring their dimensions and so forth. They also check the completeness and consistency of the official records concerning NREGA (for instance if the actual work dimensions and expenditure correspond to those laid down in the technical estimate). Further, they conduct door-to-door interviews in which they ask beneficiaries about hours worked, wages received, material supplied at the worksite and other complaints and compare the gained information to the entries in the official NREGA records. Their findings are then read out and discussed in a public hearing called *Jan Sunwai* which all beneficiaries are encouraged to attend, in addition to other stakeholders such as the local NREGA officials and Gram Panchayat council members. The fact that audits are performed by a team of auditors and the presence of all other stakeholders at the public hearing ensures checks and balances of all stakeholdes. Because of initially low participation in the Jan Sunwais, it was decided in March 2014 to only hold public hearings if at least 30 percent of the beneficiaries participated (Karuna et al. 2019). The hearing is chaired by a Zilla Panchayat (the next highest government level above GPU) member. At the end of the hearing, actions to be taken are agreed on. These include further investigations, disciplinary action against erring officials, and recovery of embezzled or otherwise improperly used funds. Whether these actions have been carried out is then verified at the end of the social audit process during an 'exit conference' at the district level. Furthermore, the progress achieved in resolving issues is read out at the beginning of the following Jan Sunwai (RMDD et al. 2015). Roughly thirteen percent of missing funds are recovered (Karuna et al. 2019)[18]. Finally, all irregularities detected during the social audit are recorded in social audit reports.

3.2.3 Potential effects of social audits

If external social audits (in the following called 'social audits' for brevity) are effective, I expect some movement in the detected irregularities compared to the first audit - either an *increase* due to higher detection which is not counteracted to the same extent by offenders reducing corrupt or inefficient actions, or a *decrease* if social audits deter corruption (Afridi and Iversen 2014). However, irregularities may also decrease if perpetrators learn to shift corruption to less intensely monitored activities (Olken and Pande 2012).

If social audits have the desired deterring effect on corruption and mismanagement, the effect on program outcomes can still be ambiguous: for instance, with respect to financial program scope, effective social audits would on the one hand prevent unnecessary expenditures and embezzlement, as a consequence of which expenditure would decrease (Olken 2006). However,

[18] This is comparable to a fifteen percent recovery rate of user-conducted audits in Andhra Pradesh (Karuna et al. 2019).

expenditure will increase if social audits raise NREGA employment, and material expenditure can increase independently if audits lead to better quality but expensive material purchases. Employment generation is the primary goal of NREGA. The demand for employment is expected to rise for two reasons: first, disadvantaged population groups are less likely to be excluded through corruption and discrimination when decision-makers have to justify decisions publicly (Kruglanski and Freund 1983). Second, effective social audits will likely raise workers' expectations regarding the probability of receiving their full wages in a timely manner. However, *recorded* NREGA employment may decrease if social audits uncover records of fake workers or exaggerated hours worked. A further goal of NREGA is the creation of infrastructure assets. One indicator for this is the work completion status. While a high number of ongoing works signals a high short-term *capacity* for employment and asset creation, the number of suspended and completed works indicates whether this work effort culminates in the creation of usable assets. For this reason, the program guidelines stipulate that works should be completed within a year. If social audits enforce this, they will decrease the number of ongoing and suspended works and increase the number of completed works. Countervailing effects can materialize if there is a rise in detected ghost works, suspension of inefficient works or if social audits detect works previously labeled wrongly as completed. Ongoing works may also rise if suspended or "completed" works are taken up again after social audits discovered they were not up to standards, or if audits decrease embezzlement of materials needed for construction.

External social audits may have different effects on outcomes related to each of the two main goals of NREGA – employment generation and construction of infrastructure assets. I expect external social audits to perform well for material- and works-related outcomes, as external auditors are able to detect most material- and works-related irregularities without cooperation of beneficiaries (Olken 2006; Afridi and Iversen 2014). However, for some irregularities, auditors depend to some extent on the willingness and ability of beneficiaries to detect, share information on, and pursue certain issues. The issues that beneficiaries are most concerned with and knowledgeable about are typically those related to labor (Aiyar and Mehta 2015; Afridi and Iversen 2014; Khwaja 2004). Thus, while employment outcomes are expected to improve due to the high individual stake of beneficiaries in these outcomes (Afridi and Iversen 2014; Olken 2007), it is also possible that external social audits do not involve beneficiaries sufficiently to detect and pursue labor issues.

3.3 Data

3.3.1 Data sources and description of variables

The analysis makes use of two data sets: first, for the estimation of effects of social audits on the detection of irregularities in the program implementation I construct an unbalanced panel data set for the years 2011/12 to 2016 from 391 social audit reports. From these reports, I extract irregularities detected during each audit in Sikkim in a particular year. Reports from end of 2016 on are excluded for comparability because they differ slightly in format from the earlier reports. Second, for the estimation of effects of social audits on NREGA program outcomes, I create a balanced year-GPU panel by using information on social audit incidence from a list of all social audits conducted between 2010 and 2017 supplied by the Ministry of Rural Development. This is merged with data from the NREGA website on material and labor costs, days worked, job card applications, and number of work projects by completion status (MRD and GOI 2021b). This dataset comprises 170 GPUs and spans the years 2011/12 to 2017/18. Both datasets are supplemented with population data from the census 2011 in order to calculate outcomes per capita. GPUs which split up or merged during the period of analysis between 2010/2011-2017/2018 were excluded[19]. Table A3.1 in the appendix tracks the sample size over the various sample restrictions.

Outcome variables

Irregularities

Irregularities in program implementation are an indicator of active waste of government resources (Zamboni and Litschig 2018). Therefore, I use detected irregularities listed in the audit reports to measure the extent of detected corruption and inefficiencies in the program (Cisneros and Kis-Katos 2022; Zamboni and Litschig 2018). Following Afridi and Iversen (2014), I categorize irregularities broadly into total, labor, and material and work project-related. In addition, I add embezzlement complaints as a direct measure of detected corruption frequency. For comparability, a few issues which were not yet recorded in the earliest social audit reports were excluded. Table A3.2 in the appendix, which lists all main outcomes and explanatory variables in detail, also specifies all individual irregularities included and excluded in the categories above. The total number of irregularities is the sum of irregularities related to labor, works and material. The twenty-two labor-related irregularities encompass access to the program, such as denial, delay or fee charged for registration and job cards, incorrect or fake registrations or job cards, as well as irregularities in the work application, allocation and remuneration. The twenty-four work and material irregularities concern issues with the work project, construction sites and material, and related documents at different implementation stages: firstly, the work planning process, which encompasses work selection and the technical estimate. Second, the work implementation, which includes issues such as non-existent workers or works, irregular prices, quantity and quality of materials, violation of work standards, and false or late work completion certificates. Third, transparency and monitoring of works. Embezzlement forms a separate category which subsumes embezzlement cases from both labor- and works-related irregularities.

NREGA program outcomes

As for any public program, the financial scope of the program is seen as one measure of success. To explore the impact on this aim, and following Afridi and Iversen (2014), I include total NREGA expenditure as an outcome. At the same time, NREGA intends to allocate as much

---

[19] These GPUs are Tingchim Mangshila, Tingchim Chaday, Mangchila Tibuk, Naitam Nandok, Bhusuk Naitam, Nandok Saramsa, Dodak and Buriakhop. Despite a slightly different period of analysis, I apply the same restrictions to both datasets to retain comparable samples.

as possible to labor-related causes to preserve its function as a social safety net via the provision of short-term occupation. I investigate effects on this second aim by taking a separate look at program expenditure on labor and material, as well as the ratio of labor to material expenditure. Distinguishing between these two types of expenditure also allows to investigate whether external social audits influence material-related expenditure differently from labor-related expenditure (which is expected to depend more on information from beneficiaries).

Further, the primary aim of employment generation makes it essential to measure beneficiaries' access and use of the program's employment service. Demand for and access to the program are measured through the number of job cards applied for and granted, respectively. The extent of employment is measured through person days worked under the NREGA program.

A secondary goal of NREGA is the creation of infrastructure assets. I measure this through the number of construction projects in various completion states: ongoing, suspended, and completed[20]. Ongoing works are works in which some financial or labor activity is taking place. While in the long term, the number of work projects should be driven by labor demand, the number of ongoing works serves as an indicator for short-term employment opportunities. To ensure that the work projects culminate in the creation of usable assets, NREGA guidelines stipulate that all ongoing works should be completed within one financial year. This means that all necessary documents have been provided which prove that the aim of the work has been reached and the project has officially been closed. However, ongoing works can become suspended if there is currently no activity occurring in them despite the works being incomplete. This can happen for a variety of good reasons, for instance because the worksite was poorly chosen, the land owner withdrew his permission, the work is unfeasible or unlikely to be finished, or the work is almost complete but the whole sanctioned amount has been spent (personal communication with Telangana SAU personnel, 25.11.2021). However, because suspended works open up a possibility of embezzlement, ideally these works would either be deleted from the system (in the first two instances) or allocated some additional funds to be completed and closed (personal communication with Telangana SAU personnel, 25.11.2021). Incentives for leaving works nevertheless suspended for a prolonged period of time can be that officials plan to use leftover money as cushion in case of a sudden increase in labor demand, or for easy cash withdrawal for themselves. In order to ensure that results are not driven merely by a large GPU population, I report days worked and expenditure respectively per GPU inhabitant, irregularities and work projects per thousand GPU inhabitants, and job cards per household.

Main explanatory variables

In the analysis of irregularities, the main explanatory variable is a count variable which indicates the number of social audits which have been completed for a GPU (including the

[20] I use the absolute number of works because the calculation of shares of work types in total works leads to many missing observations due to zeros in the denominator.

social audit that the respective social audit report refers to). In the analysis of NREGA outcomes, the main explanatory variable consists of a count step variable ranging from one to four which indicates how many social audits were complete in the previous period. More precisely, once social audit number k has been conducted in a GPU, the count variable takes on a value of k in the following year and stays k in all following periods without an audit until the year after a new social audit has been conducted.

3.3.2 Descriptive statistics

Table 3.1 lays out citizens' participation in social audit public meetings in Sikkim. Average participation is close to the minimum stipulated 30 percent of active beneficiaries, which shows that the quota is relatively effective[21]. However, disadvantaged groups - scheduled castes and tribes - are less likely to participate in the public meeting than the more advantaged groups. Overall, participation is high enough for the public meeting to have an impact (almost twice the village *populations'* participation rate succeeding Olken's (2007) invitation experiment [of 2.5 percent]), but much lower than for user-conducted social audits in Andhra Pradesh, for which Shankar (2010) finds a *beneficiary* participation rate of 77 percent [in 2007].

Table 3.1: Summary statistics of social audit participation

	Absolute			As percent of GPU NREGA beneficiaries			As percent of GPU (sub-) population		
	Mean	SD	N(Audits)	Mean	SD	N(Audits)	Mean	SD	N(Audits)
Participants									
Total	115.46	(42.51)	392	27.52	(53.10)	390	4.26	(2.09)	392
By social group									
SC/ST	35.87	(23.51)	392	10.00	(16.00)	390	3.43	(2.37)	392
Non-SC/ST	72.83	(46.01)	392	16.06	(41.16)	390	4.40	(3.55)	392
Unclear	6.39	(21.25)	392	1.37	(4.09)	390	-	-	-

Notes: Participation in social audit public meetings derived from attendance lists in social audit reports. Social groups are categorized by surnames into scheduled castes and tribes (SC/ST), non-scheduled castes and tribes and 'unclear' if names did not allow for a classification. The number of beneficiaries in a GPU is sourced from the NREGA website. It is not available by social group and only refers to beneficiaries in the respective year. Differences in observations in are due to missing beneficiary data. Sample: 162 GPUs weighted by population frequency weights derived from census 2011, years 2011/12-2016/17.

[21] The unweighted percentage of beneficiaries participating in social audits lies with 31 percent even above the 30 percent quota.

Table 3.2: Summary statistics of irregularities and NREGA outcomes

	Mean	SD	N (Years*GPUs)
Panel A: Irregularities detected in social audits			
Detected irregularities			
Total	42.82	104.85	398
Work quality and material	17.47	34.00	398
Labor	24.99	85.48	398
Embezzlement	0.76	1.34	398
Main explanatory variable			
Social audit round	1.77	0.75	398
Panel B: NREGA outcomes			
Expenditure			
Total real disbursed expenditure (labor & material) (₹)	1353.27	946.33	1190
Real disbursed material expenditure (₹)	497.59	472.81	1190
Real disbursed labor expenditure (₹)	855.69	620.21	1190
Labor expenditure in total expenditure (percent)	64.64	20.51	1158
Employment			
Job cards issued	0.87	0.34	1190
Person days worked	8.15	5.21	1190
Work projects			
Ongoing	7.56	7.19	1190
Suspended	0.10	1.10	1190
Completed	5.44	5.87	1190
Main explanatory variable			
Completed social audit rounds (t-1)	0.99	1.14	1190

Notes: Panel A: Detected irregularities derived from social audit reports per thousand GPU inhabitants. Sample: 162 GPUs weighted by GPU population (source: census 2011), years 2011/12-2016/17. Differences in observations to panel B arise due to the unbalanced panel and additional sample restrictions, i.e. exclusion of GPUs and years for which no social audit report is available, or only in changed format (year 2017/18). Panel B: NREGA program expenditure per GPU inhabitant (₹, 2010 prices), job cards per GPU household, person days per GPU inhabitant, works per thousand inhabitants. Differences in observations in outcomes representing percentages arise whenever the underlying denominator equals zero. Sample: 170 GPUs weighted by population frequency weights derived from census 2011, years 2011/12 to 2017/18.

Panel A of Table 3.2 depicts the number of irregularities brought to light by social audits. GPUs have on average 1.8 social audits, each of which detects on average 43 irregularities per thousand GPU inhabitants. Labor-related irregularities are detected much more frequently than material and work quality-related irregularities, despite the fact that both categories encompass roughly the same number of issue types. 1.8 percent of detected material and labor irregularities are embezzlement cases. While this seems comparatively low at first sight, a single embezzlement case can go hand-in-hand with the leakage of large amounts of funds: a detected embezzlement case is on average related to the misuse of ₹31,956. In 79 percent of detected embezzlement cases in my sample, the embezzled funds are recovered in the aftermath of the social audit (not shown in Table 3.2). The descriptive statistics for NREGA outcomes are set out in panel B. Total NREGA program expenditure (excluding administrative costs) amounts to about ₹1353 per GPU inhabitant. Labor expenditure constitutes 65 percent of the total amount, in line with the stipulated minimum 60 percent. An overwhelming majority (87 percent) of households hold job cards and a GPU inhabitant works on average eight person days in NREGA. Each year, there are about eight ongoing NREGA work projects per thousand inhabitants. The step variable for completed social audit rounds in panel B shows an average of one audit over all year-GPU observations. More precisely (not depicted in Table 3.2): in 52 percent of observations the first audit has been completed, in 30 percent the second audit, in 15 percent the third audit and in 1 percent the fourth audit.

3.4 Empirical approach

3.4.1 Estimating effects of social audits on detected irregularities

If repeated social audits have deterrent or learning effects, they should lead to changes in the detection of irregularities, which in turn should catalyze changes in downstream NREGA program outcomes. In order to support the results regarding program outcomes, I first estimate the effect of an additional social audit on the number of irregularities detected. To this end, I exploit the phase-in of social audits in Sikkim over four years. As information on irregularities is only available in years without audits, the variation in the explanatory variable of the simple OLS regression thus stems from the fact that GPUs received the kth social audit at different points in time (Afridi and Iversen 2014). The explanatory variable is modeled as a count variable to account for the fact that social audits' effects may be cumulative or realized only in the long term (Gordon et al. 2019). The regression equation is depicted below:

$$Y_{it} = \alpha_i + \delta Audit_{it} + \varepsilon_{it} \tag{3.1}$$

where t indexes year and i Gram Panchayat Unit. The outcome, Y_{it}, is the number of irregularities per GPU inhabitant reported for a given GPU and year. $Audit_{it}$ is a count variable indicating which social audit round k took place in GPU i in year t. The variable takes on values between one and four. δ represents the average effect of an additional audit and α_i are GPU fixed effects which control for differences between GPUs in time-invariant characteristics. This may include for instance bureaucratic institutions, geography, and to some extent initial differences in size and population composition which influence the implementation of social audits or NREGA (Afridi and Iversen 2014). ε_{ikt} is the error term. Within years, SAs are conducted across all districts simultaneously and no clear correlation pattern is detected at the subjacent block administrative level. Therefore, as it is custom, standard errors are clustered at the level at which social audits are implemented, the GPU level (Abadie et al. 2017). This accounts for the fact that standard errors will be correlated over time for each GPU.

Year fixed effects are avoided because they are highly correlated with the number of social audit rounds and thus absorb the main source of variation in irregularities. Consequently, I cannot fully control for unobserved heterogeneity over time in the analysis of irregularities. However, I address this issue partially by reporting, jointly with the estimates from equation (3.1), estimates with additional period fixed effects for periods with enough variation in audit rounds for reasonable identification. These two periods are the blocks of years before and after 2015 (the year from which on all GPUs were audited yearly).

3.4.2 Estimating effects of social audits on NREGA outcomes

In contrast to irregularities, data on NREGA program outcomes is available also for years in which no social audit took place. This allows to extend the analysis to a full fixed-effects panel regression (Wooldridge 2010) that can control also for GPU-specific time trends in social audit implementation. The social audit count variable now becomes a lagged step variable: once the kth audit has been completed, the step variable takes on the value k in all following rounds without an audit, until a new audit is conducted, after which it takes on the value k+1. The corresponding regression equation is

$$y_{it} = \alpha_i + \gamma_t + \beta AuditStep_{it-1} + \rho X_{it} + \varepsilon_{it} \tag{3.2}$$

where i indexes GPU and t financial year. y_{it} represents an NREGA program outcome of GPU i in year t, divided by population. $AuditStep_{it-}$ indicates how many audits have been completed by the previous period (t-1), α_i are GPU fixed effects and γ_t are year fixed effects. The latter account for GPU-wide time trends such as economic growth, or other changes in NREGA policy which could curb corruption independently from social audits.[22] ε_{it} is the error term. As in the previous regression, standard errors are clustered at the GPU level. To avoid that results are driven by a change in overall work activity, X_{it} controls for the total number of works when the outcome is the number of specific work projects. This yields very conservative estimates as there is some endogeneity between the number works of a certain completion status and the total number of works.

Accounting for multiple inference and heterogeneity

In both regression specifications, I address two threats to internal validity, multiple inference and heteroscedasticity: due to the large number of hypotheses being tested in this paper, uncorrected p-values are more likely to falsely reject the null hypothesis of no effect. I account for this by adjusting the p-values of each regression for families of outcomes using the bootstrapping methodology by Romano and Wolf (2005). This method corrects p-values through controlling the family-wise error rate. All regressions are weighted by GPU population for two reasons. First, it increases precision in the case of heteroscedasticity (Solon et al. 2015).

[22] Afridi and Iversen (2014) in addition control for the introduction of payment through Aadhar cards with biometric identification. However, in the case of Sikkim, these were already introduced in 2010/2011, before social audits started.

Second, by giving more weight to larger villages, it makes the data representative to the population of Sikkim as a whole.

3.5 Results

3.5.1 Effects of social audits on detected irregularities and NREGA outcomes

When examining the effects of social audits, I first provide some evidence that social audits created a movement in irregularities in the program. Table 3.3 lays out the effects of social audits on the number of detected irregularities. The results show that a greater number of social audit rounds decreases detected irregularities of all categories. In the specification without period fixed effects, an additional social audit round leads on average to 42 less detected irregularities. In other words, the number of total irregularities almost halves compared to the number of irregularities detected in the first audit. However, a significant decrease in both specifications without and with period fixed effects can only be seen for material irregularities and embezzlement. Embezzlement, although relatively rare to begin with, decreases depending on the specification by 0.3 to 0.6 cases, 20 to 42 percent compared to the first audit.

Turning to the effects on NREGA program outcomes, Table 3.4 displays the effects on NREGA program expenditure. Social audits increase material expenditures by ₹81 per inhabitant. This is equivalent to roughly 18 percent of the control mean. The result is significant at the 5 percent level using conventional p-values, and with a multiple-inference-adjusted p-value of 0.1046 falls just short of the adjusted 10 percent significance level. Labor expenditure, in turn, decreases by ₹54 per person, although the result is not statistically significant. In accordance with a larger increase in material expenditure than decrease in labor expenditure, total expenditure increases slightly and the percentage of labor in total expenditure decreases by about 4 percentage points. Hence, labor expenditure remains slightly above the 60 percent of overall costs stipulated to preserve NREGA's function as a social safety net. However, neither of these effects is statistically significant from zero.

The stagnation (or even decrease) in labor expenditure is corroborated by the results regarding employment generation in NREGA, set out in Table 3.5. Both the number of job cards applied for and issued, as well as person days worked per GPU inhabitant slightly decrease. However, only the decreases in job cards are significant at the 10 percent level, and the decrease in job cards issued ceases to be significant when adjusting for multiple inference.

The work project completion status can provide insight into the short-term capacity for employment and the creation of infrastructure assets. Table 3.6 shows the effects of social audits on the number of ongoing, suspended and completed work projects, controlling for the total number of works ever conducted in a GPU. As the total number of works is correlated

Table 3.3: Irregularities detected during social audits

	Total		Labor		Work quality/ material		Embezzlement	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SA round	-41.89***	-14.43	-28.11***	-5.93	-13.77***	-8.50**	-0.63***	-0.33***
	(6.83)	(17.48)	(6.51)	(15.50)	(2.11)	(4.13)	(0.09)	(0.12)
			{0.01***}	{0.74}	{0.00***}	{0.32}		
Period fixed effects		Yes		Yes		Yes		Yes
Control mean	91.5	91.5	56.8	56.8	34.6	34.6	1.5	1.5
Observations (Audits)	398	398	398	398	398	398	398	398
Clusters (GPUs)	170	170	170	170	170	170	170	170

Notes: Linear regression. Outcomes are irregularities detected during social audits per 1,000 GPU inhabitants. Embezzlement includes both labor and material related embezzlement. Main explanatory variable is social audit (SA) round (range: 1 to 4). Control mean indicates the mean outcome in the first social audit. Estimates computed using sampling weights and GPU fixed effects in all specifications, estimates in columns 2, 4, 6, and 7 include a dummy equal to one for the period after 2015. Standard errors corrected for clustering at the GPU level in parenthesis. Multiple inference adjusted Romano-Wolf p-values (for columns 3 and 5, 4 and 6, respectively) in curly brackets. Sample: 398 audits in 170 GPUs weighted by population weights derived from census 2011, years 2011/12-2016/17.
* p<0.10 ** p<0.05 *** p<0.01

Table 3.4: Material and labor expenditure

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t-1)	27.70	81.42**	-53.72	-4.42
	(67.63)	(40.24)	(50.72)	(3.05)
		{0.10}	{0.31}	
Control mean	1217.3	464.4	752.9	64.5
Observations (Years*GPUs)	1190	1190	1190	1156
Clusters (GPUs)	170	170	170	170

Notes: Fixed effects panel regression. Outcomes are real disbursed NREGA program expenditure per GPU inhabitant (₹, 2010 prices) (columns 1 - 3) and percentage of labor in total expenditure (column 4). Main explanatory variable is the number of social audits (SA) completed up to the previous year. Control mean indicates the mean outcome if the main explanatory variable equals zero, i.e. no social audit took place in all previous years. Estimates computed using sampling weights, year and GPU fixed effects. Standard errors corrected for clustering at the GPU level in parenthesis. Multiple inference adjusted Romano-Wolf p-values (for columns 2 - 3) in curly brackets. Sample: 170 GPUs weighted by population weights derived from census 2011, years 2011/12-2017/18. Discrepancies in observations in column (4) arise when the denominator total expenditure equals zero.
* p<0.10 ** p<0.05 *** p<0.01

Table 3.5: Employment

	Job cards applied	Job cards issued	Person days
	(1)	(2)	(2)
SA round complete (t-1)	-0.07**	-0.07**	-0.38
	(0.03)	(0.03)	(0.47)
		{0.14}	{0.43}
Control mean	0.9	0.9	7.7
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: Outcomes are NREGA job cards per GPU household and NREGA person days worked per GPU inhabitant. Multiple inference adjusted Romano-Wolf p-values (for columns 2 - 3) in curly brackets. All other notes from Table 3.4 apply.
* p<0.10 ** p<0.05 *** p<0.01

Table 3.6: NREGA projects

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA round complete (t-1)	0.72	0.16**	-1.13**
	(0.62)	(0.07)	(0.46)
	{0.26}	{0.10*}	{0.08*}
Control mean	5.9	0.2	3.8
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: Outcomes are ongoing, suspended and completed NREGA work projects per 1,000 GPU inhabitants. Estimates computed controlling for the total number of works. Multiple inference adjusted Romano-Wolf p-values (for columns 1 - 3) in curly brackets. All other notes from Table 3.4 apply.
* p<0.10 ** p<0.05 *** p<0.01

with the respective outcomes, the estimates represent lower bounds of the actual effects. The number of ongoing works per thousand GPU inhabitants increases but this increase is not statistically significant. Suspended works increase significantly by 0.16 works per thousand inhabitants. The result remains significant at the ten percent level when adjusting for multiple hypotheses (the p-value equals 0.098). At the same time, 1.13 less works per thousand inhabitants are completed. The decline in completed works is significant at the 10 percent level when adjusting for multiple inference and corresponds to a large change of around 30 percent of the control mean.

3.5.2 Internal validity and robustness

An assumption of my empirical approach is that GPUs which received social audits earlier, and consequently completed a greater number of audit rounds within the period of analysis, had parallel trends in outcomes before they received social audits, compared to GPUs which received social audits at a later point in time. To test this hypothesis, I conduct a parallel trend placebo test (Kahn-Lang & Lang, 2020) in which I repeat the main analysis for NREGA outcomes for the years 2011 and 2012 for the subsample of GPUs which received their first audit after 2012 (these make up 96 percent of GPUs in the full sample). I have to revert to a subsample and NREGA outcomes for the following reasons: a balancing or parallel trends test with my full regression sample would require GPU-level data on outcomes from the time before social audits started or, in case of a balancing test, at least a balanced sample from the year in which audits are introduced. However, data on irregularities is only available for years in which social audits take place, making either test unfeasible. GPU-level NREGA outcomes are only available from 2011/12 on so that I cannot test for parallel trends before 2011 (when social audits were introduced)[23]. Moreover, only four GPUs received a social audit in 2011, so that a balancing test for this year would be based on two very differently sized samples. While this makes a balancing test for the year of introduction unreliable, it is also a strength because it ensures that the lagged social audit count variable in the analysis includes GPU-year observations in which social audits have not yet been introduced. Moreover, it allows me to retain a sufficiently large representative subsample for the placebo test when excluding the few earliest audited GPUs. The data for the placebo sample are summarized in Table A3.3 in the appendix. The placebo test results in Tables 3.7 – 3.9 show that receiving a social audit early (in 2013) or late (in 2014) has no significant effect on program outcomes before 2013.

In order to ensure that my main results and conclusions are robust to the regression specification, I conduct two robustness checks. First, I estimate equations (3.1) and (3.2) without population weights, second, I employ social audit round dummies as explanatory variables in lieu of the number of rounds[24]. The estimates without population weights (Tables A3.5 to A3.8) are overall similar, with all point estimates having the same sign as the estimates in section 3.5.1, except the estimate for ongoing works (which is, however, not significant in

[23] Neither is it possible to conduct parallel trends tests for the time before audit introduction with related outcomes, e.g. general unemployment, non-NREGA infrastructure or unskilled wages, due to lack of disaggregated data at the GPU level. While e.g. the economic census for the years 1990, 1998, 2005 and 2013 surveyed employment for some villages, these constitute only a fraction of villages in Sikkim. Moreover, the data dates either from a time before the onset of NREGA or after the introduction of social audits in Sikkim. The same applies to the population census 2001 and 2011.

[24] For the sake of completeness, I also report estimates for irregularities with year fixed effects in Table A3.4. However, as discussed in section 3.4.1, year fixed effects absorb most variation in audit rounds which makes identification impossible.

Table 3.7: Material and labor expenditure, placebo test

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t+1)	-57.89	3.28	-61.17	-0.92
	(98.90)	(69.62)	(58.75)	(3.40)
		{0.96}	{0.53}	
Control mean	1229.9	475.2	754.7	63.9
Observations (Years*GPUs)	984	984	984	952
Clusters (GPUs)	164	164	164	164

Notes: See Table 3.4. Main explanatory variable is the number of social audits completed in the following year (equal to the second lead of the original explanatory variable in Tables 4-6). Control mean indicates the mean outcome if the main explanatory variable equals zero, i.e. no placebo social audit took place in all previous years. Sample: 164 GPUs weighted by population weights derived from census 2011, years 2011/12 and 2012/13, excludes 6 GPUs which received their first social audit in 2011/12 or 2012/13.

* p<0.10 ** p<0.05 *** p<0.01

Table 3.8: Employment, placebo test

	Job cards applied	Job cards issued	Person days
	(1)	(2)	(3)
SA round complete (t-1)	0.00	0.00	0.07
	(0.04)	(0.03)	(0.48)
		{0.98}	{0.98}
Control mean	0.9	0.9	7.7
Observations (Years*GPUs)	984	984	984
Clusters (GPUs)	164	164	164

Notes: See Table 3.5. Main explanatory variable is the number of social audits completed in the following year (equal to the second lead of the original explanatory variable in Tables 4-6). Control mean indicates the mean outcome if the main explanatory variable equals zero, i.e. no placebo social audit took place in all previous years. Sample: 164 GPUs weighted by population weights derived from census 2011, years 2011/12 and 2012/13, excludes 6 GPUs which received their first social audit in 2011/12 or 2012/13.

* p<0.10 ** p<0.05 *** p<0.01

Table 3.9: NREGA projects, placebo test

	Ongoing (1)	Suspended (2)	Completed (3)
SA round complete (t-1)	0.78	-0.06	-0.77
	(0.66)	(0.14)	(0.53)
	{0.46}	{0.69}	{0.39}
Control mean	6.0	0.2	3.7
Observations (Years*GPUs)	984	984	984
Clusters (GPUs)	164	164	164

Notes: See Table 3.6. Main explanatory variable is the number of social audits completed in the following year (equal to the second lead of the original explanatory variable in Tables 4-6). Control mean indicates the mean outcome if the main explanatory variable equals zero, i.e. no placebo social audit took place in all previous years. Sample: 164 GPUs weighted by population weights derived from census 2011, years 2011/12 and 2012/13, excludes 6 GPUs which received their first social audit in 2011/12 or 2012/13.
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

either weighted or unweighted estimates). As expected, the unweighted estimates are less precise than the weighted estimates, indicating the presence of unaccounted-for heteroscedasticity in the unweighted results. For ease of interpretation, my main specification computes average effects over rounds. These might mask large differences in the effects of particular social audit rounds. For instance, if effects only materialized in the very long term, average effects might be too small to be detected. I investigate this in Tables A3.9 to A3.12 by performing all analyses with separate social audit round step variables. Overall, the round-wise results are very much in line with the average effects estimates. What stands out is that the significant increase in material expenditure is mainly driven by the last two rounds. Consonant with this, the labor-material ratio decreases more and more with each audit round until the decrease even becomes significant in the last two rounds. Effects on total expenditure differ to some extent by round but are for all but the last round insignificant.

3.5.3 Potential mechanisms

Can the changes in outcomes observed in the main results in section 3.5.1 be evaluated as beneficial? To assess this, I delve below deeper into possible mechanisms of these changes. These explorations are based on alternative outcomes and sample restrictions rather than the estimation of heterogeneous effects, because the latter would be based on comparisons of very unbalanced subsamples. Mechanisms that point towards a reduction in corruption or mismanagement, or which increase quality or efficiency are viewed as favorable developments, whereas mechanisms related to an increase in corruption or inefficiencies are deemed adverse effects.

Mechanisms for the decrease in detected irregularities

Table 3.10: Potential mechanisms for detected irregularities

	Percent recovered
	(1)
SA round complete (t-1)	10.25
	(8.03)
Control mean	72.8
Observations (Years*GPUs)	109
Clusters (GPUs)	90

Notes: See Table 3.3. Outcome is percent of embezzled funds which have been recovered. Sample: social audits with detected embezzlement cases.
* p<0.10 ** p<0.05 *** p<0.01

Social audits cause a shift in detected irregularities, but whether this is a desired effect may be interpreted ambiguously: a decrease in detected irregularities can be observed if social audits have the desired deterring effect, but also if officials become more adept at hiding irregularities over time, which may be perfectly correlated with the number of audit rounds (Tambe et al. 2016b). Therefore, the results regarding irregularities mainly serve to show that social audits either set some learning or deterrence effects in motion, while meaningful conclusions regarding the increase or decrease of corruption are difficult. Notwithstanding this, some anecdotal evidence speaks for the hypothesis that learning of corrupt bureaucrats is limited, or at least counteracted by similar learning of auditors, and that the decrease in irregularities is thus genuine. First of all, the observed decrease in embezzlement is in accordance with a decrease in missing expenditure observed by Olken (2007) in an external social audit. Second, the detected embezzled amount in Sikkim's social audits decreases with each audit but the share of recovered funds remains constant (the point estimate is even positive but imprecisely estimated (Table 3.10)). This suggests that officials at least fail to learn how to avoid repayment when getting caught.

Mechanisms for the decrease in completed works and increase in suspended works

Table 3.11 is dedicated to the examination of different mechanisms which could explain the effects of social audits on work projects. Taken together, the pattern of effects found on the work completion status (increase in ongoing works, significant increase in suspended works and significant decrease in completed works) suggests that social audits either impede the completion of works, shift corruption to suspended works, prolong the approval phase, or cause a mere relabeling (correction or mislabeling) of the work status.

The fact that the decrease in completed works goes hand-in-hand with an increase in suspended works suggests that works which would formerly have entered the system as completed are now entered as suspended. Four pathways allow for a strong substitution from completed to suspended works. Firstly, social audits may detect unsuitable works, which are better abandoned at an early stage because they are inefficient or doomed to fail. I test for this pathway by excluding GPUs from the sample in which cases of poorly selected worksites and works not conforming to standards were detected during social audits. It becomes evident from columns 1 to 3, that the decrease in "efficient" completed works is about a quarter smaller than in the full sample. That the increase in "efficient" ongoing works is also smaller indicates that these were indeed works beyond repair. The point estimate for suspended works barely changes, suggesting that works which are discontinued due to efficiency reasons are rather deleted from the system than being labeled suspended.

Second, suspended works may also rise if bureaucrats keep completed works with leftover funds suspended on the shelf in case employment demand suddenly rises. Expanding this strategy would only make sense if social audits decreased ongoing or approved works (which are not yet activated) that bureaucrats could allocate labor to in case of surging employment demand. This could happen if social audits' scrutiny led bureaucrats to adhere more to stipulated but time-consuming approval processes. However, neither ongoing (Table 3.6) nor approved works (Table 3.11, column 7) decrease. Taken together, these results indicate that the changes in work status are not driven by preparations for a sudden rise in employment demand.

A third, less favorable explanation for a simultaneous decrease in completed and increase in suspended works would be that a work is physically almost complete but is not finalized because the whole sanctioned amount has already been spent. Social audits could aggravate this if they impede the shift of funds between different projects. However, columns 5 and 6 in Table 3.11 show that controlling for the amount spent does not weaken the effect of social audits on suspended and completed works.

In a fourth mechanism, officials may keep physically complete projects with leftover funds suspended in order to embezzle small amounts under them from time to time. However, although the results in columns 4-6 show that projects with leftover funds are significantly more likely to be suspended and less likely to be closed, controlling for this has almost no influence on the effect of social audits on these outcomes. This makes an increase in corruption an unlikely pathway for the observed results for works.

Evidently, substitution from completed to suspended works cannot explain the entire change in completed works, as completed works decrease to a greater extent than suspended works increase. This pattern can arise if social audits detect works which are entered as complete in the system but either do not exist or are incomplete. Such a detection would entail a deletion of purported "completed" ghost works or a correction of the status of existing works which have been wrongly declared as 'completed' to 'ongoing'. I examine this by excluding GPUs from the sample for which social audits found irregularities that could cause a correction or deletion of works in the system, including cases of missing completion certificates. The results are set out in columns 8 to 10. The effect on completed works is almost halved compared to the original full sample regression, demonstrating the relevance of this mechanism.

A final explanation is offered by Tambe et al. (2016b), who report that in the initial years, only completed works were audited in detail, causing bureaucrats to falsely declare works as ongoing instead of completed in the following[25]. However, the auditors detected this and

[25] This is a distinct pathway from the status correction/work deletion: I find no indication that GPUs in which no ghost works etc. were detected were less likely to hide corruption.

Table 3.11: Potential mechanisms for effects on project completion status

Sample:	Efficient			Full			Full	No status correction by auditors			Since 2014/15	
Work status:	Ongoing	Suspended	Completed	Ongoing	Suspended	Completed	Approved	Ongoing	Suspended	Completed	Ongoing	Completed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SA round complete (t-1)	0.44	0.17**	-0.90*	0.89	0.15*	-1.10**	0.76	0.51	0.13*	-0.67	0.22	0.23
	(0.62)	(0.07)	(0.47)	(0.61)	(0.08)	(0.47)	(0.54)	(0.64)	(0.07)	(0.47)	(0.80)	(0.61)
	{0.49}	{0.11}	{0.11}	{0.16}	{0.14}	{0.09*}		{0.45}	{0.24}	{0.26}	{0.88}	{0.88}
10-99% of sanctioned funds disbursed				-0.28	0.18**	-0.70**						
				(0.42)	(0.09)	(0.33)						
				{0.53}	{0.16}	{0.16}						
All sanctioned funds disbursed (≥100%)				-3.54***	0.29	-0.31						
				(0.67)	(0.38)	(0.65)						
				{0.22}	{0.75}	{0.75}						
Control mean	6.0	0.2	3.8	5.9	0.2	3.8	1.9	6.0	0.2	3.8	5.7	4.9
Observations (Years*GPUs)	1141	1141	1141	1190	1190	1190	1190	1085	1085	1085	680	680
Clusters (GPUs)	163	163	163	170	170	170	170	155	155	155	170	170

Notes: See Table 3.6. Outcomes are ongoing, suspended, completed and approved works per thousand inhabitants. Columns 1 - 3 exclude GPUs in which cases of poorly selected worksite or works not conforming to standards were detected. Columns 4 - 6 control for the percentage of funds disbursed (it can exceed 100 if additional funds are granted for project completion). Columns 4 - 6 control for funds disbursed; reference category is less than 10 percent of funds disbursed. The sample in columns 8 – 10 excludes GPUs in which cases of fake or missing completion certificate, ghostworks or non-achieved physical work targets were detected during social audits. The subsample in columns 11 - 12 spans the years 2014/15 or later. No works were suspended during this period. The sub-samples consist of 163 GPUs (columns 1 - 3) and 155 (columns 8 - 10) out of originally 170 GPUs, and four (columns 11 - 12) out of seven years. Multiple inference adjusted Romano-Wolf p-values (for columns 1 - 3, 4 - 6, 8 - 10, 11 - 12 respectively) in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

Table 3.12: Potential mechanisms for effects on employment

Sample:	Excluding ghost labor			Full		Excluding late wages		
	Job cards applied	Job cards issued	Days worked	Days worked by SC/ST (percent)	Days worked by women (percent)	Job cards applied	Job cards issued	Days worked
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SA round complete (t-1)	-0.07**	-0.07**	-0.40	-0.08	-0.15	-0.07*	-0.07*	-0.48
	(0.04)	(0.04)	(0.50)	(0.82)	(0.88)	(0.04)	(0.04)	(0.57)
		{0.12}	{0.45}	{0.98}	{0.98}		{0.60}	{0.60}
Control mean	0.9	0.9	7.7	42.7	46.6	0.9	0.9	7.3
Observations (Years*GPUs)	1148	1148	1148	1155	1155	875	875	875
Clusters (GPUs)	164	164	164	170	170	125	125	125

Notes: See Table 3.5. Outcomes: number of job cards applied and issued per household, total person days worked per GPU inhabitant, person days worked by scheduled castes and tribes (SC/ST) and women as share of total days worked. The sample in columns 1 to 3 is restricted to 164 GPUs (out of originally 170 GPUs) in which no cases of ghostworkers, fake registrations or payments to non-existent persons were detected. The sample in columns 6 to 8 is restricted to GPUs in which no cases of delayed wages were detected. Discrepancies in sample size in columns 4 and 5 with respect to the first three columns arise due to missings in percent variables when total days worked are zero. Multiple inference adjusted Romano-Wolf p-values (for columns 2 - 3, and 4 - 5) in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

Table 3.13: Potential mechanisms for effects on expenditure

	Material expenditure disbursed as percent of sanctioned	Total sanctioned expenditure on material intensive works	Material expenditure	Detected less/poorer quality material supplied	Detected excessive rates/material/ technical estimate
	(1)	(2)	(3)	(4)	(5)
SA round complete (t-1)	1.01	9621.01	83.02**		
	(1.16)	(558304.50)	(40.64)		
SA round				-0.60***	-0.07
				(0.23)	(0.05)
				{0.18}	{0.19}
Ongoing works			-1.36		
			(2.04)		
Control mean	10.7	1376055.4	467.2	1.3	0.0
Observations (Years*GPUs)	1167	1190	1190	398	397
Clusters (GPUs)	170	170	170	170	170

Notes: See Tables 3.3 (for columns 4 - 5) and 4 (for columns 1 - 3). Outcomes are disbursed NREGA program expenditure on material (2010 prices) per GPU inhabitant (₹), as percent of total sanctioned (material and labor) expenditure; real sanctioned expenditure on material intensive works (sanitation, rural connectivity and "other" works); real disbursed material expenditure; number of detected cases of less or poorer quality material supplied than paid for; number of detected cases of excessive rates paid, excessive amounts of material paid for, or technical estimate exaggerated. Column 3 controls in addition for the number of ongoing works. Multiple inference adjusted Romano-Wolf p-values for columns 4 - 5 in curly brackets. Control mean for excessive rates is 0.03.

* p<0.10 ** p<0.05 *** p<0.01

reacted by auditing also ongoing works in detail, from 2014/15 on. That this strategy proved successful is demonstrated in columns 11 and 12. When restricting the sample to the years since 2014/15[26], the increase in ongoing works is strongly reduced, there are no new suspended works, and completed works increase insignificantly (indicating that the "backlog" of hidden works is now officially labeled as completed).

In sum, the observed change in the composition of work projects cannot be explained by bureaucrats expanding suspended works as backup in case of employment shocks, or by social audits complicating the redistribution of funds for work completion. Instead, completed works decrease due to the deletion of detected ghost works and correction of the completion status of unfinished works wrongly declared as completed by auditors, and due to the resumption or closure of poorly executed works. Thus, it can be concluded that social audits had a beneficial effect on works by fostering efficiency and detecting corruption. While the work composition partly also changed due to bureaucrat's attempts to avoid inspection by labeling works as ongoing or suspended rather than completed, this likely served to hide existing corruption and thus does not offset the beneficial effects outlined above. Moreover, the attempt to avoid inspection was successfully counteracted in the long run by a revised auditing strategy which included ongoing works. To this may be added that a further likely beneficial effect of external social audits on works which is not captured by the number of work projects is that social auditors detected that the *recipients* of private assets constructed under NREGA were frequently not the poor (Tambe et al. 2016b). After this was discovered, the criteria of who should benefit from such assets were tightened. While I lack information on private asset recipients, the success of this measure is supported by a significant decrease in detected cases of works not selected based on the recommendation of the Gram Sabha.

Mechanisms for stagnation or decrease in employment

Ideally, social audits would decrease discrimination, corruption and hurdles in the application for job cards and work, thus making employment in NREGA more attractive and accessible. This, however, is not mirrored in the main results, which give some weak evidence for a decrease, at best a stagnation, in employment.

A *stagnation* could indicate either that labor infringements were quite low to begin with or that social audits did a poor job at detecting those types of labor infringements most important for employment generation. However, this is at odds with the high initial number of labor issues detected in the first audit and their strong decrease in the following audits. Those irregularities which have low initial detection rates (fake registrations, ghost workers and wages paid to them) do not entail a personal loss to beneficiaries and are thus unlikely to influence beneficiaries' employment decisions.

A *decrease* in employment can be witnessed if (i) ghost workers and unworked hours are detected during social audits and removed from the records in the following year, (ii) social audits increase entry barriers at job card issuance or (iii) discourage workers from enrolling in

[26] The period *before* 2014/15 does not have sufficient variation in audit rounds for separate inference.

and using the program. Regarding the first mechanism, fake registrations, ghost workers and hours worked by them are detected very rarely (fifteen cases in six GPUs). Accordingly, restricting the sample to GPUs without these detected cases does not change the results regarding number of job cards and decreases the number of days worked only slightly (Table 3.12, columns 1 to 3). Another, here not testable, possibility is also that social audits uncovered collusion between workers and worksite supervisors to disburse money for unworked hours. Detection of this type of embezzlement is difficult, but if successful may deter beneficiaries and officers from engaging in collusion in the following year, and possibly even from enrolling in the program if they are deterred by the prospect of having to provide hard manual labor. While instances of collusion have been reported for Sikkim (Tambe et al. 2016b), this type of corruption is not recorded separately in the social audit reports, making a verification of this theory difficult.

Regarding the second and third mechanism, if *applications* for job cards remained high, this would point to *increasing* entry barriers at job card issuance, such as bribery or discrimination. However, applications decrease by the same extent as issued job cards, indicating a drop in demand, not supply of job cards. A potential reason for this drop may be, as discussed above, that detection of collusion makes enrollment in the program potentially less attractive. Further, social audits may either discourage more advantaged or disadvantaged groups from applying for the program if audits enforce a stricter bureaucratic process and documentation, resulting in more paper work for applicants. While I lack disaggregated data for job card applications by different social groups, a shift in the social composition of applicants should be mirrored by a shift in the number of days worked by certain social groups. However, discouragement of certain groups is at best a weak channel: as columns 4 and 5 in Table 3.12 show, social audits lead to a decrease in the share of days worked by disadvantaged groups (women, scheduled castes and tribes), but these changes are statistically not significantly different from zero. Another possibility is that through social audits, beneficiaries become more aware of and are discouraged by *existing* entry barriers. For instance, Narayanan and Das (2014) show that workers reduce their future labor supply in NREGA when faced with rationing of jobs and late payment of wages. Learning about entry barriers and ensuing discouragement, are more likely if social audits frequently uncover such barriers. On the one hand, I find no evidence for discouragement due to rationing of employment through bureaucrats: there are only two detected instances in which job card issuance was refused, and one case in which work was refused. On the other hand, late payment of wages is very frequent (4.4 cases per social audit). Notwithstanding this, excluding GPUs with instances of late wages from the regression does not change the results on employment outcomes significantly (Table 3.12, columns 6 to 8).

A final explanation for both stagnating or decreasing enrollment and hours worked may be low public meeting participation. Most labor irregularities can only be detected through reports from beneficiaries. A high participation of beneficiaries in the public meetings ensures the inclusion of information from beneficiaries who failed to be interviewed, and cross-checking of information provided by other beneficiaries. It further raises the likelihood of social sanctions. In this way, high beneficiary participation could improve employment conditions and thus encourage future labor supply in NREGA. In contrast, poor working conditions likely persist after social audits with low beneficiary participation, which may even discourage future labor supply of frustrated beneficiaries. On the other hand, awareness about entry barriers may increase with participation. I shed some light on the role of beneficiary involvement as a catalyst by introducing an additional explanatory variable, the cumulative percentage of the GPU population participating in social audit meetings up until the previous year. If no social audit has yet been conducted in a GPU, the variable is coded as zero. It is of course highly correlated with the number of previously conducted audit rounds, but Tables A3.13 to A3.16 reveal some weak evidence that the participation component has different effects[27]. In contrast to the number of social audits conducted, participation in public meetings may be endogenous to previous social audit outcomes, e.g. beneficiaries may be discouraged from participating if past audit results were unsatisfactory. It is also possible that unobserved employment shocks in the same period as the audit (t-1) influence both social audit participation in t-1, as well as employment in the current period t. Past outcomes can thus constitute time-varying omitted variables which are not captured by unit fixed effects. While these problems could be solved through an instrumental variable approach, I find that a potential instrument, social audit hearings taking place in the agricultural slack season, is only weakly correlated with participation and would thus lead to inconsistent estimates (Chao and Swanson 2005). In order to at least get an idea of the possible boundaries of effect sizes, I follow Angrist and Pischke (2008) and report a robustness test with a double lag of the respective outcome instead of GPU fixed effects. Positive treatment effects from the specification with unit fixed effects and with the double lag can be interpreted respectively as lower and upper bounds of the true effects (Angrist and Pischke 2008). However, while the inclusion of a double lag addresses the problem of past audit results influencing future participation, it does not address the presence of potential unobserved correlates of both social audits in t-1 and outcomes in t. Therefore, the following results still have to be interpreted with caution. While higher participation is in the lower bound estimates not associated with job cards issued, labor days increase insignificantly by 0.09 days per person (Table A3.15). The upper bound estimates even allow for a significant increase in both outcomes.

In conclusion, I can rule out that issued job cards decrease because of detected corruption (i.e. the removal of fake job cards). Instead, the demand for job cards applications decreases. However, the full cause of this drop in demand remains unclear: detection of existing barriers in job card application and issuance is low, making it unlikely that greater knowledge about such barriers discourages citizens from applying. However, I find weak evidence that discouragement of particularly vulnerable groups may be partially responsible for this demand drop. A last possible mechanism, which I could not test due to data restrictions, is that the program becomes less attractive because social auditors detect collusion between workers and

---

[27] This also applies to work status: conservative upper bound estimates (with outcome lags) of the effect of participation on work status show no effect on work status whereas the lower bound estimates (with year fixed effects) allow for a small significant negative correlation with ongoing and suspended works.

bureaucrats. I find no to weak evidence that higher beneficiary participation in public meetings could increase days worked in NREGA and job card demand.

Mechanisms for an increase in material expenditure

As officials are instructed to expand works if employment demand rises, program expenditure should rise and fall with employment demand (personal communication with Telangana SAU personnel, 25.11.2021). This is only partly in line with what is observed: on the one hand, consistent with the result that social audits influence labor supply negatively or not at all, labor expenditure decreases insignificantly. On the other hand, material expenditure increases in spite of this. Four explanations for the misaligned rise in material expenditure are possible: shifts towards more material intensive projects, material-related corruption or budgets that include higher quality material, and greater effectiveness of external auditors in addressing irregularities which influence material expenditure (compared to labor expenditure).

Table 3.13 investigates some of these mechanisms. Column 1 shows that the percentage of sanctioned material expenditure which is disbursed barely changes. Clearly, sanctioned material expenditure increases along with actual material expenditure. This indicates that material expenditure rises due to decisions taken at project approval, for instance a change in project composition. To rule out a shift towards projects with a high share of material expenditure, I estimate the effect of social audits on sanctioned expenditure of projects categorized as material-intensive[28] in column 2. I find no significant effect. Material expenditure could also increase due to the increase in ongoing projects discussed above. However, the increase in material expenditure persists when controlling for the number of ongoing projects (column 3).

Bureaucrats may shift from labor-related corruption to corruption in material procurement to avoid detection. This corruption would have to occur at a high level (sanctioned budgets). However, several observations speak against this hypothesis: first, according to Andhra Pradesh/Telangana SAU personnel, material-related issues are not systematically easier to hide than labor-related issues in user-conducted social audits and it is even less likely that such a nexus exists with skilled external auditors (Olken 2007; Afridi and Iversen 2014). Second, even if excessive material rates and quantities were easier to hide, if the detection rate remains constant, expanding this type of corruption should still be mirrored in an increase in the respective absolute number of detected cases. However, Table 3.13 reveals that additional audit rounds significantly *decrease* the detected cases of less and poor quality supply (column 4), and (albeit only marginally and insignificantly) excessive rates (column 5).

This decrease in inadequate material supply, in combination with increased sanctioned material expenditure, suggests another pathway. Over the years, the government has put an increasing focus on the creation of *durable* assets. Social audits in Sikkim take this into account through quality checks of works and material, and by letting beneficiaries rank assets in terms of

[28] As the NREGA guidelines lay down the ratios for different project types, one can roughly identify the main NREGA project categories in the data which should have a particularly high material cost share.

usefulness and durability (RMDD et al. 2015). Accordingly, detected deficiencies in these areas may increase sanctioned material expenditure in the following years due to a higher focus on sufficient and better quality material.

Finally, most irregularities in material expenditure can be detected by external auditors without cooperation from beneficiaries, while this is not the case for irregularities in labor expenditure. If beneficiary participation in public meetings is weak, this likely lowers the information gathered on and public sanctions for embezzled wages to a greater extent than for material expenditure irregularities. The results in Table A3.14 support this mechanism: greater meeting participation is associated with a small significant increase in labor expenditure which surpasses the corresponding increase in material expenditure.

To sum up, it is unlikely that the disproportionate increase in material compared to labor expenditure is due to a shift towards more material-intensive project types, or towards material- related corruption. Two potential mechanisms are that projects budgets are raised to buy expensive but better quality material, and that detection of labor irregularities requires greater beneficiary cooperation than detection of material irregularities.

3.5.4 Cost-effectiveness

Waddington et al. (2019) criticize that comparability of different community monitoring approaches is hampered by the fact that only four out of 35 studies on citizen engagement for public service delivery, evaluated in their systematic review, provided information on the cost or cost effectiveness of interventions. Olken (2007) estimates cost-effectiveness only for the pure government audit component, which underestimates for instance time costs of public meeting participation in social audits. Therefore, I conduct two types of rough cost-effectiveness estimates: costs per change in selected outcomes and the net monetary benefit of the program. For the first, I estimate the costs per change in those outcomes for which I find a significant and clear beneficial effect in my main analysis of the full seven years. I focus here on the direct costs of conducting a social audit. These amount in Sikkim to ₹20,650 per audited GPU (or ₹8 per capita and ₹38.42 per household, respectively). Given the observed effects of a social audit in Sikkim, the cost of decreasing total issues p.c. in Sikkim by one amounts to ₹0.19, while increasing material expenditure p.c. through social audits by one ₹ costs ₹0.1.[29]

For the second cost-effectiveness estimation, loosely based on the methodology used by Olken (2007), I weigh monetary benefits and costs of social audits against each other by calculating the net benefit of conducting an additional (second, third or fourth) social audit. These are presented in Table 3.14, which contains three different cost-effectiveness estimates: first, the

[29] I refrain from a direct comparison of these cost-effectiveness estimates to those of Afridi and Iversen (2014) since there are no comparable significant effects in the two analyses. The only outcomes with a significant effect in both analyses are the number of total and material irregularities (see appendix A3.1 for the respective cost-effectiveness estimates for user-conducted audits). However, the ambiguous interpretation of positive and negative irregularities makes a meaningful comparison difficult.

simple net benefit, which is equal to the funds gained by deterring embezzlement through an additional audit minus direct operational costs. This amounts to a positive net benefit of ₹9.71 per capita, or 122 percent of direct costs. It thus lies slightly below the net benefit of a pure government audit in Olken (2007). Second, the net social benefit, which in addition takes into account the marginal cost of levying taxes in India (based on an estimate by Ahmad and Stern (1987)), as well as the average opportunity costs of time (measured by a day's foregone NREGA wage) of the beneficiaries who attend the public social audit meeting. This results in a negative net social benefit close to zero (₹-0.64 per capita), which would imply that additional social audits just about fail to reach social cost-effectiveness. This estimate, however, ignores that social audits might also increase the recovery rate of embezzled funds. In my sample, an additional audit increases the recovery rate by ten percent, but as the sample for the recovery rate is strongly reduced to only observations with embezzlement, this effect size is too small to be statistically significant. If, in a third step, one makes the assumption that additional audits also increase the recovery rate, the net social benefits become positive[30].

The adjusted benefits of social audits in Sikkim make up 296 percent of the social costs. Compared to the negative net social benefit from Olken (2007) for the Indonesian roadbuilding program, social audits in Sikkim thus perform extremely well. What is more, there are several reasons why the overall cost-effectiveness of social audits in Sikkim may be even higher than the above estimates: firstly, it should be noted that the results are very sensitive to the time costs of beneficiaries (which vary between half a day and a full day). When assuming half a day instead of the average 75 percent work day, the resulting net social benefit of additional audits with 532 percent of social costs almost doubles. The strong influence of time costs on the net benefit also has the interesting implication that if one assumes decreasing returns to participation, there might be a tipping point where the marginal cost of participation in social audits exceeds the marginal benefits induced by a higher participation. That said, Table A3.13 shows that raising the percentage of public meeting participants in the population by one percentage point may stimulate an increase in labor expenditure of ₹20.27 per person. This additional wage income surpasses the additional time costs of ₹1.31 of such a change by far. Second, a more extensive cost-effectiveness analysis could factor in the additional benefit from superior-quality assets created by the deterrence of embezzlement of material funds (Olken 2007). These may well exceed the pure costs of funds, as many assets (such as roads) are used by a multitude of individuals. As my dataset contains no disaggregated embezzlement data, I cannot take such quality benefits into account. Lastly and most importantly, embezzlement data is only available for years in which social audits took place, so that my estimates ignore the effect of the first social audit on embezzlement.

---

[30] Alternatively, zero cost-effectiveness is reached when public meeting participation decreases from 4.3 to 3.8 percent of the population or by lowering the public meeting length to a 67 percent day.

Table 3.14: Cost-effectiveness of social audits

	External social audit Sikkim	Pure external audit (Olken 2007)
Monetary benefits		
Decrease in embezzled amount	₹17.70	
Change in amount of recovered funds	₹0-3.71	
Costs		
Average direct costs of SA	₹-7.99	
Deadweight loss due to levying taxes for SA	₹-4.71	
Average time costs of Jan Sunwai participation (75% day NREGA wage, for 4.3% of population)	₹-8	
Cost effectiveness		
Net benefit (assuming constant recovery rate of funds)	₹9.71	$0.06
As percent of costs	*122%*	*140%*
Net social benefit (assuming constant recovery rate of funds)	₹-0.64 [₹-2.52; ₹1.24]	
Net social benefit (assuming increased recovery rate of funds)	₹3.07 ($0.05) [₹1.19 ($0.02); ₹4.95 ($008)]	-$0.02
As percent of social costs	*296% [104%; 532%]*	*93.6%*

Notes: Benefits and costs reported per capita. Costs listed as negative numbers. Decrease in embezzlement is the estimated decrease in the detected embezzled amount through social audits in rupees (own estimation, see Table A3.17). Change in recovered amount is calculated as mean amount embezzled detected in first audit times the estimated change in share recovered. Net social benefit calculated from Olken (2007) refers to net (social) benefit of reduction in corruption compared to (social) cost of treatment when raising the audit probability from 4 to 100 percent (external audits plus public meeting with low beneficiary attendance and eliciting information from beneficiaries). For comparison, it is adjusted for inflation and converted from village to per capita level. NREGA wage is the average daily unskilled NREGA wage of ₹174.78 between 2011/12 and 2018/19. Time costs are based on the share of GPU population participating in public social audit meetings in Table 3.1. Upper and lower bounds of social benefits using maximum (one day) and minimum (half a day) public meeting length in squared brackets.

3.5.5 External vs. user-conducted approach and external validity

Comparison of external to user-conducted approach

Does Sikkim's external social audit approach yield results different from those of the user-conducted approach? My results for external social audits are consonant with insignificant effects on total expenditure and employment described by Afridi and Iversen (2014) for user-conducted social audits in the same program. However, user-conducted social audits experience an (insignificant) *increase* in person days worked, which suggests that compared to external social audits, beneficiaries in user-conducted social audits may be slightly better informed and engaged in the monitoring due to their involvement in the social audit implementation and evidence collection. Notwithstanding this, my finding in section 3.5.3 gives hope that labor outcomes might be improved even within external social audits by raising the participation in public meetings. Furthermore, the results regarding work projects in Sikkim indicate that external social audits can in the long term play an important role in improving the quality of works, an aspect not yet investigated for user-conducted social audits in NREGA. Finally, the direct costs of conducting external social audits are reportedly lower than for user-conducted audits (Tambe et al. 2016a) and even social cost effectiveness is likely higher in the external design compared to the user-conducted design: social costs hinge strongly on the time costs of the social audit participants, which are by definition higher in the external design. Considering that comparable effects (including days worked in NREGA) are small and insignificant in both designs, the clear advantages of external social audits regarding costs and potentially work quality may very well offset a possible poorer performance in terms of employment.

External validity

The results obtained in this paper are representative for the complete state of Sikkim since the analysis incorporates almost all existing GPUs in the state. To which extent can conclusions from these results be transferred beyond that, to other regions of India, or even other countries and programs? India is a marvelously diverse country, with a vast number of cultures and languages. Sikkim's culture is likely different from those of South Indian states. However, with eleven official languages, tribal and non-tribal, predominantly Hindu but also Buddhist population, Sikkim is, in a way, a miniature representation of Indian diversity. Moreover, weighting the regressions by population ensures that the results are not driven by very small outlier villages and could thus be recreated in other areas with slightly larger administrative entities. In any case, NREGA is a strictly rural program, so no large discrepancies in the size of administrative entities are to be expected. Indeed, the set-up of NREGA and the political institutions which implement it (such as the Gram Panchayat system) are by design largely identical across India.

If my estimates yield similar results to evaluations of other, similar interventions in different settings, this supports the assumption that my results are not spurious (internally valid) and applicable to other contexts (externally valid) (Banerjee et al. 2015). As established above, my results are similar to those of user-conducted social audits in NREGA except for potentially better labor outcomes (Afridi and Iversen 2014). Broadly in line with results of Molina (2013) and Gordon et al. (2019) for user-conducted social audits of other programs, I find positive effects on provision of at least some services. Finally, the improvement in works-related outcomes and embezzlement complements Olken (2007)'s findings that external social audits reduce missing funds in the provision of public road infrastructure. In addition, the differential influence of higher meeting participation on labor and material expenditure corroborate similar findings of Olken (2007) for missing funds.

3.6 Policy implications and concluding remarks

The findings of this paper are relevant both to academia and policymakers: firstly, my findings add to the limited evidence on the effects of social audits in general; second, I expand the literature on the effects of external social audits to program outcomes such as public asset and service generation and use, and third, I evaluate an external social audit over an extended period of seven years.

I find that external social audits only partially live up to the high expectations imposed on them. On the one hand, they are a cost-effective intervention which, in its current form has positive effects on the efficient creation of infrastructure assets, for instance by eliminating corruption in the form of ghost works. This effect is likely sustainable, as auditors adapt the auditing process to changing practices of bureaucrats to hide corruption. On the other hand, external social audits do not on average increase the number of completed infrastructure assets or improve employment outcomes. Especially the latter is a notable drawback compared to user-conducted social audits, for which existing literature indicates that they may be marginally more successful at upholding demand for employment. Consistent with the previous theoretical and empirical literature, this supports the hypothesis that cooperation of beneficiaries may be important when it comes to the evaluation of certain irregularities about which they are particularly well informed or concerned. In line with this, my analysis provides some suggestive evidence that the external social audit approach has the potential to raise labor expenditure if participation in the public social audit meetings was further increased. The additional wage expenditure generated to workers by such higher participation would surpass the time costs of participation by far.

The above observations are important for policy makers and researchers for several reasons. Most importantly, they highlight the importance to update the audit process to new forms of corruption. Second, outreach to disadvantaged population groups should be paid particular attention to, and the mechanism between social audits and a decrease in program enrollment needs to be investigated further. Does the program for instance become less attractive due to greater administrative hurdles or because of reduced opportunities of collusion between workers and implementers? Finally, my findings indicate that it would be worthwhile to causally investigate the nexus between public meeting participation and program outcomes, including whether returns to participation are decreasing. The high sensitivity of cost-effectiveness to time costs of participation suggest that a critical mass of participants may be sufficient, and striving for full participation in public hearings or greater involvement of beneficiaries beyond those hearings, such as in user-conducted audits, may be counter-productive. Presumably because NREGA officially promotes the user-conducted design (Karuna et al. 2019), only one other Indian state, Assam, has unofficially adopted the external social audit design to date[31]. However, my results show that policy makers face a trade-off between potentially slightly better labor outcomes (user-conducted approach) and cost-effectiveness and sustainability through

[31] Assam formed an SAU in 2016. While villagers have been selected and trained officially to conduct social audits they are not involved in the audits in practice.

the long-term learning of auditors (external approach). Given that comparable effects on employment and other outcomes differ only marginally between the two designs, the marked advantages of external social audits might well outweigh its poorer effects on labor outcomes.

A.3 Appendix

A3.1 Further methodology

Empirical approach for estimates with participation

$$y_{it} = \alpha_i + \gamma_t + \beta_1 AuditStep_{it-1} + \beta_2 ParticipStep_{it-1} + \rho X_{it} + \varepsilon_{it} \quad (2)$$

Where i indexes GPU and t financial year. y_{it} represents an NREGA program outcome of GPU i in year t, divided by population, $AuditStep_{it-1}$ indicates how many audits have been completed by the previous period (t-1), α_i are GPU fixed effects and γ_t are year fixed effects. ε_{it} is the error term. Standard errors are clustered at the GPU level. X_{it} controls for the total number of works (only relevant when the outcome is the number of work projects). $ParticipStep_{it-1}$ indicates the cumulative percentage of participants in social audit public meetings up to the previous year.

Empirical approach for estimates by round

Detected irregularities

I estimate the effect of the second, third and fourth SA on the number of irregularities detected (compared to the first audit), using a simple difference methodology for estimation. The regression equation is

$$Y_{ikt} = \alpha_k Round_{ikt} + \delta_i + \varepsilon_{ikt} \quad (4.3)$$

where t indexes year, i Gram Panchayat Unit, k the social audit number (between 2 and 4, thus excluding the first audit). The outcome, Y_{ikt}, is the number of irregularities per GPU population reported in the kth audit for a given GPU and year. $Round_{ikt}$ is a dummy variable indicating whether social audit round k took place in GPU g in year t. δ_i are GPU fixed effects. α_k represents the effect of the kth audit relative to the first audit and ε_{ikt} is the error term. Standard errors are clustered at the GPU level.

NREGA outcomes

The social audit round dummies now continue to take on the value of one in all years following the completion of the respective audit round. This captures the fact that the effects of audits likely accumulate over time and allows to directly see the additional benefit from an additional audit compared to the previous audit. The corresponding regression equation is as follows:

$$y_{it} = \alpha_i + \gamma_t + \sum_{k=1}^{4} \beta_k D_{it}^k + \varepsilon_{ikt} \quad (4.4)$$

where i indexes GPU, k social audit number, t financial year, y_{it} represents an NREGA program outcome of GPU i in year t, divided by population, D_{it}^k indicates whether audit k occurred in a previous period (t-1 or earlier), α_i are GPU fixed effects and γ_t are year fixed effects. ε_{ikt} is the error term. Standard errors are clustered at the GPU level.

Methodology for cost effectiveness calculation in Table 3.14

Net benefit = reduction in amount embezzled per GPU inhabitant through SA

$$- \frac{\text{costs per SA}}{\text{average GPU population 2010−18}}$$

Net social benefit= net benefit - time costs - deadweight loss of levying tax

where time costs are equal to the average daily NREGA wage times the time spent at public meeting (in days), times the average participation rate in public meetings

Net social benefit (assuming significant change in amount recovered) = net social benefit + change in amount of recovered funds

where change in amount of recovered funds is equal to the change in share of recovered funds through SA times the mean amount embezzled at baseline.

A rough cost effectiveness calculation of the change in total irregularities for social audits in Andhra Pradesh using the average of the effects computed by Afridi and Iversen (2014) and costs per social audit per mandal of ₹250,000 (roughly ₹12,910.31 per GPU) reported by Karuna et al. (2019) indicates that *raising* total irregularities in one GPU in Andhra Pradesh by one costs ₹20,188.14.

A3.2 Additional tables

Table A3.1: Data restrictions and number of observations

Data restriction	Dataset (Observations)			
	SA irregularities/reports		NREGA outcomes	
	(Years*GPUs)	(GPUs)	(Years*GPUs)	(GPUs)
Unrestricted	581	176	1,246	178
Exclude reports with new format	414	176	-	-
Exclude GPUs which split or merged	419	172	1,176	172
Exclude GPUs for which no population weights are available	415	170	1,190	170

Note*:* The irregularities dataset comprises data from social audit (SA) reports for the financial years in which social audits took place between 2011/12 and 2016/17. The dataset on NREGA outcomes comprises a balanced sample spanning 2011/12 to 2017/18.

Table A3.2: Description of variables

Variable	**Scale**	**Description**	**Source**
Outcome variables			
Irregularities, total	Per 1000 GPU inhabitants	Irregularities in the NREGA program found during social audits. Corresponds to the sum of irregularities listed below in the categories labor-related, and work-s and material-related (does not consider irregularities found for social audits themselves). Irregularities excluded because they were not yet included in the early audits are: job card not in custody of job card holder but kept by mate, wage slip not given to worker, bank passbook not in custody of worker, unemployment allowance not paid, less than correct amount of unemployment allowance, contractor used in execution of work, labor displacing machines used in work project, no/inadequate facilities at the worksite, annual shelf of works not prepared in Gram Sabha, no grievance redressal within seven days of complaint,	SA reports, population: Census 2011

		records/information not provided to social audit team two weeks before the village meeting, supervisor nominated by district program coordinator is not present in the social audit meeting, and no proactive disclosure of information on the wall of the Gram Panchayat.	
Irregularities, labor-related	Per 1000 GPU inhabitants	Sum of irregularities found during social audits in the following categories: denial of registration, registration of bogus households, fee charged for registration, non-issuance of job cards (after registration), delay of job cards, fake job cards, fee charged for job card, absence of photo on job card, job card entries not updated/incoherent, non-acceptance of work application, wrong/no date on the work application, work not given on time, wage not paid, wage paid late, wage underpaid, wage paid to wrong person, wage paid in the name of non-existing (ghost) worker, unemployment allowance paid late, workers do not receive dated receipt for work application, 33% women quota not satisfied in work allotment, transport allowance not paid, payment of wages for non-existing (ghost) projects.	SA reports, population: Census 2011
Irregularities, works- and material-related	Per 1000 GPU inhabitants	Sum of irregularities found during social audits in the following categories: selection of work not based on village council recommendation, priority of works not maintained, poor selection of worksite, exaggerated or inaccurate technical estimate, inclusion of unnecessary expenditure in estimate, excessive rates and material, recording of nonexistent (ghost) workers, recording of nonexistent (ghost) works, work not conforming to work specifications/prescribed standard, supply of less than sanctioned/poor quality materials and tools, shelf of projects not prepared in the Gram Sabha (village council meeting), no citizen information board at the worksite (which should display the sanctioned amount, work	SA reports, population: Census 2011

		dimensions etc.), muster rolls not publicly available at the worksite, worksite material register not properly maintained, daily individual measurement of work not conducted, final measurement of the work not done by the junior engineer, vigilance committee did not make regular worksite visits, complaints made were not addressed within seven days, measurement book not maintained properly, taking and/or recording of improper measurement, issuing of false/late completion certificates, work not conforming to specifications/standards, data recorded in confusing or incomplete manner, photographs taken before, during and after the work are not available for public display and scrutiny during the Social Audit.	
Irregularities, embezzlement cases	Per 1000 GPU inhabitants	No. of cases of mis-utilization of funds.	SA reports, population: Census 2011
Expenditure, total	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure of a GPU in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011
Expenditure, material	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure disbursed on material in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011
Expenditure, Labor	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure disbursed on labor in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011,
Expenditure, percentage of labor	Percent	Percentage of labor expenditure in total expenditure. Should be above 60% according to NREGA guidelines.	NREGA website, population: Census 2011
Job cards, applied for	Per GPU household	Number of NREGA job cards applied for.	NREGA website, population: Census 2011
Job cards, issued	Per GPU household	Number of NREGA job cards issued.	NREGA website,

			population: Census 2011
Person days worked in NREGA	Per GPU inhabitant	Number of person days worked in NREGA.	NREGA website, population: Census 2011
Works, total	Per 1,000 GPU inhabitants	Total number of NREGA works.	NREGA website, population: Census 2011
Works, ongoing	Per 1,000 GPU Inhabitants	Number of ongoing NREGA works (works in which some activity/expenditure is taking place).	NREGA website, population: Census 2011
Works, suspended	Per 1,000 GPU Inhabitants	Number of suspended NREGA works (previously ongoing work in which no expenditure/activity is happening but which is not yet closed. Work can be (but is not always) resumed at a later point in time.	NREGA website, population: Census 2011
Works, completed	Per 1,000 GPU inhabitants	Number of completed NREGA works (works on which all activities have been completed and which are thus declared closed or partially closed, meaning all necessary closure documentation has been provided).	NREGA website, population: Census 2011
Main explanatory variables			
Social audit round		Number of social audit rounds conducted, including the audit round for which the respective SA report was written.	Ministry of Rural Development
Social audit round step variable (t-1)		Indicates the number of social audits that a GPU has been subject to until (and including) the previous year.	Ministry of Rural Development
Additional variables for exploration of pathways			
Embezzlement, percent recovered		Percent of funds found in social audits to be embezzled which has been returned	Computed from social audit reports
Irregularities, less/poorer quality material	Per 1000 GPU inhabitants	Number of detected cases of less or poorer quality material supplied than paid for;	Social audit reports
Irregularities, excessive rates/material/	Per 1000 GPU inhabitants	Number of detected cases of excessive rates paid, excessive amounts of material paid for, or technical estimate exaggerated.	Social audit reports

technical estimate			
Person days worked by SC/ST (percent)	Percent	Percentage of person days worked in NREGA by members of scheduled castes and tribes in total person days.	NREGA website, population: Census 2011
Person days worked by women (percent)	Percent	Percentage of person days worked in NREGA by women in total person days.	NREGA website, population: Census 2011
Works, approved	Per 1,000 GPU inhabitants	Number of technically and administratively approved NREGA works (on which work/expenditure has not yet been started)	NREGA website, population: Census 2011
Expenditure, material, disbursed as percent of sanctioned	Rupees (₹)	Real *disbursed* NREGA program expenditure on material, as percent of total (material and labor) sanctioned expenditure	Calculated from indicators from NREGA website
Expenditure, total sanctioned, on material intensive works	Rupees (₹)	Real *sanctioned* NREGA program expenditure on material intensive works. These are expenditure categories which include types of works for which NREGA guidelines assume a low labor-material ratio. These are sanitation, rural connectivity and "other" works (the latter comprises material intensive agricultural works). Base year for consumer price index is 2010.	NREGA website, population: Census 2011
Participation (alternative explanatory variable)	Per GPU inhabitant	Variable equals the participation in the last social audit until (and including) the previous year. Coded as zero if no social audit has yet been carried out in the respective GPU.	Social audit reports

Note: population based on Census 2011 and adjusted for rural population growth.

Table A3.3: Summary statistics of NREGA outcomes, placebo sample

	Mean	SD	N (Years*GPUs)
NREGA outcomes			
Expenditure			
Total real disbursed expenditure (labor & material) (₹)	1082.14	948.05	328
Real disbursed material expenditure (₹)	394.70	511.71	328
Real disbursed labor expenditure (₹)	687.44	595.46	328
Labor expenditure in total expenditure (percent)	67.25	25.93	300
Employment			
Job cards issued	0.84	0.40	328
Person days	7.24	5.15	328
Work projects			
Ongoing	6.15	4.73	328
Suspended	0.37	2.09	328
Completed	3.41	4.32	328
Main explanatory variable			
Completed social audit rounds (t+1)	0.28	0.45	328

Notes: NREGA program expenditure (₹, 2010 prices) per GPU inhabitant, job cards per GPU household, person days per GPU inhabitant, works per thousand inhabitants. Differences in observations in outcomes representing percentages arise whenever the underlying denominator equals zero. Sample: 170 GPUs weighted by population frequency weights derived from census 2011, years 2011/12 to 2017/18.

Table A3.4: Irregularities detected during social audits, controlling for year fixed effects

	Total	Labor	Work quality/material	Embezzlement
	(1)	(2)	(3)	(5)
SA round	-95.89	-97.26	1.37	-0.94
	(80.01)	(77.86)	(6.49)	(0.68)
		{0.64}	{0.86}	
Year fixed effects	Yes	Yes	Yes	Yes
Control mean	91.5	56.8	34.6	1.5
Observations (Audits*GPUs)	398	398	398	398
Clusters (GPUs)	170	170	170	170

Notes: See Table 3.3. Estimates computed using year fixed effects.
* p<0.10 ** p<0.05 *** p<0.01

Table A3.5: Issues detected during social audits, unweighted

	Total		Labor		Work quality/material		Embezzlement	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SA round	-43.37***	-1.14	-27.52***	6.79	-15.85***	-7.92*	-0.79***	-0.52***
	(6.20)	(13.49)	(5.40)	(10.31)	(2.30)	(4.33)	(0.10)	(0.16)
			{0.01***}	{0.58}	{0.00***}	{0.31}		
Period fixed effects		Yes		Yes		Yes		Yes
Control mean (first audit)	91.5	91.5	56.8	56.8	34.6	34.6	1.5	1.5
Observations (Audits*GPUs)	398	398	398	398	398	398	398	398
Clusters (GPUs)	170	170	170	170	170	170	170	170

Notes: See Table 3.3. Observations not weighted by population.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.6: Material and labor expenditure, unweighted

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t-1)	41.81	132.12**	-90.31	-4.79
	(90.74)	(58.63)	(59.69)	(2.92)
		{0.06*}	{0.15}	
Control mean	1404.6	530.0	874.6	64.7
Observations (Years*GPUs)	1190	1190	1190	1156
Clusters (GPUs)	170	170	170	170

Notes: See Table 3.4. Observations not weighted by population.
* p<0.10 ** p<0.05 *** p<0.01

Table A3.7: Employment, unweighted

	Job cards applied	Job cards issued	Person days
	(1)	(2)	(3)
SA round complete (t-1)	-0.07	-0.06	-0.73
	(0.04)	(0.04)	(0.55)
		{0.28}	{0.28}
Control mean	1.0	0.9	8.9
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: See Table 3.5. Observations not weighted by population.
* p<0.10 ** p<0.05 *** p<0.01

Table A3.8: NREGA projects, unweighted

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA round complete (t-1)	-0.18	0.17**	-0.55
	(0.73)	(0.08)	(0.55)
	{0.81}	{0.14}	{0.46}
Control mean	6.8	0.3	4.3
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: See Table 3.6. Observations not weighted by population.
* p<0.10 ** p<0.05 *** p<0.01

Table A3.9: Effect of separate audit rounds on irregularities detected during social audits

	Total		Labor		Work quality/material		Embezzlement	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SA round 2	-67.81***	-78.81**	-49.84***	-64.22*	-17.44***	-14.59***	-0.86***	-0.88***
	(11.88)	(35.02)	(10.48)	(34.11)	(3.74)	(3.04)	(0.13)	(0.21)
			{0.00***}	{0.21}	{0.00***}	{0.14}		
SA round 3	-71.05***	-84.56*	-44.41***	-61.26	-26.64***	-23.30***	-1.06***	-1.08***
	(13.76)	(47.34)	(12.61)	(45.10)	(4.67)	(7.27)	(0.19)	(0.30)
			{0.05**}	{0.30}	{0.02**}	{0.18}		
SA round 4	-58.40***	-74.47	-33.14***	-53.19	-25.26***	-21.28***	-1.66***	-1.69***
	(9.17)	(47.56)	(7.77)	(45.73)	(3.21)	(5.68)	(0.29)	(0.44)
			{0.01**}	{0.36}	{0.00***}	{0.15}		
Round 2 vs. 3	0.78	0.74	0.59	0.83	0.06	0.11	0.13	0.17
Round 3 vs. 4	0.19	0.18	0.13	0.15	0.69	0.52	0.02	0.03
Control mean	93.2	93.2	58.2	58.2	35.0	35.0	1.5	1.5
Observations (Audits*GPUs)	387	387	387	387	387	387	387	387
Clusters (GPUs)	162	162	162	162	162	162	162	162

Notes: See Table 3.3. The three main explanatory variables are dummy variables equal to one if the nth social audit (SA) has been completed in any of the previous years and zero otherwise. Round 2(3) vs. 3(4) indicates the p-value for the difference between the respective rounds. Control mean indicates the mean outcome in the first social audit round. Sample: 162 GPUs which had at least two audits, weighted by population frequency weights derived from census 2011, years 2011/12-2016/17.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.10: Effect of separate audit rounds on material and labor expenditure

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA 1 complete (t-1)	-6.84	38.63	-45.46	-2.44
	(76.46)	(54.06)	(51.76)	(3.27)
		0.65	0.65	
SA 2 complete (t-1)	-63.81	50.94	-114.74	-3.11
	(101.84)	(80.47)	(79.11)	(6.90)
		0.55	0.33	
SA 3 complete (t-1)	19.44	144.93*	-125.50	-9.56*
	(133.24)	(79.24)	(110.07)	(5.50)
		0.20	0.29	
SA 4 complete (t-1)	404.84**	351.75***	53.09	-13.39**
	(179.07)	(122.17)	(119.72)	(5.97)
		0.06	0.68	
Control mean	1217.3	464.4	752.9	64.5
Observations (Years*GPUs)	1190	1190	1190	1156
Clusters (GPUs)	170	170	170	170

Notes: See Table 3.4. The four main explanatory variables are dummy variables equal to one if the nth social audit (SA) has been completed in any of the previous years and zero otherwise.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.11: Effect of separate audit rounds on employment

	Job cards applied	Job cards issued	Person-days
	(1)	(2)	(3)
SA 1 complete (t-1)	-0.05	-0.04	-0.24
	(0.03)	(0.03)	(0.47)
		{0.34}	{0.61}
SA 2 complete (t-1)	-0.12**	-0.12**	-1.25
	(0.06)	(0.06)	(0.78)
		{0.15}	{0.17}
SA 3 complete(t-1)	-0.14**	-0.14**	-0.67
	(0.05)	(0.05)	(0.76)
		{0.07*}	{0.42}
SA 4 complete (t-1)	-0.09*	-0.09*	0.37
	(0.05)	(0.05)	(0.83)
		{0.22}	{0.69}
Control mean	0.9	0.9	7.7
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: See Table 3.5. The four main explanatory variables are dummy variables equal to one if the nth social audit (SA) has been completed in any of the previous years and zero otherwise.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.12: Effect of separate audit rounds on NREGA projects

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA 1 complete (t-1)	0.36	0.13*	-0.91*
	(0.59)	(0.07)	(0.51)
		{0.18}	{0.52}
SA 2 complete (t-1)	0.29	0.20**	-1.58**
	(0.97)	(0.09)	(0.66)
		{0.11}	{0.12}
SA 3 complete(t-1)	2.58**	0.19**	-1.36*
	(1.03)	(0.09)	(0.79)
		{0.14}	{0.21}
SA 4 complete (t-1)	1.61	0.26**	-1.71
	(1.98)	(0.11)	(1.23)
		{0.12}	{0.22}
Control mean	5.9	0.2	3.8
Observations (Years*GPUs)	1190	1190	1190
Clusters (GPUs)	170	170	170

Notes: See Table 3.6. The four main explanatory variables are dummy variables equal to one if the nth social audit (SA) has been completed in any of the previous years and zero otherwise.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.13: Effect of social audit participation on detected irregularities

	Total		Labor		Work quality/material		Embezzlement	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Participation	-0.84	-0.18	0.40	0.94	-1.24	-1.13	-0.09**	-0.08**
	(3.47)	(3.29)	(3.36)	(3.27)	(1.12)	(1.02)	(0.04)	(0.04)
			{0.92}	{0.81}	{0.60}	{0.60}		
SA round	-38.31**	-13.72	-29.80	-9.59	-8.51*	-4.12	-0.25	-0.01
	(18.75)	(26.25)	(18.52)	(24.36)	(4.77)	(7.31)	(0.19)	(0.19)
			{0.31}	{0.87}	{0.31}	{0.87}		
GPU fixed effects	Yes		Yes		Yes		Yes	
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	91.5	91.5	56.8	56.8	34.6	34.6	1.5	1.5
Observations (Audits*GPUs)	398	398	398	398	398	398	398	398
Clusters (GPUs)	170	170	170	170	170	170	170	170

Notes: See Table 3.3. Additional explanatory variable is the cumulative number of participants per GPU population at public social audit meetings until the previous audit. Estimates with period fixed effects include a dummy for the period before and after 2015. Estimates controlling for the lagged outcome instead of GPU fixed effects are not reported as they would reduce the already small sample size too much.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A3.14: Effect of social audit participation on material and labor expenditure

	Total		Material		Labor		Percentage of labor in total expenditure	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Participation (t-1)	33.92***	84.19***	14.31**	37.57***	19.61***	54.48***	0.12	-0.13
	(10.71)	(9.26)	(6.99)	(5.37)	(6.51)	(5.98)	(0.19)	(0.30)
			{0.05**}	{0.00***}	{0.01**}	{0.00***}		
SA rounds completed (t-1)	-83.08	-138.43*	35.42	-23.00	-118.51**	-140.34***	-2.90	-3.96
	(77.95)	(79.25)	(47.01)	(46.82)	(58.15)	(51.65)	(2.45)	(3.32)
			{0.46}	{0.48}	{0.11}	{0.00***}		
GPU fixed effects	Yes		Yes		Yes		Yes	
Double lag		Yes		Yes		Yes		Yes
Control mean	1217.3	1430.1	464.4	584.4	752.9	845.7	64.5	59.5
Observations (Years*GPUs)	1190	850	1190	850	1190	850	1156	814
Clusters (GPUs)	170	170	170	170	170	170	170	169

Notes: See Table 3.4. Additional explanatory variable is the cumulative number of participants per GPU population at public social audit meetings until the previous audit. Columns 2, 4, 6 and 8 include a double lag of the outcome instead of GPU fixed effects. Differences in numbers of observations arise from missings the participation variable as well as missing double lags for the first two years. Multiple inference adjusted Romano-Wolf p-values (for columns 3 and 5, 4 and 6, respectively) in curly brackets.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A3.15: Effect of social audit participation on labor

	Job cards applied		Job cards issued		Person days	
	(1)	(2)	(3)	(4)	(5)	(6)
Participation (t-1)	0.00	0.02***	0.00	0.02***	0.09	0.42***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)
			{0.71}	{0.00***}	{0.31}	{0.00***}
SA rounds completed (t-1)	-0.08**	-0.09***	-0.08**	-0.08***	-0.65	-1.11***
	(0.04)	(0.02)	(0.04)	(0.02)	(0.52)	(0.02)
			{0.12}	{0.00***}	{0.24}	{0.00***}
GPU fixed effects	Yes		Yes		Yes	
Double lag		Yes		Yes		Yes
Control mean	0.9	0.9	0.9	0.9	7.7	8.2
Observations (Years*GPUs)	1190	850	1190	850	1190	850
Clusters (GPUs)	170	170	170	170	170	170

Notes: See Table 3.5. Additional explanatory variable is the cumulative number of participants per GPU population at public social audit meetings until the previous audit. The variable takes on a value of zero if no social audit has yet taken place until the previous year. Columns 2 and 4 include a double lag of the outcome instead of GPU fixed effects. Differences in numbers of observations arise from missings in the participation variable as well as missing double lags for the first two years. Multiple inference adjusted Romano-Wolf p-values (for columns 1 and 3, 2 and 4, respectively) in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

Table A3.16: Effect of social audit participation on NREGA projects

	Ongoing		Suspended		Completed	
	(1)	(2)	(3)	(4)	(5)	(6)
Participation (t-1)	-0.20*	-0.09	-0.04**	0.00	-0.05	0.02
	(0.11)	(0.12)	(0.02)	(0.00)	(0.08)	(0.08)
	{0.16}	{0.55}	{0.16}	{0.30}	{0.51}	{0.72}
SA rounds completed (t-1)	1.39**	0.49	0.29**	0.00	-0.94*	-0.69*
	(0.66)	(0.45)	(0.11)	(0.00)	(0.52)	(0.39)
	{0.09*}	{0.30}	{0.09*}	{0.72}	{0.09*}	{0.06*}
GPU fixed effects	Yes		Yes		Yes	
Double lag		Yes		Yes		Yes
Control mean	5.9	5.8	0.2	0.0	3.8	4.2
Observations (Years*GPUs)	1190	850	1190	850	1190	850
Clusters (GPUs)	170	170	170	170	170	170

Notes: See Table 3.6. Main explanatory variable is the cumulative number of participants per GPU population at public social audit meetings until the previous audit.. The variable takes on a value of zero if no social audit has yet taken place until the previous year. Columns 2, 4 and 6 control for a double lag of the outcome instead of GPU fixed effects. Differences in numbers of observations to the main specifications arise from missings the participation variable as well as missing double lags for the first two years. Multiple inference adjusted Romano-Wolf p-values (for columns 1, 3 and 5; 2, 4 and 6, respectively) in curly brackets.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table A3.17: Effect of social audit number on costs of embezzlement

	Embezzled amount		Percent recovered	
	(1)	(2)	(3)	(4)
SA rounds completed (t-1)	-17.70***	-18.54*	10.25	-1.80
	(4.50)	(10.08)	(8.03)	(18.76)
Period fixed effects		Yes		Yes
Control mean	37.1	37.1	72.8	72.8
Observations (Years*GPUs)	319	319	109	109
Clusters (GPUs)	158	158	90	90

Notes: Outcomes are embezzled amount per GPU population and percent of embezzled amount recovered. Differences in observations arise from the fact that recovered amount is only available for observations for which embezzlement occured. All other notes from Table 3.3 apply.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

4 The spatial spillover effects of social audits

Abstract

Social audits, a form of community monitoring during which the public program performance reported by providers is cross-checked against information from beneficiaries in public gatherings, can be an effective tool to combat corruption and inefficiencies in program implementation. Notwithstanding this, the true effects of social audits may be underestimated due to the presence of spatial spillover effects of social audits to neighboring communities. I investigate such spillover effects in the context of a large work program, the Indian National Rural Employment Guarantee Act (NREGA), using a panel fixed effects spatial lag of X (SLX) model. Consistent with the theory that social audits in neighboring communities influence program outcomes in a manner similar to that of having an audit in one's own community, I observe a significant increase in ongoing and suspended work projects, and a decrease in the percentage of labor expenditure in neighboring communities. However, I find no evidence that an existing direct, potentially adverse effect, a decline in program enrollment, spills over to neighboring communities.

4.1 Introduction

Social audits, a type of community monitoring in which official information is compared in public hearings to information gained from beneficiaries of public programs, are viewed as a remedy against both corruption and inefficiencies in the provision of public programs (Ringold et al. 2011; Grandvoinnet et al. 2015). Monitoring public programs through social audits has been shown to reduce leakages in public road building programs (Olken 2007), improve both nutrition-related outcomes (Gordon et al. 2019) and increase efficiency in the creation of public assets but also to have some unintended side effects such as audit avoidance by bureaucrats and decreased outreach (see chapter 3 of this dissertation). However, none of these studies investigate whether these effects of social audits spill over to neighboring localities. Neglecting such spillover effects may lead to biased estimates of the true effect of the policy (Miguel and Kremer 2004) and mask unintended effects of public policies (Cisneros and Kis-Katos 2022).

In the case of social audits, *beneficial* spillover effects of audited communities to nearby communities may be relevant for several reasons: a social audit may, first, lead potential offenders in neighboring municipalities to adjust their expectation of detection (Olken and Pande 2012; Avis et al. 2018; Berman et al. 2017) or, second, change local social norms equilibria (Angelucci and Di Maro 2016), both of which may discourage potential offenders to engage in irregular behavior. Third, news about social audits' effectiveness in neighboring communities may influence potential beneficiaries' trust in the public program's effectiveness, leading to their increased (or decreased) participation in the program (for instance, Bobonis and Finan (2009) find such neighborhood peer effects of a public program on school enrollment). Fourth, beneficiaries may learn on how to best spot and remove irregularities by observing other communities (Berman et al. 2017). Fifth, positive externalities can arise if communities profit from their neighbors' through social audits improved program outcomes (Angelucci and Di Maro 2016). In contrast, *adverse* spillover effects may arise if government officials and suppliers learn from audits in their neighboring community which forms of corruption and mismanagement are most prone to be detected, and adapt their future behavior to avoid detection (Cisneros and Kis-Katos 2022; Angelucci and Di Maro 2016). They may also initially over-estimate consequences of the audits and update their expectations and behavior after observing their neighbors (Cisneros and Kis-Katos 2022).

Knowledge about geographic spillovers can guide policy makers towards the optimal information policy, as well as placement and timing of social audits, in particular during phase-in. If spillover effects are overwhelmingly beneficial, they may be enhanced by improving information flow between the relevant stakeholders in neighboring communities. On the other hand, the detection of adverse spillover effects is the precondition for counter-action. For instance, if bureaucrats learn from bureaucrats in audited communities how to hide corruption, an implication could be to audit neighboring clusters of communities at the same time to give bureaucrats as little time as possible to communicate before being audited themselves. Despite the potential importance of spillover effects of social audits to neighboring communities, they have so far not been investigated.

The literature on development economics and policy evaluation has in the last decades seen a growing number of studies on spillover effects of public development programs on individuals or localities that are not directly targeted (Banerjee et al. 2017). Spillover effects of public development programs have been estimated for a range of interventions such as deworming (Miguel and Kremer 2004), unconditional (Haushofer and Shapiro 2016) and conditional cash transfers (Angelucci and Giorgi 2009; Bobonis and Finan 2009; Barrera-Osorio et al. 2011; Alix-Garcia et al. 2013), a public work program (Merfeld 2019) and agricultural insurance (Mobarak and Rosenzweig 2014). Most of these evaluations focus on cash transfer programs and evaluate spillover effects through close social networks (e.g. siblings and peers within families, schools or communities) rather than spatial proximity. However, Alix-Garcia et al. (2013) stress the importance of *spatial* spillover effects of public programs. Evidence for both beneficial and adverse spatial spillovers of public programs has been found for schooling and deforestation in a conditional cash transfer program (Alix-Garcia et al. 2013; Bobba and Gignoux 2019) and for district-level migration and wages for the public work program NREGA (National Rural Employment Program) (Merfeld 2019).

Existing evidence on spatial spillover effects of community monitoring policies other than social audits, such as information and citizen engagement, and of pure government audits targeting corruption indicates that spillover effects of these policies reinforce both beneficial and adverse direct policy effects (Gonzalez and Komisarow 2020; Avis et al. 2018; Cisneros and Kis-Katos 2022). Gonzalez and Komisarow (2020) investigate spillover effects of community monitoring on crime in US communities. In addition to reducing crime in treated blocks, citizen patrols decreased crime in unmonitored neighboring blocks but also led to a partial relocation of crime to blocks further away from the monitored areas. Berman et al. (2017) find that roads in Afghanistan which were monitored during construction by village volunteers deteriorated more slowly than unaudited roads, and the average effect was more pronounced if more villages along a road were treated. However, although the authors report that some auditors tried to involve the public (ranging from the dissemination of audit results during Friday prayers at the mosque in one village to bombing corrupt contractors' equipment in another) their community monitoring intervention does not systematically involve public hearings open to all beneficiaries and is thus missing a core element of social audits.

Regarding spillover effects of fiscal audits targeting corruption without direct community involvement, Avis et al. (2018) find that random fiscal audits in Brazil decrease corruption (although only in the presence of local media), and this effect is reinforced by spillovers from audited neighboring municipalities. Cisneros and Kis-Katos (2022) evaluate unintended effects of the same fiscal audits on deforestation. Not only do they find that audits increase deforestation in the three years following the audit, but this direct unintended effect is re-enforced by similar effects on neighboring municipalities which did not have a recent audit themselves.

The only existing study estimating spatial spillover effects in the context of social audits is Olken (2007). In order to *avoid* spillovers between communities, he randomizes his main intervention, raising the likelihood of fiscal audits with public meetings, at the sub-district level. He then estimates spillover effects on nearby sub-districts with a low audit probability to provide support for his randomization design. He finds no spillover effects from treated sub-districts to untreated neighboring sub-districts. However, Olken explicitly randomizes the main intervention at the sub-district level precisely because he expects spatial spillovers to occur only at the subjacent village level. In addition, the only source of spillover effects which he discusses is officials' perception of the probability to be audited. However, the involvement of beneficiaries and other citizens in social audit public hearings suggests that information transmission between other stakeholders in nearby communities may also play a role.

This paper is the first to investigate the presence of spatial spillover effects of social audits between nearby communities. I estimate spillover effects of social audits of the National Rural Employment Guarantee Act (NREGA)[32]. NREGA is a large-scale public work program which aims to provide temporary employment to the poor rural population and improve long-term livelihoods through the construction of infrastructure assets. I exploit the phase-in of its social audits in the Indian state of Sikkim over four years to estimate the effects of an additional social audit in nearby communities on several important program outcomes: program enrollment, employment, expenditure and work completion status. The estimation applies a fixed effects panel spatial lag of X (SLX) model with standard errors clustered at the nearest neighbor community. Communities (called Gram Panchayat Units or GPUs) are defined as nearest neighbors if they are closest in terms of spatial distance to the original GPU. The analysis is based on the balanced year-GPU panel dataset introduced in chapter 3 of this dissertation. The panel dataset uses information on year- and GPU-wise occurrence of social audits and combines it with official NREGA data on program outcomes in the period 2011/12 to 2018/19. For the estimation of spillover effects, the dataset is augmented with geospatial information on village location. Identifying villages belonging to the same GPU allows me to pinpoint the location of GPUs and thus to calculate the distance between GPUs.

My results show that spatial spillover effects of social audits to neighboring communities are reinforcing effects observed in the audited communities. In particular, social audits in neighboring communities significantly increase both the number of ongoing and suspended works at conventional significance levels. These results could be attributed to two possible mechanisms: reduction of irregularities in anticipatory obedience by bureaucrats who witness the consequences of being audited in their neighbors, or passing on of strategies to hide corruption by declaring works as ongoing. Further, I find evidence that social audits in neighboring communities significantly decrease the percentage of labor in NREGA expenditure, which could threaten NREGA's aim to act as security net for the unemployed poor. The estimates of spillover effects are robust to the non-inclusion of population weights but not to adjusting p-values conservatively for multiple inference. A notable result of my analysis is that a further potentially adverse direct effect of social audits, declining program enrollment, does

[32] Geographic spillovers of the NREGA program itself have been estimated by Merfeld (2019), who shows that casual wages and private-sector employment increase more for villages located near districts in which NREGA was present. However, Merfeld does not evaluate spillovers of social audits.

not spill over to neighboring communities. I attribute this to the weak NREGA-related informational network of villagers not yet involved in the program.

The primary contribution of this paper to the existing literature is that it is the first study to estimate spatial spillovers of social audits on neighboring communities. In this way, it addresses biases in the estimation of the overall program effect of social audits and shows that total effects, which take into account direct and spillover effects of social audits, can surpass the direct effect of social audits by up to 66 percent. These findings generate useful information for researchers and policy makers alike. First, by showing that the program's effects surpass its direct effects, which could lend further support to the program but has implications for how social audit interventions should be evaluated. For instance, evaluations using unaudited neighbors as counterfactuals will underestimate the effects of social audits. Second, by highlighting the absence of unintended spillover effects on enrollment but also the possible presence of adverse spillovers related to hiding of corruption. A further contribution of this paper is that it embeds the findings regarding spillover effects in a discussion of access to information of the relevant stakeholders in audited and neighboring communities.

Notwithstanding this, information transmission channels and whether changes in work status are due to hiding behavior of bureaucrats need to be investigated more rigorously in the future in order to accurately predict spillover effects beforehand and to take relevant measures in case of adverse spillover effects. I also argue that changes in the way NREGA is implemented, for example the introduction of joint worksites for several GPUs, can lead to changes in spillover effects, which calls for a constant re-evaluation of appropriate policy responses.

The remainder of the paper is structured as follows. Section 4.2 briefly describes the NREGA program and social audits in Sikkim and derives anticipated spillover effects. Section 4.3 describes the data. Section 4.4 sets forth the empirical strategy. Results are discussed in section 4.5 while section 4.6 concludes.

4.2 Background

4.2.1 The NREGA program

NREGA aims to improve a diverse set of outcomes, of which the two most important ones are employment and public and private asset generation for the poor. To serve the first and primary goal, NREGA offers a maximum of hundred days of employment in unskilled manual labor at the national minimum wage to each rural household in India. NREGA thus aims to provide a safety net for the poor through a focus on rural areas, in which 86 percent of India's poor live (Alkire et al. 2014), and through the specific work conditions: manual labor, minimum wage and a women's quota ensure that the program is relatively more attractive to the poor and disadvantaged population. The safety net is further strengthened through several provisions such as an unemployment allowance in case no work can be allocated to an NREGA applicant, a work injury or death allowance, and stipulations regarding the distance to the allocated work which ensure that also poor beneficiaries can access it without migrating. At the same time, by offering employment in the construction and extension of physical infrastructure, NREGA strives to improve rural infrastructure, in particular for the poor (Samuel and Srinivasan 2016). Nevertheless, as employment generation is the primary goal, a minimum of 60 percent of the program funds are to be spent on labor (MRD and GOI 2013).

The NREGA program was rolled out district-wise in three waves between 2006 and 2009 and is active in all Indian districts since 2009. Choice of projects and work allocation are performed by the local communities (GPUs). In many Indian states, the state level ('line') department responsible for NREGA may also implement its own NREGA projects which nearby GPUs then send workers to. However, a circumstance which simplifies the isolation of spillover effects between different GPUs in the Sikkimese context is that line departments in Sikkim do not implement their own NREGA work projects. Instead, GPUs implement the work projects in the field, assisted by district councils ('Zilla Panchayats') (Tambe et al. 2012). Because of this, NREGA in Sikkim comprises many micro level works rather than large higher level works (RMDD 2014). In some instances, GPUs decide to collaborate in NREGA projects with other state schemes which plan and operate at a higher level (e.g. blocks or clusters of GPUs) to create specific assets. In these collaborations, the unskilled manual labor is typically paid with NREGA funds while the collaborating agency pays or supplies the material (RDD and GOS 2019; RMDD 2014; MRD and GOI 2021a; Bela 2022). In the financial year 2014/15, 67 percent of funding of collaborative projects in Sikkim was contributed by the collaborating line departments (DRD and MRD 2014), but only 1.8 percent of ongoing and completed works were executed as collaborative projects, called 'works under convergence', in 2014/15 (MRD and GOI 2022a). This rose to 10 percent in 2017/18 (MRD and GOI 2022a). However, due to the decentralized nature of NREGA in Sikkim, the decision to collaborate, and choice of the respective collaborative project as an NREGA project is made individually by each GPU. As a result, until 2022, collaborations only took the form of small scale work projects at the GPU level (mainly construction of livelihood assets for the poor, such as housing, or of local public assets) (MRD and GOI 2022b). Large projects spanning several GPUs, and implemented by the higher level collaborating organization instead of by GPUs were only conducted after an agreement for collaboration with the Border Roads Organization in 2022 (MRD and GOI 2021a, 2022b).

As described in detail in chapter 3, NREGA offers bureaucrats and suppliers ample opportunity for corruption and mismanagement through manipulation of worker records and wages, collusion with material suppliers, bribes or discrimination in work allocation, nepotism in work selection, or sloppy work construction. As a response to this, the Indian government has been trying to institutionalize social audits of NREGA in all states (MRD and GOI 2011) and since 2009 requires all states to set up an independent social audit unit (SAU) charged with organizing social audits in the specific state. However, two states, Andhra Pradesh and Sikkim, pioneered in setting up their own social audit unit even before these stipulations came into force. The independence and long existence of their SAUs ensures the comparability and availability of long-term quality data on social audits. However, while Sikkim conducts social audits at the community (GPU) level, social audits in Andhra Pradesh are conducted at the sub-district (mandal) level. As Olken (2007) argues that spillover effects of social audits are much more likely to occur between communities than sub-districts, my analysis of spillover effects focuses on the state of Sikkim.

4.2.2 The social audit process in Sikkim

Social audits in Sikkim are carried out by a non-governmental organization (NGO), the Voluntary Health Association of Sikkim (VHAS). As Sikkim's SAU, VHAS' staff and finances are independent from the Rural Development Department and the GPUs which implement NREGA. Social audits in Sikkim always follow the same procedure (Karuna et al. 2019): apart from interviewing the beneficiaries about the amount and timing of wages they received, hours worked, availability of work, compliance with work regulations and corruption encountered (both related to material procurement and access to employment), the auditors also examine the quality and extent of the physical infrastructure created under NREGA. The gathered information is then compared to information from official NREGA documentation. Social audits in Sikkim are external in the sense that the relevant information to be verified in the public hearing is collected exclusively by district-level SAU personnel. Nevertheless, they are *social* audits because information is also collected by questioning beneficiaries and not merely examining official documents, and because all irregularities detected during the social audits are presented and verified at a public meeting (called *Jan Sunwai*) with all stakeholders present, including both program implementers and program beneficiaries. Since 2014, these public social audit meetings are only allowed to take place if they ensure a minimum attendance of 30 percent of NREGA beneficiaries. The public hearing closes with an agreement on how to resolve the discovered irregularities and potentially punish erring officials. The actions taken are then followed-up at the district level and in the next social audit in the GPU (RMDD et al. 2015). The current, statewide social audit design under the SAU was rolled out in a staggered fashion: each year, a subset of GPUs was randomly selected to be audited (personal communication with SAU Sikkim personnel, 2022). The first four GPUs were audited in 2011/2012, three in 2012/2013, 89 in 2013/14, 92 in 2014/15, and as of 2015/16 on, yearly audits of all 176 GPUs in Sikkim were conducted.

4.2.3 Information transmission channels

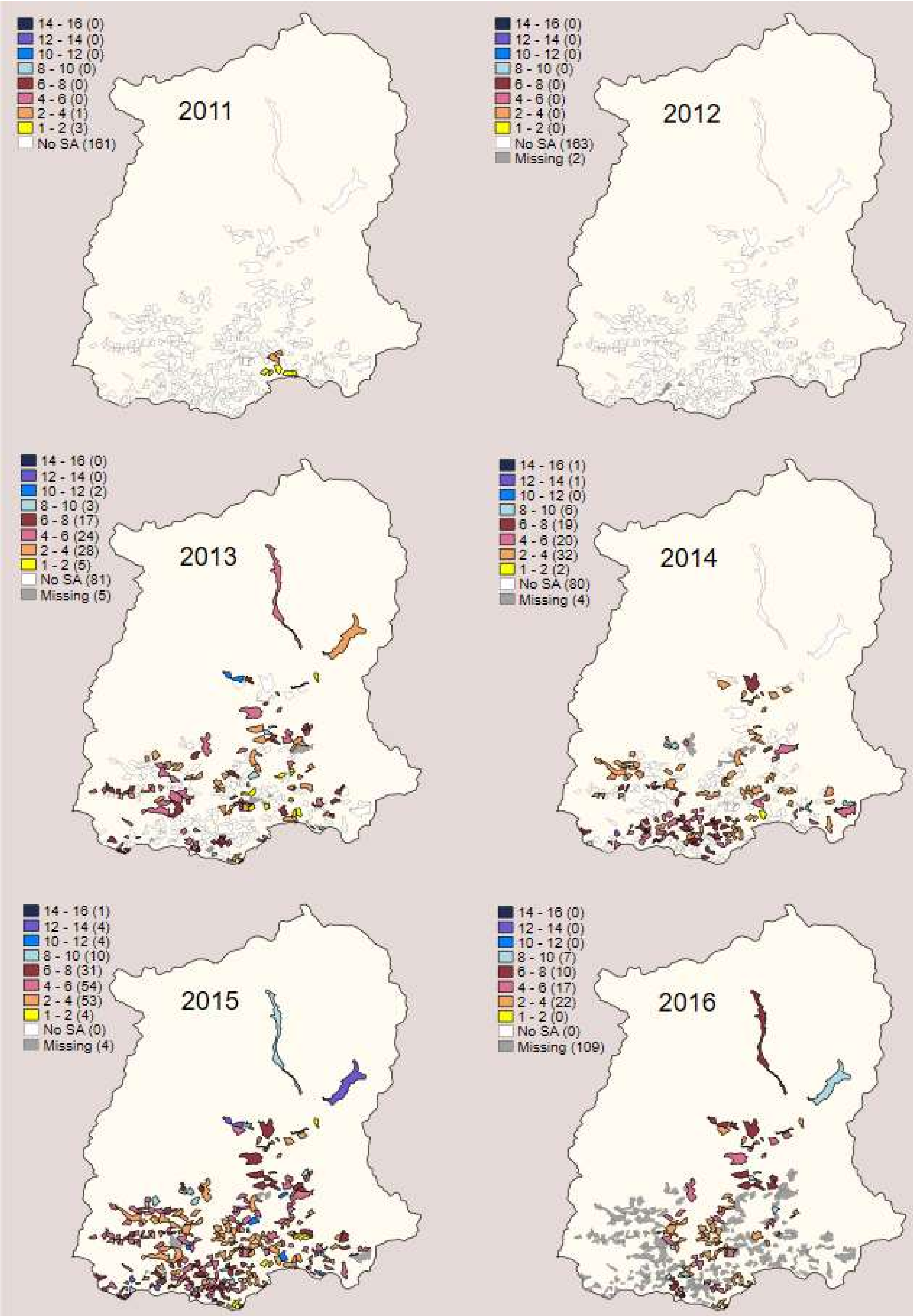
As laid out in the introduction, a major mechanism of spillover effects from audited to neighboring communities can be the transfer of information about incidence and content of social audits and NREGA entitlements between audited communities and their neighbors. This information could ultimately change incentives, expectations and local social norms equilibria in the neighboring communities (Olken and Pande 2012; Angelucci and Di Maro 2016). Information related to social audits can reach neighboring communities either through direct channels, such as the media (Avis et al. 2018), or through interactions with stakeholders in the audited community (Cisneros and Kis-Katos 2022; Angelucci and Di Maro 2016; Avis et al. 2018). For instance, Avis et al. (2018) show that partisan networks are an important channel of spillover effects of fiscal audits to neighboring communities. In the following, I provide some descriptive statistics which give an overview to which extent information transmission is possible at the GPU level in the setting of my analysis.

An important source of information about social audit results for stakeholders in the audited community are the social audit public hearings. Accordingly, dissemination of audit-related information likely increases with the share of the GPU population who attends these hearings. The participation in those hearings over the various years is mapped in Figure 4.1. As can be seen in Figure 4.1, participation per GPU inhabitant was initially low but increased to an acceptable level from 2013 on, once more GPUs received social audits.

Apart from this, communication infrastructure can facilitate information transmission within and, more importantly, between GPUs. Table 4.1 provides descriptive statistics of such communication infrastructure and interaction forums of GPUs. It is based on census data which was collected in February 2011, before the SAU conducted the first social audits (Office of the Registrar General and Census Commissioner 2011). Most GPUs have access to newspapers and either landline or mobile phone coverage. In addition, there exist forums in which villagers from different GPUs can interact: while only 29 percent of GPUs have an agricultural marketing society, 61 percent of villagers from GPUs without a society can easily reach one within a radius of 5 km. In addition, about half of the GPUs have either their own market or a market in close proximity to their GPU. Both nearby markets and society meetings provide opportunities for information exchange between villagers of neighboring GPUs. Notwithstanding this, half the GPUs do not have the opportunity to interact in nearby markets. Moreover, actual access of the poor to those interaction opportunities and information transmission channels may be limited. This is illustrated by the fact that interaction forums targeted exclusively to the poor, such as self-help groups, are scarce. These limitations may particularly hamper information exchange for potential and actual beneficiaries of NREGA. Moreover, in contrast to other Indian states, and to the setting in Avis et al. (2018), Sikkim did not have any community radio stations at the time, which have been shown to be particularly relevant for conveying local information (Rai 2020; Avis et al. 2018).

In addition, information transfer can occur through joint projects, in which the collaborative agency may link NREGA functionaries in different GPUs. In contrast, a connection of beneficiaries from neighboring GPUs through joint work projects is likely less strong in the context of Sikkim because, as has been discussed above, collaborative works do not involve a joint worksite for beneficiaries from different GPUs.

Figure 4.1: Map of participation in social audit hearings per GPU population by year



Note: Colors indicate participation in percent of GPU population. Number of GPUs in each category in brackets. SA indicates social audit. Based on social audit reports until financial year 2015/16. Missings before 2016 due to missing reports or missing participation information in reports.

Table 4.1.: Potential for information flow

	Mean	SD	N (GPUs)
Newspaper availability			
Newspaper selling point in GPU	77.07	42.04	164
Distance newspaper seller (if none in GPU) under 5 km	12.02	32.52	164
Distance newspaper seller (if none in GPU) 5 km or more	10.92	31.18	164
Phone connection			
No phone coverage	0.63	7.94	164
Either landline or mobile coverage (not both)	25.77	43.74	164
Landline and mobile coverage	73.59	44.08	164
Road connection			
Footpath	100.00	0.00	164
Tar road	71.70	45.05	164
Social networks and interaction opportunities outside NREGA			
Self-help group	1.48	12.08	164
Agricultural marketing society in GPU	28.61	45.19	164
Distance agricult. marketing society (if none in GPU) under 5 km	60.84	48.81	164
Distance agricult. marketing society (if none in GPU) 5 km or more	10.55	30.72	164
Regular market in GPU	5.91	23.58	164
Distance market (if none in GPU) under 5 km	40.22	49.03	164
Distance market (if none in GPU) 5 km or more	53.87	49.85	164

Notes: Data source is census 2011, all variables are defined in percent (of GPUs). Discrepancies in the number of observations (N) to regression sample arise from a missing observation in census data. SD indicates standard deviation.

4.2.4 Effects of social audits on neighboring communities

Based on spillover effects on neighborhoods or neighboring villages found for other community monitoring programs, I expect social audits in neighboring GPUs to reinforce the direct effect of having a social audit in one's own GPU (Gonzalez and Komisarow 2020; Avis et al. 2018; Cisneros and Kis-Katos 2022). While spillover effects could arise from the mere incidence of social audits in neighboring communities, for example through readjustment of the expected auditing probability by bureaucrats, the existing evaluations of similar policies also provide evidence for spillover effects through learning from the concrete experiences and outcomes of social audits in neighboring communities (Cisneros and Kis-Katos 2022; Avis et al. 2018). For instance, if audits change particular outcomes through a behavioral change in community j, actors in neighboring community i can learn about this behavior and adopt it, thus influencing outcomes in their own community in a similar way. Conversely, if actors in the audited community do not learn from audits themselves, they cannot pass the relevant information to stakeholders in other communities. A potential indicator of whether actors in the audited communities change their behavior is the observed direct effect of social audits on outcomes in audited communities in the following year. In chapter 3, significant direct effects of social audits in Sikkim are found on enrollment in the program through job cards (a drop in demand), material expenditure (an increase) and work projects (a shift away from completed to suspended and ongoing projects).

A second factor which may catalyze spillover effects on a particular outcome is whether the relevant information about social audit results and potential adaptation mechanisms reaches those actors in neighboring communities who can also act on it, i.e. who have the ability to directly influence this outcome (Cisneros and Kis-Katos 2022; Avis et al. 2018; Björkman Nyqvist et al. 2017). In the context of NREGA, Citizens who are not yet NREGA beneficiaries can directly influence the number of job cards applications (and through their applications the number of job cards issued). Beneficiaries can directly influence the number of days worked in NREGA and labor expenditure by applying for employment under NREGA, and by noticing and filing a complaint about irregularities in their wages. Bureaucrats in turn can influence most outcomes directly (work completion status, job cards issued and program expenditure), except for job card applications.

While there exists some evidence that NREGA program beneficiaries are generally better informed about the program than non-beneficiaries (Ravallion et al. 2015), due to data limitations, I cannot observe the information flow between villages and towards specific stakeholders. However, different results can be indicative of different information channels. Thus, to improve the understanding of the information channels behind my results, I make the following assumptions, based on which stakeholders wield a strong influence over specific outcomes, and on whether the information they receive from neighboring communities is potentially action-stimulating (roughly approximated by the presence of a significant direct effect on outcomes in the audited communities). Firstly, a spillover on job card applications, possibly in combination with job card issuance, likely points towards non-beneficiaries having high access to social-audit-related information in neighboring communities. Second, if I detect spillovers on days worked in NREGA and labor expenditure, then NREGA beneficiaries likely have comprehensive access to such information, although the *absence* of an effect on these outcomes with weak or no direct effects would most likely be due to the absence of actionable information rather than weak information transmission. Third, if I detect spillover effects on work completion, expenditure and job cards issued (if not driven by applications and demand for employment), bureaucrats are likely well informed about social-audit-related information. While this framework is a first step towards understanding the mechanisms behind my results, it cannot account for spillover effects occurring through channels which do not involve explicit information transmission, such as general equilibrium effects.

4.3 Data

For the estimation of spatial spillover effects of social audits, I combine two main data sources: first, a panel data set assembled in the second essay of this dissertation, based on data on NREGA program outcomes from 2011/12 to 2017/18 publicized by NREGA (MRD and GOI 2021b) and information gained from the Ministry of Rural Development regarding the years in which social audits were conducted in Sikkim between 2010/11 and 2016/17[33]. Second, for the estimation of spillover effects, I augment this dataset with geospatial information on village location based on village boundaries sourced from DataMeet (2021) which are based geolocations collected in Census 2011 and other government sources. I then associate villages with the respective GPU they are part of, based on the Census 2011 village directory, and calculate the midpoint between villages belonging to the same GPU. This allows me to estimate the distances between midpoints of GPUs and consequently determine the nearest neighbor GPU. I exclude GPUs which split during the period of analysis (see chapter 3 for a list of these GPUs). In addition, five GPUs could not be matched unequivocally to the geospatial dataset because of splits before the period of analysis, incorrect or missing geolocations, so that the final dataset encompasses 165 GPUs.[34] The sample size under the respective restrictions is provided in Table A4.1 in the appendix.

I focus on the main outcomes evaluated in the second essay of this dissertation in order to compare the main effects found in chapter 3 with those from my analysis. These main outcomes are a series of NREGA program outcomes (described in detail in the previous chapter and Table A4.2 in the appendix) which encompass program expenditure, indicators of employment and asset creation through NREGA. Program expenditure is measured through real total, material and labor expenditure, as well as the percentage of labor expenditure in total expenditure. Asset creation is captured by work project completion status (ongoing, suspended or completed). Finally, demand for and access to the program, as well as employment are measured by the number of job cards applied for and issued and person days worked, respectively. All variables are adjusted for GPU population based on population data from the Indian Census 2011 (Office of the Registrar General and Census Commissioner 2011).

The main explanatory variables are the number of social audits a GPU has completed up to the previous year (see chapter 3 for a detailed description), as well as its spatial lag, the number of social audits the nearest neighbor GPU has completed till the previous year. A nearest neighbor to GPU_i is defined as the GPU_j with the smallest distance between its own and GPU_i's midpoint. The spatial lag is thus essentially a count step variable for the number of social audit rounds which the closest GPU has completed in the previous year.

4.4 Empirical approach

4.4.1 Estimation of spillover effects of social audits

My empirical approach exploits variation in the introduction of social audits in different GPUs in Sikkim. The map in Figure 4.2 visualizes the introduction of social audits in Sikkim across years. As can be seen from the map, introduction was slow in the first two years but took up

[33] In contrast to chapter 3, I do not explore effects on detected irregularities in chapter 4 since the unbalanced irregularities dataset would lead to an incomplete set of neighbors.

[34] GPUs additionally excluded from the dataset are: Martam (incorrect location), Rongli Changeylakha (missing location), Gnathang and Kyongosla (split, but treated as one in the shapefiles), Pakyong (missing location).

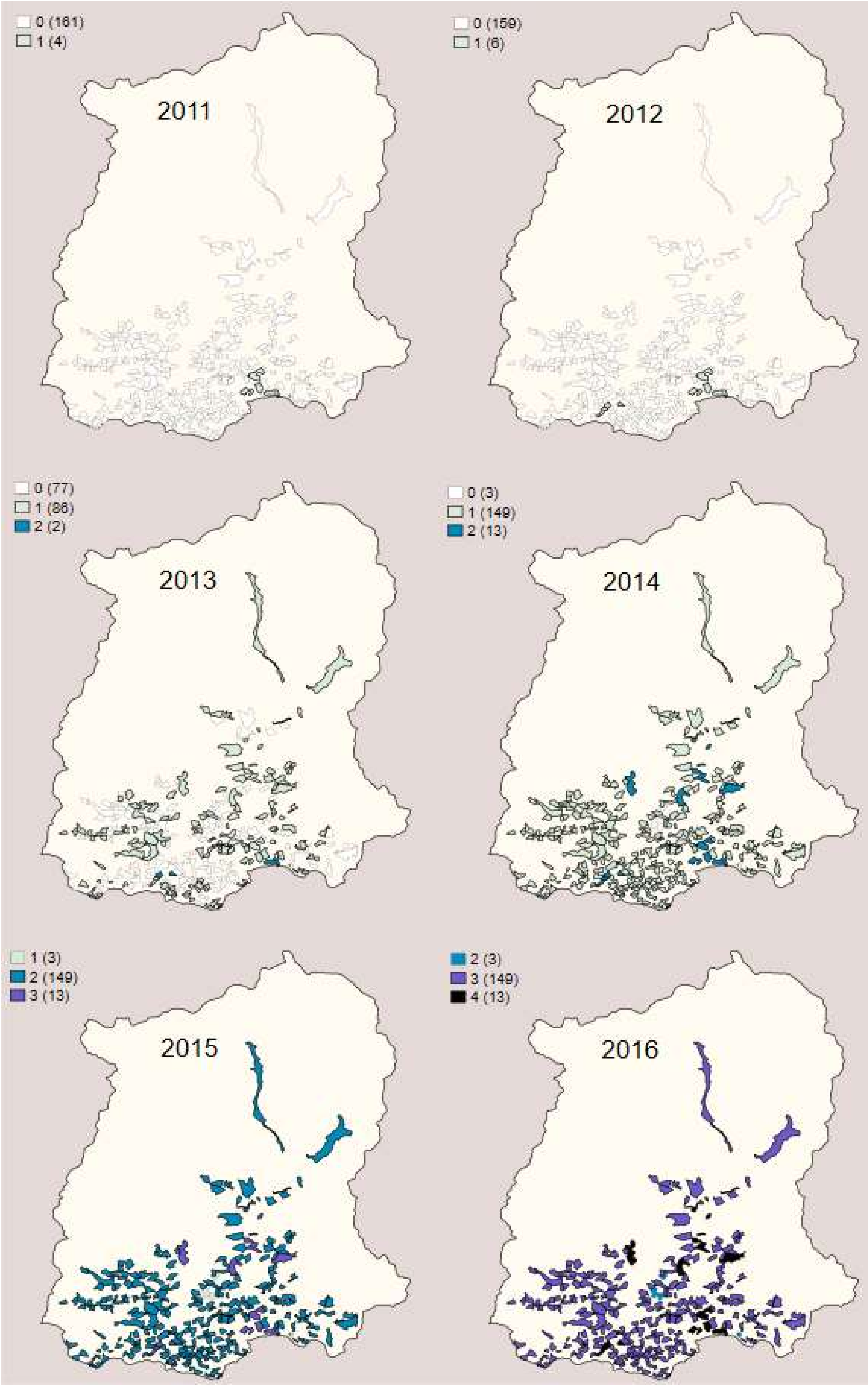
pace and was spaced out evenly across the state from the year 2013 on. From 2015/16 on, all GPUs were audited yearly. The explanatory variable for the direct effect of social audits is defined as in the previous chapter, as the number of social audits completed up to the previous year in a GPU. The main focus of this chapter lies in the estimation of the spillover effect from an additional social audit in a neighboring GPU. A GPU's outcomes could potentially be influenced by social audits of more than one GPU in its vicinity. However, in contrast to the existing studies on fiscal audits and community monitoring (Cisneros and Kis-Katos 2022; Avis et al. 2018; Gonzalez and Komisarow 2020), in the Sikkimese setting, a community's direct neighbors cannot be defined by a common community border because GPU boundaries do not necessarily touch. One approach would be to arbitrarily choose a particular distance as a 'neighborhood cut-off' and calculate the spillover effect from all audited GPUs within that distance (Gonzalez and Komisarow 2020). However, as distances between nearest neighbors vary greatly in my setting, this approach either leads to so-called island observations (observations without neighbors) for short cut-offs or to a very large distance cut-off. The first leads to insufficient power and potentially biases the results by excluding more remote nearest neighbors. The latter leads to a complete set of neighbors for all GPUs but potentially also includes a number of higher order (not only nearest) neighbors which have no influence at all because they are quite far away and GPUs are more likely oriented towards closer neighbors if they have any. To avoid these drawbacks, I apply a nearest neighbor approach, in which a GPU_j is defined as a nearest neighbor to GPU_i when its midpoint is closest to the midpoint of GPU_i in terms of geographic distance. While this approach can in principle also be extended to an arbitrary number of k nearest neighbors, I focus on the first nearest neighbor because its spillover effect is expected to be strongest. Although some of the higher order neighbors which I thus exclude may also exert some influence, Elhorst (2014) notes that punishment for a potential misspecification is weak: in case of weak spatial dependence, coefficient estimates will be close to the true estimates if the spatial weight matrix (a matrix which reflects which GPUs are defined as neighbors) is incorrectly specified, while in case of a strong spatial dependence, the probability that coefficients will be distorted is small. Notwithstanding this, I provide some estimates which evaluate the relevance of second nearest neighbors as a robustness test in section 4.5.1.

Based on the above reasoning, I employ a panel spatial lag of X (SLX) model (Elhorst 2014) to estimate the effect of an additional social audit on NREGA outcomes of the nearest neighbor GPU[35]. The spatial spillover effect is captured by a spatial lag of the explanatory variable, which indicates how many social audits took place in the nearest neighbor GPU until the previous period. The corresponding spatial regression equation is

$$y_{it} = \alpha_i + \gamma_t + \beta_1 AuditStep_{it-1} + \beta_2 AuditStep^{Neigh}{}_{it-1} + \rho X_{it} + \mu_{it} \tag{4.1}$$

[35] An alternative to the nearest neighbor approach would be to investigate the effect of distance to joint borders with units of the opposite treatment status as in Merfeld (2019). However, in contrast to the NREGA roll-out at district level (Merfeld 2019), social audits are implemented at GPU level. GPU borders do not necessarily touch, making a distance-to-border analysis impracticable.

Figure 4.2: Map of completed social audit rounds in each GPU by year



Notes: 165 GPUs. Colors indicate no. of audits. No. of GPUs in the respective category in brackets.

where i indexes GPU and t financial year. y_{it} represents an NREGA program outcome of GPU i in year t, $AuditStep_{it-1}$ indicates how many audits have been completed in GPU_i by the previous period (t-1). ${AuditStep^{Neigh}}_{it-}$ denotes the spatial lag of $AuditStep_{it-1}$. It is equal to the number of social audits completed up to the previous year by the nearest neighbor of GPU_i. The GPU and year fixed effects, α_i and γ_t, account for time-invariant heterogeneity between GPUs and GPU-invariant time-trends, respectively[36]. Regressions with work projects as outcome control in addition for the number of work projects, denoted by X_{it}. μ_{it} is the error term clustered at the nearest neighbor level[37]. This captures correlation of error terms within the GPU and its nearest neighbor (Cameron and Miller 2015; Bertrand et al. 2004). β_1 reflects the direct effect of having completed an additional social audit in the previous year, while β_2 captures the respective spillover effect of a social audit in a neighboring community. This effect may encompass both a direct spillover effect induced by the mere incidence of social audits in the neighbor community and an indirect spillover effect through changes in the neighbor's outcomes through those audits. For example, the former can occur through a reassessment of audit risk, changing incentives and norms. The latter can arise for instance if social audits improve infrastructure (e.g. roads and flood control) in the audited communities which could enhance income earning opportunities outside NREGA and thus decrease demand for employment in NREGA in their neighbor communities. Regressions with work projects as NREGA outcome control in addition for the total number of works.

The specification in equation (4.1) has several advantages: my narrow research question, which focuses on the spatial spillover effects induced by the explanatory variable social audits, allows me to refrain from explicitly including a spatial lag of the dependent variable as additional regressor. This reduces the risk of over-parametrization often encountered in General Nested Models (GNM), which simultaneously estimate spatial correlation in the error terms, dependent and independent variables (Elhorst 2014; Rüttenauer 2022; Gibbons and Overman 2012)[38]. Moreover, the inclusion of a spatially lagged outcome in a panel fixed-effects regression would entail endogeneity issues. While such endogeneity issues can be addressed through a maximum likelihood approach, the latter has the downside that instruments derived from maximum likelihood require strong and to some extent arbitrary assumptions about the data structure (Wooldridge 2010). Moreover, a maximum likelihood approach in combination with a panel fixed-effect regression requires a bias-correction which makes the estimates difficult to compare to existing, non-spatial panel fixed effects estimates of direct effects (Elhorst 2014). While unaccounted-for spatial dependence in outcomes can lead to omitted variable bias, I account for this by clustering the error terms at the neighborhood level. This accounts for any omitted

[36] I avoid a traditional event study design with an additional time lag of the outcome (Cisneros and Kis-Katos 2022) as this would (i) unnecessarily reduce the number of observations in my sample, owing to the fact that outcomes are not available before 2012 and (ii) exclude the years in which the majority of GPUs was yet unaudited.

[37] Implemented via the acreg networks command in stata (Colella et al. 2019).

[38] For the same reason, to avoid over-parametrization, I refrain from estimating heterogeneous effects which distinguish between the effect of having a social audit only in the neighboring GPU or both in one's own and one's neighboring GPU in the same year.

variable bias at the neighborhood level through spatially correlated explanatory variables not explicitly accounted for in the model, including omitted variable bias caused by spatial dependence in outcomes that is not caused by social audits (Cameron and Miller 2015; Zangger 2019).

4.4.2 Descriptive statistics and challenges to internal validity

Table 4.2 summarizes all outcome and explanatory variables for the regression sample and for the sub-sample of GPUs which are a nearest neighbor to another GPU. Expenditure per 100,000 GPU inhabitants equals ₹1367.17. 65 percent of this expenditure flows into labor costs. Almost 90 percent of households own a job card, and GPU inhabitants work on average eight days per year in NREGA. When taking the average over all observations (GPUs in years with and without audits), the average number of audits completed is one audit.

A potential concern in my analysis is that communities whose nearest neighbor received a greater number of audits are on different trajectories than communities whose nearest neighbor received less audits. As outlined in chapter 3, I cannot perform a balancing or parallel trends test for my full sample due to lack of data predating the introduction of social audits. However, I show in chapter 3 that a placebo test for the period 2011 to 2012 for the subsample of GPUs audited from 2013 on indicates that GPUs which received social audits earlier were on a similar trajectory as GPUs which received social audits at a later point in time. Such a subsample placebo test is not possible for the spatial estimates since excluding the GPUs audited in 2011 and 2012 would deprive some of the remaining GPUs of their nearest neighbor. However, if the parallel trends assumption holds in chapter 3, it likely also holds in the spatial regression for the following reasons: firstly, the spatial sample is almost identical to the one in chapter 3, as it includes all but five of the GPUs included in the analysis in chapter 3. Second, time trends of geographically very close communities are likely more similar (Gonzalez and Komisarow 2020). By restricting the analysis of spillover effects to those of the first nearest neighbor (and not neighbors which are even further away), I increase the likelihood that the characteristics of GPUs and their neighbor GPUs are similar to the sample in chapter 3 (and thus, that the results of the placebo test in chapter 3 hold also in the setting with neighbors). I underpin this by reporting summary statistics for nearest neighbor GPUs in Table 4.2 which shows that the subset of GPUs which are nearest neighbor to another GPU have on average similar socio-economic characteristics as the GPUs in the full sample.[39]

The fairly large number of outcomes evaluated in this paper increases the likelihood of falsely significant results. To account for this, I adjust the spatial estimates of the two explanatory variables of interest (the number of social audits in own GPU, and neighboring GPU, respectively) for multiple inference using the Benjamini and Hochberg (1995) method. This

[39] I do not provide p-values of differences between the two samples because the nearest neighbor GPUs form a subsample of all GPUs and the usual tests for differences are designed for mutually exclusive subsamples.

Table 4.2: Summary statistics NREGA outcomes

	All GPUs			Subsample: nearest neighbor GPUs		
	Mean	SD	N (Years*GPUs)	Mean	SD	N (Years*GPUs)
Expenditure						
Total real disbursed expenditure (labour & material) (₹)	1367.17	942.03	1155	1420.57	988.65	826
Real disbursed material expenditure (₹)	501.14	469.82	1155	518.72	489.72	826
Real disbursed labor expenditure (₹)	866.03	620.16	1155	901.85	646.65	826
Labor expenditure in total expenditure (percent)	64.72	20.56	1124	65.28	20.17	799
Employment						
Job cards applied for	0.89	0.34	1155	0.89	0.36	826
Job cards issued	0.88	0.34	1155	0.88	0.37	826
Person days	8.27	5.18	1155	8.58	5.47	826
Work projects						
Ongoing	7.71	7.20	1155	7.72	7.20	826
Suspended	0.10	1.12	1155	0.14	1.34	826
Completed	5.53	5.89	1155	5.87	6.21	826
Main explanatory variable						
Completed social audit rounds (t-1)	0.99	1.14	1155	0.98	1.14	826
Neighbor GPU's completed social audit rounds (t-1)	0.98	1.14	1155			

Notes: SD indicates standard deviation, N number of observations. NREGA program expenditure (₹, 2010 prices) per 100,000 GPU inhabitants (total expenditure includes labor and material expenditure), job cards per GPU household, person days per GPU inhabitant, works per 1000 inhabitants. Differences in observations in outcomes representing percentages arise whenever the underlying denominator equals zero. Full sample: 165 GPUs. Subsample of GPUs which are nearest neighbor to another GPU: 118 GPUs. Weighted by population frequency weights derived from census 2011, years 2011/12 to 2017/18.

method leads to fairly large adjusted p-values compared to other methods such as the bootstrapping methodology by Romano and Wolf (2005). In addition, Benjamini and Hochberg (1995) adjusted p-values increase greatly with the number of tests. This number is high in the present case since p-values are adjusted for two explanatory variables. However, the more efficient bootstrapping adjustments such as those by Romano and Wolf (2005) are not feasible in spatial regression because the random resampling leads to an incomplete set of neighbors.

4.5 Results

4.5.1 Spillover effects of social audits

In the second essay of this dissertation, I show that social audits can have important adverse and beneficial impacts on some NREGA outcomes, in particular those related to material and work projects. However, there remains the question whether these effects spill over to neighboring GPUs, or whether effects in neighboring GPUs may differ due to different expectations or lack of interaction and communication between bureaucrats or citizens. Both direct and spillover effects to neighboring GPUs are laid out in Tables 4.3 to 4.5. Overall, the results indeed point towards reinforcing effects of social audits in neighboring communities: most point estimates for spillover effects have the same sign as the direct effects (total expenditure being the only exception). However, only some of these spillover effects are statistically significant. In the following, I describe the results for the various outcomes in detail.

Table 4.3 presents the results of social audits in nearest neighbor communities on NREGA expenditure. Significant spillover effects are only found for the percentage of labor expenditure in total expenditure in column 4. Social audits in nearby communities decrease the percentage of labor in total expenditure by 4.4 percentage points (significant at the conventional 10 percent level). Spillover effects thus reinforce the direct effects of social audits on expenditure, with the total (direct and spillover) effect equaling an eight percentage point decrease in the percentage of labor in total expenditure. The subsequent Table 4.4 shows that social audits have no significant spillover effects on labor outcomes. In particular, despite a significant direct negative effect on job cards issued, the spillover effect for job cards is a precisely estimated zero effect. Finally, Table 4.5 displays spillover effects of social audits on work projects. Spillover effects are reinforcing the direct effects for all work projects. In particular, columns (1) and (2) show significant spillover effects on ongoing and suspended works. A social audit in a neighboring GPU significantly increases the number of ongoing works by 1.15 works per thousand inhabitants and increases the number of suspended works by 0.16 works, respectively. The estimates cease to be statistically significant when adjusting the p-values for multiple inference. However, considering that the adjustment by Benjamini-Hochberg is very conservative, the adjusted p-value is with 0.21 still comparatively low.

Table 4.3: NREGA expenditure

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t-1)	23.28	66.55	-43.27	-4.02
	(67.76)	(48.67)	(45.77)	(3.71)
		{0.37}	{0.67}	
Neighbor's SA round complete (t-1)	-31.91	17.94	-49.85	-4.35*
	(75.90)	(35.78)	(62.92)	(2.51)
		{0.67}	{0.67}	
Control mean	1233.9	466.8	767.1	64.8
Observations (Years*GPUs)	1155	1155	1155	1124
Clusters (Nearest neighbors)	118	118	118	118

Notes: Fixed effects panel regression. Outcomes are real disbursed NREGA program expenditure (₹, 2010 prices) per GPU inhabitant (columns 1-3) and percentage of labor in total expenditure (column 4). The main explanatory variable, neighbor's SA round complete (t-1), indicates the number of social audits in the nearest neighbor GPU completed up to the previous year. Control mean indicates the mean outcome if the main explanatory variable equals zero, i.e. no social audit took place in all previous years. Estimates computed using sampling weights, year and GPU fixed effects. Standard errors corrected for clustering at the nearest neighbor level in parenthesis. Sample: 165 GPUs weighted by population weights derived from census 2011, years 2011/12-2017/18. Multiple hypothesis adjusted standard errors by Benjamini-Hochberg for the two main coefficients in columns 2 and 3 in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

Table 4.4: Employment

	Job cards applied	Job cards issued	Person-days
	(1)	(2)	(3)
SA round complete (t-1)	-0.07***	-0.06***	-0.28
	(0.02)	(0.02)	(0.34)
		{0.00***}	{0.86}
Neighbor's SA round complete (t-1)	-0.01	0.00	-0.30
	(0.02)	(0.02)	(0.38)
		{0.97}	{0.86}
Control mean	0.9	0.9	7.9
Observations (Years*GPUs)	1155	1155	1155
Clusters (Nearest neighbors)	118	118	118

Notes: Outcomes are NREGA job cards per GPU household and NREGA person days worked per GPU inhabitant. Multiple hypothesis adjusted standard errors by Benjamini-Hochberg for the two main explanatory variables in columns 2 and 3 in curly brackets. All other notes from Table 4.3 apply.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table 4.5: NREGA projects

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA round complete (t-1)	0.58	0.14*	-1.10**
	(0.54)	(0.08)	(0.43)
	{0.29}	{0.21}	{0.07*}
Neighbor's SA round complete (t-1)	1.15*	0.16**	-0.64
	(0.63)	(0.08)	(0.45)
	{0.21}	{0.21}	{0.29}
Control mean	6.1	0.2	3.9
Observations (Years*GPUs)	1155	1155	1155
Clusters (Nearest neighbors)	118	118	118

Notes: Outcomes are ongoing, suspended and completed NREGA work projects per thousand GPU inhabitants. Multiple hypothesis adjusted standard errors by Benjamini-Hochberg for the two main coefficients in columns 1 to 3 in curly brackets. All other notes fom Table 4.3 apply.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

4.5.2 Robustness

To test the robustness of my results, I compare the latter to estimates without spatial dependence in chapter 3. Moreover, I test whether my results are robust to the inclusion of population weights by repeating the main regression in equation (4.1) without weights. Lastly, I extend equation (4.1) to estimates with two instead of one nearest neighbors.

Compared to the estimates unadjusted for spatial dependence from chapter 3, standard errors in the main spatial estimates are comparable and point estimates of the direct effects are as expected of similar size or slightly smaller (a small decrease can be caused by omitted variable bias in specifications neglecting neighborhood effects). The estimated unweighted neighborhood effects, laid out in Tables A4.3-A4.5 in the appendix, all have the same sign and similar size as the weighted estimates in the previous section. An exception is the unweighted effect for ongoing works which is smaller and loses significance in the unweighted estimates. While the neighborhood effect on the percentage of labor in total expenditure is of roughly the same size, it is less precisely estimated and therefore also loses significance. Indeed, standard errors of unweighted estimates are overwhelmingly larger than in the weighted estimates, indicating unaddressed heteroscedasticity in the unweighted results.

Table 4.6: NREGA expenditure, nearest two neighbors

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t-1)	15.78	67.26*	-51.48	-4.42
	(58.63)	(40.33)	(44.37)	(3.34)
		{0.38}	{0.66}	
Neighbors' SA round complete (t-1)	3.86	32.41	-28.55	-3.47
	(92.89)	(60.49)	(65.27)	(4.26)
		{0.66}	{0.66}	
Control mean	1241.5	469.7	771.8	64.8
Observations (Years*GPUs)	118	118	118	118
Clusters (Nearest neighbors)	1155	1155	1155	1124

Notes: See Table 4.3. The main explanatory variable, neighbor's SA round complete (t-1), indicates the average number of social audits completed in the two nearest neighbor GPUs up to the previous year.

* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table 4.7: Employment, nearest two neighbors

	Job cards applied	Job cards issued	Person-days
	(1)	(2)	(3)
SA round complete (t-1)	-0.08***	-0.07***	-0.36
	(0.02)	(0.02)	(0.34)
		{0.00***}	{0.57}
Neighbors' SA round complete (t-1)	0.04	0.04	0.12
	(0.03)	(0.03)	(0.51)
		{0.45}	{0.82}
Control mean	0.9	0.9	7.9
Observations (Years*GPUs)	118	118	118
Clusters (Nearest neighbors)	1155	1155	1155

Notes: See Table 4.4. The main explanatory variable, neighbor's SA round complete (t-1), indicates the average number of social audits completed in the two nearest neighbor GPUs up to the previous year.

* p<0.10 ** p<0.05 *** p<0.01

Table 4.8: NREGA projects, nearest two neighbors

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA round complete (t-1)	0.76	0.15*	-1.17***
	(0.55)	(0.08)	(0.43)
	{0.49}	{0.28}	{0.04*}
Neighbors' SA round complete (t-1)	0.67	0.18*	-0.11
	(0.87)	(0.10)	(0.61)
	{0.85}	{0.28}	{0.85}
Control mean	6.1	0.2	3.9
Observations (Years*GPUs)	118	118	118
Clusters (Nearest neighbors)	1155	1155	1155

Notes: See Table 4.5. The main explanatory variable, neighbor's SA round complete (t-1), indicates the average number of social audits completed in the two nearest neighbor GPUs up to the previous year.

* p<0.10 ** p<0.05 *** p<0.01

The way spatial dependence between communities is modeled could potentially influence my results (Elhorst 2014). In particular, if social audits of higher order neighbors had a similar effect as audits of first neighbors, my first neighbor specification might underestimate the effects of social audits. To test whether this is the case, I repeat my analysis taking also spatial dependence of second nearest neighbors into account. The regression equation corresponds to equation (4.1), except that the spatial lag $AuditStep^{Neigh}$ is now defined as the arithmetic mean of the number of audits that the two nearest neighbors have completed until the previous period. Such weighting of neighbors is recommended to facilitate interpretation of the spatial lag as a weighted average (Elhorst 2014). Thus, β_2 now captures the spillover effect of an additional social audit of either of the two nearest neighbors. The effects in Tables 4.6-4.8 are mostly in line with my assumption that social audits in first nearest neighbors have a stronger spillover effect than social audits of communities that are further away. The estimates for ongoing works and the percentage of labor expenditure loose significance. Only the point estimate for suspended works remains significant and is similar in size to the original estimate, indicating that the second nearest neighbor has a similar influence as the first nearest neighbor on work suspension. In this case, my estimate of the spillover effect based on the first nearest neighbor represents a lower bound of the spillover effect of neighbors in general

4.5.3 Discussion

I observe strong spillover effects for work projects, in particular a significant shift towards ongoing and suspended works, and a significant effect on the percentage of labor in total expenditure - but no significant effects on absolute expenditure, completed works or labor outcomes. All significant and five out of seven insignificant point estimates of spillover effects have the same sign as the direct effects. This overall similar pattern is consistent with the hypothesis in section 4.2.4 that social audits in neighboring communities affect NREGA outcomes in the same direction as having a social audit in one's own community. Such a reinforcing effect is in line with externalities of social audits and stakeholders learning from audited neighboring GPUs about audit probability, their rights and which irregularities are likely detected, and a resulting deterrent effect or detection avoidance by bureaucrats.

Strong spillover effects on work completion but not employment are consistent with a strong communication between bureaucrats of neighboring GPUs but not between beneficiaries. Both the direct and the indirect effect on ongoing works have similar standard errors, but the spillover effect is much larger and can thus be detected at the ten percent significance level. The larger size may be due to different underlying mechanisms for audited GPUs and their neighbors. To explore this, and to evaluate whether the spillover effects on works are beneficial, I consider several mechanisms which may increase ongoing and suspended works through neighboring audits (also allowing for a decrease in completed works) which were found to be relevant in the previous chapter. In chapter 3, I find a pattern for direct effects on work completion status which is similar to the one found here for spillover effects. I trace this direct effect back to the removal of inefficiencies and corruption, but to some extent also to learning of bureaucrats, who attempted to hide corruption from auditors. Hiding was done by declaring

Table 4.9: NREGA projects after policy change to detailed audits of ongoing works

	Ongoing	Completed
	(1)	(2)
SA round complete (t-1)	-0.25	0.39
	(0.76)	(0.58)
	{0.75}	{0.75}
Neighbor's SA round complete (t-1)	2.13***	-0.92
	(0.74)	(0.56)
	{0.02**}	{0.31}
Control mean	6.1	5.1
Observations (Years*GPUs)	660	660
Clusters (Nearest neighbors)	118	118

Notes: See Table 4.5. The sample is restricted to the financial years 2014/15 and later. During that period, no new works were suspended. The sub-sample consists of 4 out of originally 7 years. Multiple hypothesis adjusted standard errors by Benjamini-Hochberg for the two main coefficients (for columns 1 and 2) in curly brackets.
* p<0.10 ** p<0.05 *** p<0.01

factually completed works as ongoing, as auditors initially only scrutinized completed works. However, this strategy only paid off until the year 2014/15 in which the auditing strategy was revised to audit also ongoing works in detail. To test whether hiding behavior drives the results, I restrict the sample to the years from 2014/15 on (Table 4.9). Interestingly, in contrast to the direct effect, the shift to ongoing works in neighboring GPUs becomes *more* pronounced from 2014 on. This points to either incomplete communication between bureaucrats about the changed auditing policy or about corrupt practices. For instance, bureaucrats may initially learn how to hide corruption from their neighbors but fail to learn about the change in auditing policy, so that they continue to label works as ongoing. Bureaucrats in other Indian states are reported to be well-connected across communities and to exchange information on social audit results and sanctions amongst each other (personal communication with Telangana SAU personnel, 25.11.2021). In addition, information exchange could be intensified by collaborative projects, in which higher-level collaborative agencies can connect implementers in several GPUs. However, it is not clear why bureaucrats should communicate about the initial policy but not the subsequent policy change. While my results on works' completion status do not rule out hiding behavior, corruption avoidance strategies may be a more sensitive conversation topic than other social audit-related information. Thus, the continued increase in ongoing and suspended works after the policy change could also be explained by bureaucrats never learning from their neighbors how to hide corruption in ongoing works, so that they have no reason to decrease "excess" ongoing works once these are also being audited. If that is the case, the spillover effect on works must be due to other mechanisms than hiding. For instance, bureaucrats may, in anticipatory obedience, reduce types of inefficiencies and corruption in their own GPU that have been detected in the neighboring community. The following mechanisms have been shown to be relevant for the direct effect of social audits in chapter 3: a reduction in fake completion certificates and resumption, suspension or deletion of works which are not up to standards or doomed to fail. Of these, the revival of poor quality works would be particularly in line with a strong spillover effect on ongoing and suspended but less on completed works[40].

The overall (direct plus spillover) effect on the labor-material ratio is not negligible, as it causes labor expenditure to fall below the stipulated 60 percent of total expenditure[41]. The spillover effect on the labor percentage is significant despite an insignificant direct effect and insignificant separate spillover effects on absolute material and labor expenditure. This can be explained by the fact that the negative spillover on labor expenditure surpasses the positive spillover on material expenditure. This leads to a large spillover in the percentage of labor expenditure which is only slightly greater than the main effect, but has a smaller standard error and is thus detectable at the common ten percent significance level. The decline in labor expenditure through neighboring social audits would be consistent with lower labor costs following a decrease in labor days or wage irregularities. The spillover effect on the percentage of labor expenditure is likely a lower bound. The reason is that we might observe only part of the actual spillover on material expenditure because material in collaborative projects is usually financed by the collaborating organization and thus not observable through GPU expenditure. This could also explain why the spillover effect on material expenditure is relatively small compared to the large direct effect on the same outcome. Alternatively, a small, insignificant spillover could point towards knowledge gaps of bureaucrats about social audit results in neighboring communities. The smaller spillover effect on material expenditure relative to labor expenditure also explains the reversed sign for the (insignificant) spillover effect on total expenditure, the only outcome for which the spillover effect runs in the opposite direction of the direct effect.

I find evidence for no or only weak (insignificant) neighborhood effects on program enrollment and employment. The spillover effects on job cards applied for and issued are precisely estimated zeros. This is surprising, as I find a strong direct effect on job cards. A possible

[40] I cannot explicitly test for all the mechanisms discussed in chapter 3, owing to the fact that the mechanisms which turned out to be the most relevant in chapter 3 require sample restrictions that create an incomplete set of neighbors. Mechanisms that I *can* test for (bureaucrats increase ongoing and suspended works (i) to expand embezzlement opportunities of leftover funds in these works, (ii) to expand employment possibilities in case of employment shocks, or (iii) because they run out of funds) contribute neither in chapter 3 nor in this chapter to the explanation of the observed pattern in work projects (see Table A4.6).

[41] This remains the case even when not weighting the observations by GPU population.

explanation is a lack of information of either bureaucrats or unenrolled citizens regarding social audit outcomes in other communities. While bureaucrats can in principle have great power over the issuance of job cards, I show in chapter 3 that direct effects of social audits on job cards issued seem to be overwhelmingly driven by a decrease in applicants, rather than by bureaucrat behavior. Hence, the absence of any spillover on job cards suggests that potential job card applicants are unaware of the information which leads the citizens in the nearest neighbor GPU to decrease their demand for job cards. In contrast, while a null effect on person days worked cannot be ruled out, the observed insignificant decrease reinforces the direct effect and is thus consonant with a transmission of information between beneficiaries of nearby communities. For instance, information about negative audit outcomes in neighboring communities may reduce the trust of those working in NREGA, thus leading workers in both audited and nearby communities to decrease their labor supply. Lower spillover effects of non-beneficiaries compared to beneficiaries could be explained by the fact that villagers without a job card, and thus not enrolled in NREGA, are less likely to communicate about the program and social audit results than current beneficiaries. This is in line with greater knowledge of beneficiaries compared to non-beneficiaries about NREGA rights found in the state of Bihar (Ravallion et al. 2015), and further supported for the Sikkimese context by a survey of Panda et al. (2009), in which 84.3 and 74.0 percent of beneficiary respondents agree that working together in NREGA improved their information level and created a forum for workers, respectively. Accordingly, once a job card is issued, beneficiaries from nearby communities are more likely to communicate about NREGA and social-audit-related issues, which may influence worker's labor supply.

Overall, my results confirm Olken's (2007) suspicion that spillover effects of social audits are likely to appear at a low spatial level. This is further supported by the fact that the reinforcing spillover effects found in this paper are consonant with those of different community monitoring policies on crime in the U.S. (Gonzalez and Komisarow 2020) and on road construction in Afghanistan (Berman et al. 2017), and spillovers on corruption of Brazilian government audits *without* direct community involvement (Avis et al. 2018). However, in contrast to the setting of Gonzalez and Komisarow, a simultaneous spatial displacement of corruption from one GPU to an even further off GPU is unattractive for transgressors in Sikkim, as bureaucrats cannot easily transfer their workplace to another GPU, and it was clear at the introduction of social audits that all GPUs would be audited eventually. Moreover, the absence of adverse spillover effects on enrollment (and potentially hiding of corruption) contrasts findings by Cisneros and Kis-Katos (2022) of such adverse spillover effects in the context of pure (non-social) government audits.

4.6 Conclusion

Taking into account spillover effects of public policies is essential to reduce biases in the estimation of the effects of such policies and may yield important information about the presence of beneficial and adverse spillover effects. Knowledge of these effects in turn provides the foundation for further policy action, in which beneficial spillover effects can be strengthened and adverse spillover effects reduced, for instance through appropriate policy placement and information strategies. This paper is the first to estimate spillover effects of social audits of a large public program on program outcomes in neighboring communities.

I find that spillover effects from audited to neighboring communities reinforce direct effects of social audits in the audited GPUs. In particular, I find that a greater number of completed social audits in nearby communities significantly increases ongoing and suspended work projects, and decreases the percentage of labor expenditure. Neglecting spillover effects can thus lead to an underestimation of the overall effect of social audits by up to 66 percent[42]. Moreover, I show that an important direct *adverse* effect of social audits on program enrolment does not spill over to nearby communities, while I cannot rule out that adverse spillovers occur in the form of hiding of work projects from auditors. The observed results suggest that certain stakeholder groups, in particular villagers not yet enrolled in the work program, have low access to information about social audit results and the NREGA program in neighbor communities.

These findings provide the starting point for further research, which should investigate in depth the concrete channels behind the spillovers detected in this paper: whether spillovers occur through collaborative work projects or social interactions outside of NREGA, through social pressure, voluntary behavioural change, or the re-assessment of the risk of inspection, and who are the primary disseminators of information on social audit results. The resulting insights can be applied to determine whether the spillover effects on ongoing works are beneficial or adverse effects, and how to hone policy responses to these effects. Moreover, it will facilitate predictions of spillover effects on outcomes outside of NREGA, which are beyond the scope of this paper. For instance, context equilibrium effects through supply and demand interactions may materialize in the form of regional changes in non-NREGA wages and employment opportunities (Angelucci and Di Maro 2016; Merfeld 2019).

All in all, the findings in this paper provide the foundation for tentative policy recommendations: firstly, evaluations of newly introduced pilot social audits need to take into account potential spillover effects, by making sure not to use neighboring villages as control villages, e.g. by random assignment of social audit introduction at block instead of GPU level. Second, if bureaucrats in neighboring communities alter their behavior owing to a long-term change in local norms and the resulting rising public pressure, costs of social audits could in theory be reduced and adverse effects minimized by selecting a random subsample of GPUs to be audited each year. However, such a policy should be considered carefully: if spillovers occur because bureaucrats reassess the risk of being audited, lowering the probability of inspection by auditing less GPUs may set adverse incentives (Olken 2007). Moreover, as the full effect of social audits is the combined effect of the direct and spillover effect, auditing only a subsample will likely not only lower the costs but also the overall effect of social audits. This, however, is an important third result of my analysis: that the overall effects of social audits are more pronounced than it seems at first sight, which can provide further justification for the

[42] For ongoing works.

continuation of social audits if spillover effects are overwhelmingly beneficial. A fourth takeaway is that the large overall decrease in the percentage of labor expenditure threatens NREGA's function as a security net for the unemployed. To ensure that this function is maintained, communities need to take measures to raise labor expenditure. Finally, policy makers have to take into account the potential effects of changing contexts. The spillover effects found in my analysis will likely increase in the future, in particular for labor-related outcomes, since Sikkim signed in 2022 an agreement to collaborate in the construction of large scale road projects. These are co-financed by NREGA but directly implemented by the Border Roads Organization. Such large-scale projects may lead to even greater synergies and to worksites in which laborers from neighboring GPUs work together, thus giving rise to stronger or different spillover effects. Similarly, spillover effects may already be larger in other Indian states, where such macro level NREGA projects are more common. These diverse and changing contexts create a need to continuously adapt policies and incentives to new and different settings.

A.4 Appendix

Table A4.1: Data restrictions and number of observations

Data restriction	Observations	
	(Years*GPUs)	(GPUs)
Unrestricted	1,246	178
Exclude GPUs which split or merged during period of analysis	1,190	170
Exclude GPUs with inseparable, incorrect or unclear geolocation	1,155	165

Note*:* The dataset on NREGA outcomes comprises a balanced sample spanning 2011/12 to 2017/18. Source of the spatial data are village-level shapefiles from DataMeet.

Table A4.2: Description of variables

Variable	**Scale**	**Description**	**Source**
Outcome variables			
Expenditure, total	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure of a GPU in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011
Expenditure, material	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure disbursed on material in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011
Expenditure, Labor	Rupees (₹) per GPU inhabitant	Total real NREGA expenditure disbursed on labor in the financial year (base year of consumer price index is 2010).	NREGA website, population: Census 2011,
Expenditure, Percentage of labor	Percent	Percentage of labor expenditure in total expenditure. Should be above 60% according to NREGA guidelines.	NREGA website, population: Census 2011
Job cards, issued	Per GPU household	Number of NREGA job cards issued.	NREGA website, population: Census 2011
Person days worked in NREGA	Per GPU inhabitant	Number of person days worked in NREGA.	NREGA website, population: Census 2011

Works, total	Per 1,000 GPU inhabitants	Total number of NREGA works (over time).	NREGA website, population: Census 2011
Works, ongoing	Per 1,000 GPU Inhabitants	Number of ongoing NREGA works (works in which some activity/expenditure is taking place).	NREGA website, population: Census 2011
Works, suspended	Per 1,000 GPU Inhabitants	Number of suspended NREGA works (previously ongoing work in which no expenditure/activity is happening but which is not yet closed. Work can be (but is not always) resumed at a later point in time.	NREGA website, population: Census 2011
Works, completed	Per 1,000 GPU inhabitants	Number of completed NREGA works (works on which all activities have been completed and which are thus declared closed or partially closed, meaning all necessary closure documentation has been provided).	NREGA website, population: Census 2011
Main explanatory variables			
Social audit round step variable (t-1)		Indicates the number of social audits that a GPU has been subject to until (and including) the previous year.	Ministry of Rural Development
Neighbors' social audit round step variable (t-1)		Indicates the number of social audits that the nearest neighbor GPU has completed until (and including) the previous year.	Ministry of Rural Development, geolocation from DataMeet

Table A4.3: NREGA expenditure

	Total	Material	Labor	Percentage of labor in total expenditure
	(1)	(2)	(3)	(4)
SA round complete (t-1)	11.85	95.14*	-83.29	-4.12
	(79.42)	(52.46)	(54.56)	(3.02)
		{0.28}	{0.35}	
Neighbor's SA round complete (t-1)	-6.19	52.33	-58.52	-4.07
	(80.86)	(56.10)	(53.10)	(2.95)
		{0.35}	{0.35}	
Control mean	1233.9	466.8	767.1	64.8
Observations (Years*GPUs)	1155	1155	1155	1124
Clusters (Nearest neighbors)	118	118	118	118

Notes: See Table 4.3. The main explanatory variable, neighbor's SA round complete (t-1), indicates the number of social audits in the nearest neighbor GPU completed up to the previous year. Observations not weighted by population.

* p<0.10 ** p<0.05 *** p<0.01

Table A4.4: Employment

	Job cards applied	Job cards issued	Person-days
	(1)	(2)	(3)
SA round complete (t-1)	-0.07***	-0.07***	-0.72*
	(0.02)	(0.02)	(0.42)
		{0.01***}	{0.26}
Neighbor's SA round complete (t-1)	0.02	0.02	-0.17
	(0.02)	(0.02)	(0.44)
		{0.61}	{0.70}
Control mean	0.9	0.9	7.9
Observations (Years*GPUs)	1155	1155	1155
Clusters (Nearest neighbors)	118	118	118

Notes: See Table 4.4. The main explanatory variable, neighbor's SA round complete (t-1), indicates the number of social audits in the nearest neighbor GPU completed up to the previous year. Observations not weighted by population.

* p<0.10 ** p<0.05 *** p<0.01

Table A4.5: NREGA projects

	Ongoing	Suspended	Completed
	(1)	(2)	(3)
SA round complete (t-1)	-0.14	0.14	-0.56
	(0.64)	(0.09)	(0.53)
	{0.83}	{0.52}	{0.73}
Neighbor's SA round complete (t-1)	0.65	0.18**	-0.72
	(0.72)	(0.09)	(0.63)
	{0.73}	{0.28}	{0.73}
Control mean	6.1	0.2	3.9
Observations (Years*GPUs)	1155	1155	1155
Clusters (Nearest neighbors)	118	118	118

Notes: See Table 4.5. The main explanatory variable, neighbor's SA round complete (t-1), indicates the number of social audits in the nearest neighbor GPU completed up to the previous year. Observations not weighted by population.

* p<0.10 ** p<0.05 *** p<0.01

Table A4.6: Mechanisms for neighborhood effects on NREGA projects

Sample:	Full sample			Full sample
Work type:	Ongoing	Suspended	Completed	Approved
	(1)	(2)	(3)	(4)
SA round complete (t-1)	0.73	0.13*	-1.05**	0.69**
	(0.54)	(0.08)	(0.43)	(0.32)
	{0.18}	{0.18}	{0.08*}	
Neighbor's SA round complete (t-1)	1.25**	0.17**	-0.67	0.07
	(0.63)	(0.08)	(0.44)	(0.28)
	{0.18}	{0.18}	{0.18}	
10-99% of sanctioned funds disbursed	-0.19	0.18*	-0.74**	
	(0.44)	(0.09)	(0.35)	
All sanctioned funds disbursed (≥100%)	-3.77***	0.33	-0.25	
	(0.90)	(0.35)	(0.75)	
Control mean	6.1	0.2	3.9	1.9
Observations (Years*GPUs)	1155	1155	1155	1155
Clusters (Nearest neighbors)	118	118	118	118

Notes: See Table 4.5. Outcomes are ongoing, suspended, completed and approved works. Columns 1-3 control for the percentage of funds disbursed (the percentage can exceed 100 if additional funds are granted for project completion). Reference category for additional controls regarding funds disbursed is less than 10 percent of funds disbursed. Outcome in column 4 is approved works (on which no activity takes place yet) per 1000 inhabitants. Multiple hypothesis adjusted standard errors by Benjamini-Hochberg for the two main coefficients for columns 1 to 3 in curly brackets.

* p<0.10 ** p<0.05 *** p<0.01

References

Abadie, Alberto; Athey, Susan; Imbens, Guido W.; Wooldridge, Jeffrey (2017): When should you adjust standard errors for clustering? National Bureau of Economic Research (w24003).

Afridi, Farzana; Iversen, Vegard (2014): Social audits and MGNREGA delivery. Lessons from Andhra Pradesh (IZA Discussion Paper, 8095).

Afridi, Farzana; Iversen, Vegard; Sharan, Mamidipudi R. (2017): Women political leaders, corruption, and learning. Evidence from a large public program in India. In *Economic Development and Cultural Change* 66 (1), pp. 1–30.

Ahmad, Ehtisham; Stern, Nicholas (1987): Alternative sources of government revenue: illustrations from India, 1979-80. In David Newbery, Nicholas Stern (Eds.): The theory of taxation for developing countries. New York: Oxford University Press.

Ahuja, Ravi (2020): Minoritarian labour welfare in India: the case of the Employees' State Insurance Act of 1948. In Lutz Leisering (Ed.): One hundred years of social protection. The changing social question in Brazil, India, China, and South Africa. Basingstoke: Palgrave Macmillan (Global dynamics of social policy), pp. 157–188.

Aiyar, Yamini; Mehta, S. Kapoor (2015): Spectators or participants? Effects of social audits in Andhra Pradesh. In *Economic and Political Weekly* 50 (7), pp. 66–71.

Alhassan, Robert Kaba; Nketiah-Amponsah, Edward; Ayanore, Martin Amogre; Afaya, Agani; Salia, Solomon Mohammed; Milipaak, Japiong; Ansah, Evelyn Korkor; Owusu-Agyei, Seth (2019): Impact of a bottom-up community engagement intervention on maternal and child health services utilization in Ghana: a cluster randomised trial. In *BMC Public Health* 19 (1), pp. 1–11.

Alix-Garcia, Jennifer; McIntosh, Craig; Sims, Katharine R. E.; Welch, Jarrod R. (2013): The ecological footprint of poverty alleviation: evidence from Mexico's Oportunidades program. In *The Review of Economics and Statistics* 95 (2), pp. 417–435.

Alkire, Sabina; Chatterjee, Mihika; Conconi, Adriana; Seth, Suman; Vaz, Ana (2014): Poverty in rural and urban areas. Direct comparisons using the global MPI 2014. OPHI.

Anand, Sudhir; Bärnighausen, Till (2007): Health workers and vaccination coverage in developing countries: an econometric analysis. In *The Lancet* 369 (9569), pp. 1277–1285.

Anand, Utkarsh (2015): India has 31 lakh NGOs, more than double the number of schools. In *The Indian Express*, 8/1/2015. Available online at https://indianexpress.com/article/india/india-others/india-has-31-lakh-ngos-twice-the-number-of-schools-almost-twice-number-of-policemen/, accessed on 8/7/2022.

Andrews, Rhys; Boyne, George A.; Enticott, Gareth (2006): Performance failure in the public sector: misfortune or mismanagement? In *Public Management Review* 8 (2), pp. 273–296.

Angelucci, Manuela; Di Maro, Vincenzo (2016): Programme evaluation and spillover effects. In *Journal of Development Effectiveness* 8 (1), pp. 22–43.

Angelucci, Manuela; Giorgi, Giacomo de (2009): Indirect effects of an aid program. How do cash transfers affect ineligibles' consumption? In *American Economic Review* 99 (1), pp. 486–508.

Angrist, Joshua D.; Pischke, Jörn-Steffen (2008): Mostly harmless econometrics. An empiricist's companion. Princeton: Princeton University Press.

Asfaw, Abay; Lamanna, Francesca; Klasen, Stephan (2010): Gender gap in parents' financing strategy for hospitalization of their children. Evidence from India. In *Health Economics* 19 (3), pp. 265–279.

Atkin, David (2009): Working for the future: female factory work and child health in Mexico. Yale. Unpublished manuscript.

Avis, Eric; Ferraz, Claudio; Finan, Frederico (2018): Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians. In *Journal of Political Economy* 126 (5), pp. 1912–1964.

Balarajan, Yarlini; Selvaraj, Sakthivel; Subramanian, S. V. (2011): Health care and equity in India. In *The Lancet* 377 (9764), pp. 505–515.

Banerjee, Abhijit; Banerji, Rukmini; Berry, James; Duflo, Esther; Kannan, Harini; Mukerji, Shobhini; Shotland, Marc; Walton, Michael (2017): From proof of concept to scalable policies: challenges and solutions, with an application. In *Journal of Economic Perspectives* 31 (4), pp. 73–102.

Banerjee, Abhijit; Duflo, Esther (2011): Poor economics. A radical rethinking of the way to fight global poverty. New York: Public Affairs.

Banerjee, Abhijit; Duflo, Esther; Glennerster, Rachel; Kinnan, Cynthia (2015): The miracle of microfinance? Evidence from a randomized evaluation. In *American Economic Journal: Applied Economics* 7 (1), pp. 22–53.

Banerjee, Abhijit; Niehaus, Paul; Suri, Tavneet (2019): Universal basic income in the developing world. In *Annual Review of Economics* 11, pp. 959–983.

Banerjee, Abhijit Vinayak; Duflo, Esther; Glennerster, Rachel; Kothari, Dhruva (2010): Improving immunisation coverage in rural India. Clustered randomised controlled evaluation of immunisation campaigns with and without incentives. In *The BMJ* 340, c2220.

Barrera-Osorio, Felipe; Bertrand, Marianne; Linden, Leigh L.; Perez-Calle, Francisco (2011): Improving the design of conditional transfer programs: evidence from a randomized education experiment in Colombia. In *American Economic Journal: Applied Economics* 3 (2), pp. 167–195.

Bastagli, Francesca; Hagen-Zanker, Jessica; Harman, Luke; Barca, Valentina; Sturge, Georgina; Schmidt, Tanja (2019): The impact of cash transfers. A review of the evidence from low-and middle-income countries. In *Journal of Social Policy* 48 (3), pp. 569–594.

Bela, Himali (2022): MGNREGA workers to be employed in BRO works. In *Sikkim Express*, 2/19/2022. Available online at http://www.sikkimexpress.com/news-details/mgnrega-workers-to-be-employed-in-bro-works, accessed on 7/25/2022.

Benjamini, Yoav; Hochberg, Yosef (1995): Controlling the false discovery rate. A practical and powerful approach to multiple testing. In *Journal of the Royal Statistical Society: series B (Methodological)* 57 (1), pp. 289–300.

Berman, Eli; Callen, Michael; Condra, Luke N.; Downey, Mitch; Ghani, Tarek; Isaqzadeh, Mohammad (2017): Community monitors vs. leakage: experimental evidence from Afghanistan. Mimeo.

Bertrand, Marianne; Duflo, Esther; Mullainathan, Sendhil (2004): How much should we trust differences-in-differences estimates? In *The Quarterly Journal of Economics* 119 (1), pp. 249–275.

Besley, Timothy; Ghatak, Maitreesh (2006): Public goods and economic development. In *Understanding Poverty* 19, pp. 285–303.

Björkman Nyqvist, Martina; Walque, Damien de; Svensson, Jakob (2017): Experimental evidence on the long-run impact of community-based monitoring. In *American Economic Journal: Applied Economics* 9 (1), pp. 33–69.

Black, Robert E.; Allen, Lindsay H.; Bhutta, Zulfiqar A.; Caulfield, Laura E.; Onis, Mercedes de; Ezzati, Majid; Mathers, Colin; Rivera, Juan; Group, Maternal and Child Undernutrition Study (2008): Maternal and child undernutrition. Global and regional exposures and health consequences. In *The Lancet* 371 (9608), pp. 243–260.

Bobba, Matteo; Gignoux, Jérémie (2019): Neighborhood effects in integrated social policies. In *The World Bank Economic Review* 33 (1), pp. 116–139.

Bobonis, Gustavo J.; Finan, Frederico (2009): Neighborhood peer effects in secondary school enrollment decisions. In *The Review of Economics and Statistics* 91 (4), pp. 695–716.

Cameron, A. Colin; Miller, Douglas L. (2015): A practitioner's guide to cluster-robust inference. In *Journal of Human Resources* 50 (2), pp. 317–372.

Chao, John C.; Swanson, Norman R. (2005): Consistent estimation with a large number of weak instruments. In *Econometrica* 73 (5), pp. 1673–1692.

Chêne, Marie (2014): The impact of corruption on growth and inequality. Transparency International (Anti-corruption helpdesk).

Christensen, Garret; Miguel, Edward (2018): Transparency, reproducibility, and the credibility of economics research. In *Journal of Economic Literature* 56 (3), pp. 920–980.

Cisneros, Elías; Kis-Katos, Krisztina (2022): Unintended environmental consequences of anti-corruption strategies. Social Science Research Network (SSRN) (SSRN working paper, 3899498).

Cleland, John; Conde-Agudelo, Agustin; Peterson, Herbert; Ross, John; Tsui, Amy (2012): Contraception and health. In *The Lancet* 380 (9837), pp. 149–156.

Colella, Fabrizio; Lalive, Rafael; Sakalli, Seyhun Orcan; Thoenig, Mathias (2019): Inference with arbitrary clustering. Institute of Labor Economics (IZA). Bonn (IZA discussion papers, 12584).

Currie, Janet (2009): Healthy, wealthy, and wise. Socioeconomic status, poor health in childhood, and human capital development. In *Journal of Economic Literature* 47 (1), pp. 87–122.

DataMeet (2021): Villages maps provided by Indian Village Boundaries Project. Available online at http://projects.datameet.org/indian_village_boundaries/, accessed on 2/5/2021.

De, Prabal K. (2017): Causal effects of maternal schooling on child immunization in India. In Kristian Bolin, Björn Lindgren, Michael Grossman, Dorte Gyrd-Hansen, Tor Iversen, Robert Kaestner, Jody Sindelar (Eds.): Human Capital and Health Behavior. Bingley: Emerald Publishing Limited (Advances in Health Economics and Health Services Research, 25).

De, Prabal K.; Timilsina, Laxman (2020): Cash-based maternal health interventions can improve childhood vaccination. Evidence from India. In *Health Economics* 29 (10), pp. 1202–1219.

Desai, Sonalde; Vashishtha, Prem; Joshi, Omkar (2015): Mahatma Gandhi National Rural Employment Guarantee Act: a catalyst for rural transformation. National Council of Applied Economic Research. New Delhi.

Dhirar, Nonita; Dudeja, Sankalp; Khandekar, Jyoti; Bachani, Damodar (2018): Childhood morbidity and mortality in India. Analysis of National Family Health Survey 4 (NFHS-4) findings. In *Indian Pediatrics* 55 (4), pp. 335–338.

DRD; MRD (2014): Abstract of state convergence plans received so far by the Ministry for FY 2014-15. Department of Rural Development; Ministry of Rural Development; Government of India. Available online at https://nrega.nic.in/netnrega/convergence/conindex.aspx, accessed on 7/28/2022.

Duflo, Esther (2017): The economist as plumber. In *American Economic Review* 107 (5), pp. 1–26.

Duggal, Ravi (2001): Evolution of health policy in India. CEHAT.

Dupas, Pascaline (2011): Health behavior in developing countries. In *Annual Review of Economics* 3 (1), pp. 425–449.

Elhorst, J. Paul (2014): Spatial econometrics. From cross-sectional data to spatial panels. Heidelberg, New York, Dordrecht, London: Springer (Springer Briefs in Regional Science).

Falcao, Vanita Leah; Khanuja, Jasmeet; Matharu, Sonal; Nehra, Shikha; Sinha, Dipa (2015): Report on the study of the Indira Gandhi Matritva Sahyog Yojana. Centre for Equity Studies. New Delhi.

Ferraz, Claudio; Finan, Frederico; Moreira, Diana B. (2012): Corrupting learning. Evidence from missing federal education funds in Brazil. In *Journal of Public Economics* 96 (9-10), pp. 712–726.

Gaarder, Marie M.; Glassman, Amanda; Todd, Jessica E. (2010): Conditional cash transfers and health. Unpacking the causal chain. In *Journal of Development Effectiveness* 2 (1), pp. 6–50.

Ghosh, Prabhat; Kochar, Anjini (2018): Do welfare programs work in weak states? Why? Evidence from a maternity support program in India. In *Journal of Development Economics* 134 (C), pp. 191–208.

Gibbons, Stephen; Overman, Henry G. (2012): Mostly pointless spatial econometrics? In *Journal of Regional Science* 52 (2), pp. 172–191.

Glassman, Amanda; Duran, Denizhan; Fleisher, Lisa; Singer, Daniel; Sturke, Rachel; Angeles, Gustavo; Charles, Jodi; Emrey, Bob; Gleason, Joanne; Mwebsa, Winnie (2013): Impact of conditional cash transfers on maternal and newborn health. In *Journal of Health, Population, and Nutrition* 31 (4 Supplement 2), 48-66.

Glick, David: Choosing Among Health Providers. How conditional cash transfers affect long-term behavior. In : Glick, David 2017 - Essays in development economics and public program provision. Brown University. Department of Economics.

Gonzalez, Robert M.; Komisarow, Sarah (2020): Community monitoring and crime. Evidence from Chicago's Safe Passage Program. In *Journal of Public Economics* 191 (104250).

Gordon, Jessica; Tranchant, Jean-Pierre; Casu, Laura; Mitchell, Becky; Nisbett, Nicholas (2019): APPI/SPREAD Collective Action for Nutrition Social Audit Programme Odisha, India. Final evaluation report. Institute of Development Studies (IDS). Brighton.

Grandvoinnet, Helene; Aslam, Ghazia; Raha, Shomikho (2015): Opening the black box. The contextual drivers of social accountability. World Bank. Washington DC (New frontiers of social policy).

Grépin, Karen A.; Habyarimana, James; Jack, William (2019): Cash on delivery. Results of a randomized experiment to promote maternal health care in Kenya. In *Journal of Health Economics* 65, pp. 15–30.

Haushofer, Johannes; Shapiro, Jeremy (2016): The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. In *The Quarterly Journal of Economics* 131 (4), pp. 1973–2042.

IIPS (2010a): District Level Household and Facility Survey.

IIPS (2010b): District Level Household and Facility Survey (DLHS-3), 2007-08. Country report. International Institute for Population Sciences (IIPS). Mumbai, India.

IIPS; IFG (2018): Demographic and Health Survey (DHS) / National Family Health Survey (NFHS-4), 2015-16: India.

ILO (2022): World social protection report 2020–22. International Labour Organization. Geneva.

Janvry, Alain de; Sadoulet, Elisabeth (2016): Development economics. Theory and practice. London EN, New York NY: Routledge.

Kahn-Lang, Ariella; Lang, Kevin (2020): The promise and pitfalls of differences-in-differences. Reflections on 16 and pregnant and other applications. In *Journal of Business and Economic Statistics* 38 (3), pp. 613–620.

Kalra, Aditya; MacAskill, Andrew (2015): After uproar, India ups budget for some social welfare sectors. Reuters. Available online at https://www.reuters.com/article/india-budget-idUKKCN0Q51DY20150731, updated on 32.7.2015, accessed on 8/29/2015.

Kapur, Devesh; Nangia, Prakirti (2015): Social protection in India: a welfare state sans public goods? In *India Review* 14 (1), pp. 73–90.

Karuna, M.; Dheeraja, C.; Kumar Sinha, Rajesh; Srinivas, Sajja; Shahameed, Ali; Shashidhar, R. (2019). Status of social audits in India. National Institute of Rural Development and Panchayati Raj. Hyderabad.

Khwaja, Asim Ijaz (2004): Is increasing community participation always a good thing? In *Journal of the European Economic Association* 2 (2-3), pp. 427–436.

Kruglanski, Arie W.; Freund, Tallie (1983): The freezing and unfreezing of lay-inferences: effects on impressional primacy, ethnic stereotyping, and numerical anchoring. In *Journal of Experimental Social Psychology* 19 (5), pp. 448–468.

Kumar, Anand (2005): The welfare state system in India. In : Welfare states and the future: Springer, pp. 336–363.

Lahariya, Chandrakant (2014): A brief history of vaccines & vaccination in India. In *The Indian Journal of Medical Research* 139 (4), p. 491.

Lakha, Salim (2011): Accountability from below. The experience of MGNREGA in Rajasthan (India). Asia Research Institute, National University of Singapore. Singapore (Working Paper Series, 171).

Lim, Stephen S.; Dandona, Lalit; Hoisington, Joseph A.; James, Spencer L.; Hogan, Margaret C.; Gakidou, Emmanuela (2010): India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities. An impact evaluation. In *The Lancet* 375 (9730), pp. 2009–2023.

Maluccio, John; Flores, Rafael (2005): Impact evaluation of a conditional cash transfer program. The Nicaraguan Red de Protección Social. International Food Research Institute. Washington, DC (Research report (International Food Research Institute), 141).

Maluccio, John A.; Hoddinott, John; Behrman, Jere R.; Martorell, Reynaldo; Quisumbing, Agnes R.; Stein, Aryeh D. (2009): The impact of improving nutrition during early childhood on education among Guatemalan adults. In *Economic Journal* 119 (537), pp. 734–763.

Manley, James; Gitter, Seth; Slavchevska, Vanya (2013): How effective are cash transfers at improving nutritional status? In *World Development* 48, pp. 133–155.

Mansuri, Ghazala; Rao, Vijayendra (2013): Can participation be induced? Some evidence from developing countries. In *Critical Review of International Social and Political Philosophy* 16 (2), pp. 284–304.

Merfeld, Joshua D. (2019): Spatially heterogeneous effects of a public works program. In *Journal of Development Economics* 136, pp. 151–167.

Miguel, Edward; Kremer, Michael (2004): Worms. Identifying impacts on education and health in the presence of treatment externalities. In *Econometrica* 72 (1), pp. 159–217.

Ministry of Health and Family Welfare (2011a): Indira Ghandi Matritva Sahyog Yojana. A conditional maternity benefit scheme. Training module.

Ministry of Health and Family Welfare (2011b): Janani Suraksha Yojana. Guidelines for implementation.

Ministry of Women and Child Development (2011): Indira Ghandi Matritva Sahyog Yojana (IGMSY) - a conditional maternity benefit scheme. Implementation guidelines for state governments/UT administrations. New Delhi.

Ministry of Women and Child Development (2019): Annual reports.

Mobarak, Ahmed Mushfiq; Rosenzweig, Mark (2014): Risk, insurance and wages in general equilibrium. National Bureau of Economic Research. Cambridge (NBER working papers, 19811).

Modugu, Hanimi Reddy; Kumar, Manish; Kumar, Ashok; Millett, Christopher (2012): State and socio-demographic group variation in out-of-pocket expenditure, borrowings and Janani Suraksha Yojana (JSY) programme use for birth deliveries in India. In *BMC Public Health* 12 (1), pp. 1–19.

Molina, Ezequiel (2013): Bottom up institutional reform: Evaluating the impact of the citizen visible audit program in Colombia. Unpublished manuscript.

Molina, Ezequiel; Carella, Laura; Pacheco, Ana; Cruces, Guillermo; Gasparini, Leonardo (2016): Community monitoring interventions to curb corruption and increase access and quality of service delivery in low-and middle-income countries. A systematic review. In *Campbell Systematic Reviews* 12 (1), pp. 1–204.

Molina, Ezequiel; Carella, Laura; Pacheco, Ana; Cruces, Guillermo; Gasparini, Leonardo (2017): Community monitoring interventions to curb corruption and increase access and quality in service delivery: a systematic review. In *Journal of Development Effectiveness* 9 (4), pp. 462–499.

MRD and GOI (2011): MGNREGS audit of scheme rules. Source: Ministry of Rural Development; Government of India. In *The Gazette of India* 2 (3(i)).

MRD and GOI (2013): Mahatma Ghandi National Rural Employment Guarantee Act, 2005. Operational guidelines 2013. 4th ed. Ministry of Rural Development; Government of India. New Delhi.

MRD and GOI (2021a): Advisory on convergence of Mahatma Ghandhi NREGS and Border Roads Organisation (BRO). Ministry of Rural Development; Government of India. New Delhi.

MRD and GOI (2021b): MGNREGA public data portal. Ministry of Rural Development; Government of India. New Delhi. Available online at https://nregarep2.nic.in/netnrega/dynamic2/dynamicreport_new4.aspx, accessed on 11/10/2020.

MRD and GOI (2022a): R6.13 Work category wise no. of work under convergence entered in MIS. Ministry of Rural Development; Government of India. New Delhi. Available online at http://mnregaweb4.nic.in/netnrega/MISreport4.aspx?, updated on 2022, accessed on 7/27/2022.

MRD and GOI (2022b): R6.18 Dynamic report for monitoring and details of works. Ministry of Rural Development; Government of India. New Delhi. Available online at http://mnregaweb4.nic.in/netnrega/MISreport4.aspx?, updated on 2022, accessed on 7/27/2022.

Muralidharan, Karthik; Niehaus, Paul; Sukhtankar, Sandip (2016): Building state capacity: evidence from biometric smartcards in India. In *American Economic Review* 106 (10), pp. 2895–2929.

Mwabu, Germano (2007): Health economics for low-income countries. In *Handbook of development economics* 4, pp. 3305–3374.

Nandi, Arindam; Laxminarayan, Ramanan (2016): The unintended effects of cash transfers on fertility. Evidence from the Safe Motherhood Scheme in India. In *Journal of Population Economics* 29 (2), pp. 457–491.

Narayanan, Sudha; Das, Upasak (2014): Women participation and rationing in the employment guarantee scheme. In *Economic and Political Weekly* 49 (46), pp. 46–53.

Niehaus, Paul; Sukhtankar, Sandip (2013): The marginal rate of corruption in public programs: evidence from India. In *Journal of Public Economics* 104, pp. 52–64.

Niti Aayog; DMEO (2017): Quick evaluation study on Indira Gandhi Matritva Sahyog Yojana (IGMSY). Niti Aayog; Development Monitoring and Evaluation Office (DMEO). Delhi.

Office of the Registrar General; Census Commissioner (2011): Census of India.

Olken, Benjamin A. (2006): Corruption and the costs of redistribution. Micro evidence from Indonesia. In *Journal of Public Economics* 90 (4-5), pp. 853–870.

Olken, Benjamin A. (2007): Monitoring corruption. Evidence from a field experiment in Indonesia. In *Journal of Political Economy* 115 (2), pp. 200–249.

Olken, Benjamin A. (2009): Corruption perceptions vs. corruption reality. In *Journal of Public Economics* 93 (7-8), pp. 950–964.

Olken, Benjamin A.; Pande, Rohini (2012): Corruption in developing countries. In *Annual Review of Economics* 4 (1), pp. 479–509.

Panda, B.; Dutta, A. K.; Prusty, S. (2009): Appraisal of NREGA in the States of Meghalaya and Sikkim. Rajiv Ghandi Institute of Management. Shillong.

Pelissery, Sony (2020): Social policy in India: one hundred years of the (stifled) social question. In Lutz Leisering (Ed.): One hundred years of social protection. The changing social question in Brazil, India, China, and South Africa. Basingstoke: Palgrave Macmillan (Global dynamics of social policy), pp. 121–156.

PIB (10/20/2010): Implementation of Indira Gandhi Matritva Sahyog Yojana on pilot basis in 52 districts approved. New Delhi. Press Information Bureau, Government of India.

PIB (12/28/2017): Maternity benefits under Pradhan Mantri Matru Vandana Yojana. New Delhi. Press Information Bureau, Government of India.

PIB (12/6/2019): Maternity benefits under PMMVY. New Delhi. Press Information Bureau, Government of India.

Pinker, Robert A. (2022): Social service https://www.britannica.com/topic/social-service (Encyclopedia Britannica). Available online at https://www.britannica.com/topic/social-service., updated on 8/9/2022, accessed on 8/31/2022.

Powell-Jackson, Timothy; Hanson, Kara (2012): Financial incentives for maternal health. Impact of a national programme in Nepal. In *Journal of Health Economics* 31 (1), pp. 271–284.

Powell-Jackson, Timothy; Mazumdar, Sumit; Mills, Anne (2015): Financial incentives in health. New evidence from India's Janani Suraksha Yojana. In *Journal of Health Economics* 43, pp. 154–169.

Quisumbing, Agnes R.; Maluccio, John A. (2000): Intrahousehold allocation and gender relations: new empirical evidence. (FCND Discussion Paper, 84).

Rahman, Mohammad Mahbubur; Pallikadavath, Saseendran (2018): How much do conditional cash transfers increase the utilization of maternal and child health care services? New evidence from Janani Suraksha Yojana in India. In *Economics and Human Biology* 31, pp. 164–183.

Rai, Kunal (2020): Sikkim seeks centre's approval for setting up community radio stations. Sikkim Express. Available online at http://www.sikkimexpress.com/news-details/sikkim-seeks-centres-approval-for-setting-up-community-radio-stations, updated on 12/5/2020, accessed on 9/4/2022.

Ranganathan, Meghna; Lagarde, Mylene (2012): Promoting healthy behaviours and improving health outcomes in low and middle income countries. A review of the impact of conditional cash transfer programmes. In *Preventive Medicine* 55 Supplement, S95-S105.

Ravallion, Martin; van de Walle, Dominique; Dutta, Puja; Murgai, Rinku (2015): Empowering poor people through public information? Lessons from a movie in rural India. In *Journal of Public Economics* 132, pp. 13–22.

RDD; GOS (2019): Annual Report 2018-19. Rural Management and Development Department; Government of Sikkim.

Ringold, Dena; Holla, Alaka; Koziol, Margaret; Srinivasan, Santhosh (2011): Citizens and service delivery. Assessing the use of social accountability approaches in human development sectors. Washington DC.

Rivera, Juan A.; Sotres-Alvarez, Daniela; Habicht, Jean-Pierre; Shamah, Teresa; Villalpando, Salvador (2004): Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children. A randomized effectiveness study. In *Jama* 291 (21), pp. 2563–2570.

RMDD (2014): Convergence user manual of horticulture plantations. Convergence of Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS) with schemes of Horticulture and Cash Crop Development Department (HCCDD). Rural Management and Development Department Sikkim.

RMDD; GOS; SAU Sikkim (2015): Social audit handbook. An instrument for making the programme accountable to the people. 3rd ed. Rural Management and Development Department; Government of Sikkim; SAU Sikkim. Gangtok.

Rodrik, Dani; Rosenzweig, Mark R. (2010): Handbook of development economics. 5th ed. Amsterdam, Boston: Elsevier/North-Holland (Handbooks in economics).

Romano, Joseph P.; Wolf, Michael (2005): Exact and approximate stepdown methods for multiple hypothesis testing. In *Journal of the American Statistical Association* 100 (469), pp. 94–108.

Rosenzweig, Mark R.; Schultz, T. Paul (1983): Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight. In *Journal of Political Economy* 91 (5), pp. 723–746.

Ruggiero, Cara F.; Hohman, Emily E.; Birch, Leann L.; Paul, Ian M.; Savage, Jennifer S. (2020): The Intervention Nurses Start Infants Growing on Healthy Trajectories (INSIGHT) responsive parenting intervention for firstborns impacts feeding of secondborns. In *The American Journal of Clinical Nutrition* 111 (1), pp. 21–27.

Rüttenauer, Tobias (2022): Spatial regression models: a systematic comparison of different model specifications using Monte Carlo experiments. In *Sociological Methods and Research* 51 (2), pp. 728–759.

Samuel, Miriam; Srinivasan, Sekar (2016): Toward livelihood security through the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA). In Julie L. Drolet (Ed.): Social development and social work perspectives on social protection. London: Routledge, pp. 175–194.

Scott, John; Marshall, Gordon (2009): A dictionary of sociology. 3rd rev. ed. / edited by John Scott and Gordon Marshall. Oxford: Oxford University Press (Oxford paperback reference).

Seitz, Victoria; Apfel, Nancy H. (1994): Parent-focused intervention. Diffusion effects on siblings. In *Child Development* 65 (2), pp. 677–683.

Sen, Amartya (1988): The concept of development. In Hollis Chenery, Thirukodikaval Srinivasan (Eds.): Handbook of development economics, vol. 1. 1 volume: Elsevier Science Publishers B.V., pp. 9–26.

Sen, Amartya (1999): Development as freedom. Oxford: Oxford University Press.

Shankar, Shylashri (2010): Can Social Audits Count? In *Australian National University, Australia South Asia Research Centre Working Paper* 9.

Sinha, Dipa; Nehra, Shikha; Matharu, Sonal; Khanuja, Jasmeet; Leah Falcao, Vanita (2016): Realising universal maternity entitlements. Lessons from Indira Gandhi Matritva Sahyog Yojana. In *Economic and Political Weekly* 51 (34), pp. 49–55.

Solon, Gary; Haider, Steven J.; Wooldridge, Jeffrey M. (2015): What are we weighting for? In *Journal of Human Resources* 50 (2), pp. 301–316.

Stecklov, Guy; Winters, Paul; Todd, Jessica; Regalia, Ferdinando (2007): Unintended effects of poverty programmes on childbearing in less developed countries. Experimental evidence from Latin America. In *Population Studies* 61 (2), pp. 125–140.

Subba, Ash Bahadur; Mishra, P. K.; Jha, Sudha Kumari (2018): Institutionalization of social audit and its impact in MGNREGA: experiences from Sikkim. In *International Journal of Innovative Studies in Sociology and Humanities* 3 (11), pp. 18–27.

Tambe, Sandeep; Arrawatia, M. L.; Ganeriwala, Anil (2012): Managing rural development in the mountain state of Sikkim, India: experiences, innovative approaches, and key issues. In *Mountain Research and Development*.

Tambe, Sandeep; Subba, Ash Bahadur; Basi, Jigme; Pradhan, Sarika; Rai, B. B. (2016a): Measuring the effectiveness of social audits. Experiences from Sikkim, India. In *Development in Practice* 26 (2), pp. 184–192.

Tambe, Sandeep; Subba, Ash Bahadur; Basi, Jigme; Rai, B. B. (2016b): Decentralising accountability: anti-corruption experiment from Sikkim. In *Economic and Political Weekly* (51), Article 95-101, p. 52.

UN DESA (2015): The millennium development goals report 2015. United Nations Department of Economic and Social Affairs. New York.

UNDP (2015): Human development report 2015. Work for human development. United Nations Development Program. New York.

UNICEF; WHO (2004): Low birthweight. Country, regional and global estimates. New York.

Varghese, Feba; Narayanan, N. C.; Agnihotri, S.B., Godbole, Girija (Eds.) (2019): Transparency, accountability and participation through social audit: case of MGNREGA in Sikkim, India. Proceedings of the 13th International RAIS Conference on Social Sciences and Humanities: Scientia Moralitas Research Institute.

Waddington, Hugh; Sonnenfeld, Ada; Finetti, Juliette; Gaarder, Marie; John, Denny; Stevenson, Jennifer (2019): Citizen engagement in public services in low-and middle-income countries: A mixed-methods systematic review of participation, inclusion, transparency and accountability (PITA) initiatives. In *Campbell Systematic Reviews* 15 (1-2), e1025.

WHO (2010): Nutrition Landscape Information System (NLIS). Country profile indicators. Interpretation guide. World Health Organization (WHO).

WHO (2011): Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity (WHO/NMH/NHD/MNM/11.1). Vitamin and Mineral Nutrition Information System. World Health Organization (WHO). Geneva.

WHO (2012): World Health Statistics 2012. Geneva: World Health Organization.

WHO (2014): Global nutrition targets 2025: Low birth weight policy brief (WHO/NMH/NHD/14.5). Geneva.

WHO (2018): World health statistics 2018. Monitoring health for the SDGs. Geneva: World Health Organization.

WHO (2021): Full immunization coverage among one-year-olds (%). The Global Health Observatory (GHO). Indicator Metadata Registry List. World Health Organization (WHO). Available online at https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3317, accessed on 6/4/2021.

Wooldridge, Jeffrey M. (2010): Econometric analysis of cross section and panel data. Cambridge MA, London: MIT Press.

World Bank (1999): Principles and good practice in social policy. Issues and areas for public action. Washington DC.

World Bank (2019): World Development Indicators (WDI). Available online at https://datacatalog.worldbank.org/dataset/world-development-indicators, updated on 4/24/2019, accessed on 5/15/2019.

World Bank (2022a): Current health expenditure (% of GDP) - high income, low and middle income. World Bank. Available online at https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=XD-XO, updated on 1/30/2022, accessed on 8/31/2022.

World Bank (2022b): Government expenditure on education, total (% of government expenditure) - high income, low and middle income. World Bank. Available online at https://data.worldbank.org/indicator/SE.XPD.TOTL.GB.ZS?locations=XD-XO, updated on 6/1/2022, accessed on 8/31/2022.

World Bank (2022c): Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population) - low and middle income, high income. World Bank. Available online at https://data.worldbank.org/indicator/SI.POV.DDAY?locations=XO-XD&year_high_desc=true, accessed on 8/13/2022.

Zamboni, Yves; Litschig, Stephan (2018): Audit risk and rent extraction. Evidence from a randomized evaluation in Brazil. In *Journal of Development Economics* 134, pp. 133–149.

Zangger, Christoph (2019): Making a place for space: using spatial econometrics to model neighborhood effects. In *Journal of Urban Affairs* 41 (8), pp. 1055–1080.