



Assessing the impacts of governance reforms on health services delivery: a quasi-experimental, multi-method, and participatory approach

Alan Zarychta¹ · Krister P. Andersson² · Elisabeth D. Root³ · Jane Menken² · Tara Grillos⁴

Received: 10 April 2018 / Revised: 6 May 2019 / Accepted: 7 June 2019 / Published online: 27 June 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Despite considerable advances in developing new and more sophisticated impact evaluation methodologies and toolkits, policy research continues to suffer from persistent challenges in achieving the evaluation trifecta: identifying effects, isolating mechanisms, and influencing policy. For example, evaluation studies are routinely hampered by problems of establishing valid counterfactuals due to endogeneity and selection effects with respect to policy reform. Additionally, robust evaluation studies often must contend with heterogeneity in treatment, staggered timing, and variation in uptake. And finally, on practical grounds, researchers frequently struggle to involve policymakers and practitioners throughout the research process in order to engender the type of trust needed for policy influence. While it can be difficult to generalize about appropriate evaluation methodologies across contexts, prominent policy interventions like governance reforms for improving health services delivery nonetheless demand rigorous and comprehensive evaluation strategies that can produce valid results and engage policymakers. Drawing on illustrations from our research on health sector decentralization in Honduras, in this paper we present a quasi-experimental, multi-method, and participatory approach that addresses these persistent challenges to policy evaluation.

Keywords Impact evaluation · Policy analysis · Causal inference · Mixed methods

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10742-019-00201-8>) contains supplementary material, which is available to authorized users.

✉ Alan Zarychta
azarychta@uchicago.edu

¹ University of Chicago, Chicago, USA

² University of Colorado Boulder, Boulder, USA

³ The Ohio State University, Columbus, USA

⁴ Purdue University, West Lafayette, USA

1 Introduction

Considerable advances in developing more and more sophisticated impact evaluation methodologies have occurred over the last 15 years as part of the results agenda in international development and health services delivery (e.g., Pearl 2001; Dixit 2007; Morgan and Winship 2007; Rubin 2008; Imbens and Wooldridge 2009; Tannahill and Kelly 2013; White 2013; Jung and Pirog 2014; Bamberger 2015; and Imbens and Rubin 2015). Nonetheless, substantive policy studies continue to suffer from persistent challenges in achieving what can be called the impact evaluation trifecta: causal identification, isolation of key causal mechanisms, and substantive influence on the policy process. Many studies rightfully claim they have one or perhaps two of these features, but only a small number are able to achieve all three in a single context.¹ On account of this, there exist relatively few comprehensive examples of how all three components can be systematically developed into a study's design and then implemented ahead of data collection, during fieldwork, alongside analysis, and through iterated dissemination of results.

The primary goal of this paper is to present our experiences developing and implementing a systematic approach for impact evaluation that jointly addresses each of the three objectives of impact evaluation: causal identification through a *quasi-experimental* research design, isolation of causal mechanisms using diverse sources of original data and *multi-method* analysis, and policy influence through a *participatory* study that directly involves policymakers throughout the research. In this article and its extended supplemental appendix, we present our comprehensive study design for assessing health sector decentralization reform in Honduras as a way to illustrate the quasi-experimental, multi-method, and participatory approach that guided the research. We also highlight innovations with respect to each of the persistent challenges using examples from our experiences carrying out that project. Toward those ends, we proceed by first discussing the major challenges of evaluating governance reforms in developing countries; next, we describe our quasi-experimental, multi-method, and participatory approach in general terms; and then, we illustrate that approach using examples from our work developing and implementing a research design to assess the effects of health sector reform in Honduras, both a critically important pilot study and subsequent nationwide study. In our discussion of these examples and in the conclusion we emphasize specific lessons and guidance that we hope can be helpful for researchers beginning their own studies of health services delivery in developing countries.

2 Challenges in evaluating governance reforms

The recurring example we utilize in this paper is health sector decentralization in Honduras. The Honduran Ministry of Health (MOH) has implemented decentralized health service delivery models at the regional and local levels since about 2007. This model combines two key changes in an effort to improve the health services reaching rural and underserved communities: first, it delegates key administrative functions over local health centers, like planning, operations, and supervision, from Regional Health Authorities to

¹ The presence of major evaluation institutions, like the International Initiative for Impact Evaluation (3ie), Innovations for Poverty Action (IPA), and the Abdul Latif Jameel Poverty Action Lab (J-Pal), has helped in bridging gaps between scholars and policymakers.

local governments or community-based organizations using performance-based contracts; and second, it shifts oversight of these newly empowered intermediary organizations from the national MOH to the Regional Health Authorities (MOH 2009, 2010).² We study this example because it is a public sector governance reform that typifies the common types of challenges researchers must confront when working to achieve the three major objectives of impact evaluation: (1) identifying effects, (2) isolating causal mechanisms, and (3) influencing future policy.

2.1 Identifying the effects of decentralization

Identifying the effects of governance reforms, like decentralization, in developing countries is particularly challenging for three reasons: (1) the ability to intervene and the choice of intervention are often shaped by a combination of domestic and external politics, (2) the assignment mechanism for these interventions can be opaque to outside observers, and (3) data limitations are often severe. This means that it can be very difficult for researchers to persuade governments to randomize the targeting of these types of reforms, and even if a randomized design is agreed upon, there are numerous political barriers that may still come up in the course of implementation that can undermine the study (King et al. 2007). As a result, these challenges often end up structuring researchers' decision-making processes as they confront tradeoffs among different quasi-experimental research designs and weigh the competing limitations of several reasonable choices, as illustrated throughout this paper.

The first challenge to identifying the effects of governance reforms is related to a particular kind of selection problem: the risk that there are background factors and potentially unobserved political processes affecting the decision to implement governance reform in the first place, or the variety of reform that is or is not tenable, and the outcomes that a specific reform aims to influence. The second challenge concerns potential selection effects stemming from treatment assignment. In this case, one worries that the criteria for intervening end up restricting the sample of subnational units and thus introducing bias. For example, political dynamics might dictate that decentralization reforms are implemented only in municipalities with influential mayors or based on patterns of support for national politicians (O'Neill 2003).

One possible approach to these sometimes-related selection problems is to directly confront the limitations of the knowledge likely to be gained from a given case. One common version of this involves explicitly identifying the scope conditions for one's results based on relevant characteristics of the setting or policy context that allowed for a particular type of reform in the first place. A second common version of this relates not to the overall case context, but instead seeks to identify known restrictions around the sub-population of administrative units or individuals targeted for reform, and analyzes how those restrictions influence the applicability or generalizability of the study's findings. Under both options, it is important that researchers remain open to the possibility that severe limits on scope conditions or severe restrictions with respect to the target population could outweigh the potential value of the findings, in which case an alternative design or entirely different study may be warranted. Qualitative interviews with key decision-makers and participant observation of the ongoing reform process *prior to finalizing one's research design* can

² A full description of the reform, as well as the MOH's priorities, is included in the supplemental appendix.

prove invaluable in revealing these types of issues. In our case, part of the rationale for a sub-regional design in the pilot study we conducted was based on the fact that there was strong commitment at both national and regional levels of government to decentralize all health centers in a particular state, but at the time of our initial field research only about half of those units had received the reform.³

Expanding beyond localized designs of this type, an alternative approach is to utilize the insights gained from the aforementioned interviews and participant observation in an effort to model the treatment decision by estimating a propensity score and using it to select a matched set of treated and control units for additional data collection (Rosenbaum and Rubin 1983; Rubin 2001; Fu et al. 2007; O'Neill et al. 2016). Matching methods can be particularly strong when used in tandem with other inferential approaches and when being able to draw on temporal data, for example the increasingly common strategy of pre-processing a given sample through matching and then employing a difference-in-differences analysis (Kreif et al. 2013; Wright et al. 2016), which we do as part of our nationwide study design. While matching methods can be useful in many circumstances, they are nonetheless limited to addressing observable confounders, do not alone remove all potential sources of bias, and tend to be data-intensive (Sekhon 2009; Lindner and McConnell 2018). Thus, they run up against the third challenge that data limitations can be quite severe in many developing countries. Instrumental variable approaches are frequently suggested as a useful alternative (Posner et al. 2001). This strategy is theoretically appealing, but in practice it often proves difficult to find a convincing instrument, namely a variable that predicts treatment but has no direct effect on the outcomes of interest. Finally, regression discontinuity designs that take advantage of threshold rules for treatment assignment can be effective if applicable, but such circumstances can be rare, and stated rules and actual practices may deviate in important ways. Furthermore, observations near a given threshold are not necessarily representative of the broader population, thereby raising additional challenges with respect to external validity. These latter points are of particular concern in less-developed settings where clientelism or outright corruption often define citizens' relationships to the public officials and state bureaucrats making these assignment decisions (Stokes et al. 2013).

2.2 Isolating causal mechanisms for the effects of decentralization

Isolating the causal mechanisms underlying the effects of public sector governance reforms like decentralization carries all of the challenges of identifying effects—identification is naturally a prerequisite for considering causal mechanisms—as well as its own additional challenges. First, the most pronounced common challenge of identification and mechanisms relates to data availability. From a quantitative perspective, causal mediation analysis is considered the gold standard. However, both the traditional approach and updated versions carry significant assumptions and nontrivial data requirements (Baron and Kenny 1986; Imai et al. 2011). As scarcity, rather than abundance, is the norm in terms of high quality data in many developing countries, this approach often proves untenable.

As a result, many researchers turn to mixed methods and bring in qualitative data, causal process tracing, or behavioral games to complement their quantitative analyses (Bennett

³ As of late-2016, all health centers in this state have been decentralized under the reform.

2008; Habyarimana et al. 2009; Brady 2010).⁴ This carries with it potential advantages as well as risks. On the positive side, nuanced contextual information about particular cases within one's sample can dramatically improve the overall causal argument and provides details and examples that clarify the processes at play in a given social situation. Similarly, behavioral games can be used to simulate the relevant aspects of a real world setting in a more controlled environment, thereby allowing the researcher to overcome some of the challenges of identifying mechanisms with observational data. While games can be limited in terms of external validity, their combination with qualitative methods allows researchers to triangulate different sources of information to achieve a more trustworthy assessment of causal mechanisms. To pair these approaches well, however, researchers need to be competent in the diverse methods they plan to use, able to think and talk about evidence from multiple methodological perspectives, and effectively incorporate all of their techniques and data sources in one holistic analysis. Some common shortcomings in these cases are that the qualitative component of the study is simply tacked-on in the latter stages of analysis, or that the logistical challenges of behavioral games and experiments serve to isolate that part of the study, both of which create missed opportunities for integration into the overall research design.

2.3 Influencing the policy process

Finally, if researchers are able to identify causal effects and convincingly isolate causal mechanisms, communicating results to policymakers in a way that they understand and which influences the policy process often proves difficult. This is partly because reward systems (and sanctioning systems) in both academia and in government make participatory design and implementation of policy evaluation difficult to achieve, despite the interests of researchers in having their work taken into account. It is also partly because it remains unclear what kinds of evidence and what modes of presentation are most likely to increase policy impact under different circumstances, as the growing literature on the knowledge-to-action (KTA) gap demonstrates (Cash et al. 2003; Brownson et al. 2018).

From the perspective of researchers, their universities and institutes tend to reward them primarily on the basis of their publication and grant records. Engaging with government agencies can be costly in terms of time and effort, meaning that a significant barrier to such engagement is researchers' perceived opportunity costs. If they spend a greater share of their time working with government officials, they may believe that they will produce fewer publications and receive fewer research grants, which will hamper their career advancement. From a government perspective, working with academics may represent a significant distraction from their normal routines, which are often driven by the on-time fulfilment of planned deliverables (products, reports, spending targets, etc.). Unless government officials are convinced that researchers can help them achieve these obligations, they are not likely to invite academics to cooperate. And even if these initial motivation problems can be overcome—perhaps as a result of both research and government organizations deciding to support and reward such cooperation—the engagement will not necessarily be productive in the short term. Social science research on “usable science” in policy-making has

⁴ Bennett (2008) defines causal process tracing as an analytical technique that examines “...evidence within an individual case, or a temporally and spatially bound instance of a specified phenomenon, to derive and/or test alternative explanations of that case...” (p. 704). See also Collier (2011) and Ricks and Liu (2018).

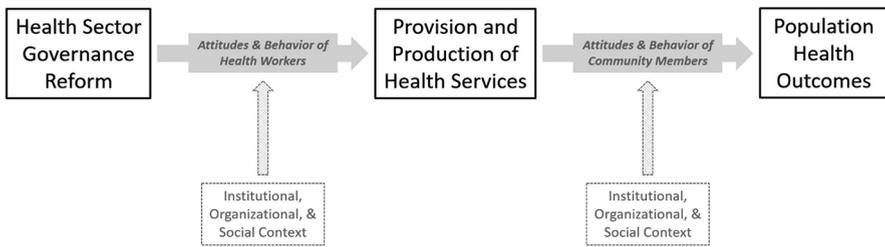


Fig. 1 Linking governance reform, health services delivery and population health outcomes

shown scientific findings are more likely to be used when decision-makers perceive these to be relevant, credible, and legitimate (Cash et al. 2002; Matson et al. 2016). This research has also shown that the key to fulfilling these expectations lies in the ability of both scientists and decision-makers to create spaces for an open and participatory process from the initial stages of the project (Clark et al. 2016). As challenging as it may be to create these spaces, without their existence, participatory policy evaluation with substantive influence is not likely to occur.

The implications are twofold. First, junior researchers should be particularly mindful of the heavy up-front costs of developing these types of collaborations that may come at the expense of early publications, and the need to actively communicate and justify the value of these investments to their senior colleagues. But second, researchers can also view these challenges as opportunities for contributing to the KTA literature by building assessments of different dissemination strategies into their broader research designs, as we have done in our work with the MOH in Honduras.

3 A quasi-experimental, multi-method, and participatory approach for assessing the health impacts of governance reforms

The preceding discussion highlighted the challenges of evaluating governance reforms like decentralization in less-developed countries. In our case, we utilized a quasi-experimental research design coupled with multi-method and participatory data collection to assess the impacts of decentralization on health services delivery in Honduras. The study design we present fully in the appendix and summarize here provides guidance and examples for how researchers can overcome many of these persistent challenges in evaluating governance reforms. In the following sections we elaborate on the three general characteristics of our methodological approach that directly respond to those challenges. Specifically, we present an approach to research design that is (1) quasi-experimental, addressing causal identification; (2) multi-method, addressing causal mechanisms; and (3) participatory, addressing policy influence as well as assisting in the first two challenges. In the subsequent section, we highlight several important lessons learned from using that approach to develop a comprehensive research design and then carrying it out through fieldwork in Honduras.

Our overall approach and the particular research design described here both stem from an underlying framework linking governance reform, health services delivery, and population health outcomes, which is shown in Fig. 1. This framework is informed by the more general Institutional Analysis and Development (IAD) framework and elaborated to apply

to the present context: governance reform of health services delivery in developing countries (Ostrom 2005; Andersson 2006; Oakerson and Parks 2011). The framework highlights a number of features related to how governance reform may be able to influence population health that guide the development of our research design.

First, the proximate consequences of governance reform will be observable in the provision and production of health services, that is to say, the policy choices over what services to make available and the subsequent creation and delivery of those services (Ostrom et al. 1961). Second, health services delivery is identified as a primary mediating factor between governance reform and health outcomes at the population or community levels; an important part of the effect of governance reform on population health outcomes is expected to come through changes in the provision and production of health services for that population. Third, while beginning with a meso- or system-level perspective, this framework also focuses attention on the key individuals in those systems: health workers who are often targets of the reform and who actually deliver or administer services at the community level, and individual members of those communities who access and receive health services (or not). In short, the attitudes, beliefs, and behaviors of these individuals are also critical channels through which system-level effects of governance reform may be realized. Lastly, the framework points out that both the overall system and its individual actors are situated within institutional, organizational, and social contexts that also moderate the ways in which governance reform, services, and health outcomes are linked together, most notably in how those contexts shape the behavioral responses of health workers and community members. The research design described in the next sections, and elaborated fully in the appendix, is oriented to assess the specific direct effects, mechanisms, and moderating conditions as shown in Fig. 1. Throughout the paper and appendix, we include details of our design choices and field research in order to highlight the numerous practical considerations underlying a successful evaluation study, issues we believe are critical yet underemphasized in the public health, public policy, and economics literatures on health services delivery and impact evaluation.⁵

3.1 Quasi-experimental

We implemented a quasi-experimental approach in our research design, namely a natural experiment in the pilot study and propensity score matching for the nationwide study. In both cases, we paired those design choices with a difference-in-differences (DID) analysis, which utilize variation in treatment assignment across subnational units and years to identify the effects of decentralization on health services and outcomes. Specifically, we take advantage of the fact that the Honduran MOH has engaged in a staggered and ongoing implementation of decentralized health services delivery models since 2007 where about one-third of the country's municipalities have been decentralized over the last 10–12 years. As a result, there is pre-post data available for decentralized municipalities, as well as a considerable number of municipalities under centralized administration available to serve as a natural or matched control group (Dunning 2012). Moreover, this staggered rollout of the reform also provides important opportunities for targeted and efficient original data collection to better characterize outcomes and help in the identification of causal mechanisms.

⁵ Dunning (2012) and Kapiszewski et al. (2014) are notable exceptions.

3.2 Multi-method

Given the challenges described previously, we draw on multiple and diverse sources of information to assess decentralization reform in Honduras. For our main analyses, we compiled existing health services data for all health centers in one state for the pilot study, and then for all municipalities nationwide, as well as data on the demographic, socio-economic, and institutional contexts characterizing the population across municipalities. Furthermore, we conducted an original household survey for a nationwide sample of municipalities to complement the existing administrative data, building on the fact that there were two preceding waves of the Demographic and Health Surveys (DHS) in Honduras.

In addition to our primary focus on compiling and collecting longitudinal data on health, society, and institutions; we also collected original survey data for households, health workers, health facilities, and other organizations; implemented a set of behavioral games and experiments (sometimes referred to just as experimental games) with health workers; and conducted semi-structured interviews and focus groups with key health administrators to investigate and isolate the mechanisms underlying the effectiveness or ineffectiveness of decentralization reform. Together, the information from these multiple sources help us to triangulate and confirm our main causal inferences about the effects of the reform on health services and outcomes, and enable us to identify the mediating factors linking reform to outcomes, as well as the contextual conditions that moderate those relationships. The numerous factors and conditions that jointly contribute to local health system performance necessitate a multi-method approach of this type in order to arrive at sound conclusions that can be utilized in the policy process.

3.3 Participatory

In our own evaluation work in Honduras, several enabling factors combined to allow us to overcome common barriers and be able to engage in a participatory policy evaluation process with local and national health authorities. From the researchers' perspective: we were interested in getting access to government data that was not in the public domain, as well as decision-makers' perspectives on motivations for promoting the reform. These two things encouraged us to pursue an open dialogue with officials from the Ministry of Health, starting locally during the pilot study and then working our way up the hierarchy as we built and reinforced trust with key linking agents along the way, those officials that could operate across levels from local to regional or regional to national. Through these dialogues with the personnel at the MOH, we learned that they were under pressure from the external donors to produce robust evidence on the effects of the policy reform, which they requested our team to help them with. What we first perceived as a very difficult barrier to overcome turned out to largely dissolve once we had identified compatible mutual interests in conducting a study to assess the impact of the decentralization reform. Through repeated interactions in pursuing identified mutual interests and tasks, we have developed a close partnership based on trust with several individuals in the MOH that has now lasted across multiple government administrations. This partnership with MOH personnel helped us gain access to restricted government data and facilitated the implementation of field-based research, and we have provided them with results relevant to their decision-making process. Long-term and close collaboration with the implementing agency, the MOH and its Regional Health Authorities, was pivotal for the design and execution of our study.

Maintaining regular communication with staff at the MOH while finalizing the study design and developing protocols was also critical for gaining the trust and buy-in of key officials, and ultimately for being able to conduct an evaluation study that was seen as legitimate. Staff members in the statistics department were invaluable in helping us obtain, understand, and organize monthly health services data for all health centers in the country between 2006 and 2017. Administrators from the MOH also provided contextual information that helped us understand how decentralization was implemented in a staggered fashion over several years. Additionally, members of the leadership team at different Regional Health Authorities were instrumental in promoting the project with health personnel; on several occasions they personally reached out to staff in local health centers in ways that legitimized our research and engendered cooperation in data collection.

We returned to Honduras each year between 2013 and 2019 to present results from the pilot study evaluation to the Regional Health Authority's leadership team, as well as preliminary findings from the nationwide study to key decision-makers in the central MOH and other stakeholders working in the Honduran health sector; the comments received during these presentations have informed revisions and additional analyses on our part. Building the relationships necessary for this type of collaboration takes time and effort by researchers, and it is not without problems, but in this case our investments paid off in intellectual and material support for the research. Finally, being viewed as legitimate by the implementing agency, and having access to the key decision-makers, has been integral for our findings to have influence on future policy decisions.

4 Illustrations from our research design for assessing the public health impacts of decentralization reform in Honduras

Structuring the development of our research design using this quasi-experimental, multi-method, and participatory approach enabled us to collect the diverse types of information needed to assess decentralization's effects on health services delivery and health outcomes in the Honduran context. What we describe here is a comprehensive research design supported with considerable external funding and carried out over multiple years. While this extent of work will not necessarily be possible in all cases, we nonetheless see the value of presenting our full research design so that other researchers can better weigh tradeoffs between competing design choices in their own studies.

Table 1 summarizes our overall research design pairing a pilot study in one state with a sample-based nationwide study, and drawing on both existing and original data in both cases. Whereas the full details of all components of this study design are provided in the appendix, here we highlight a number of key factors that were important in overcoming the persistent challenges of evaluating governance reforms, and which can also serve as relevant guidance for researchers planning their own studies. First, despite the fact that the work of developing and sustaining in-country partnerships is not often rewarded in academia as discussed earlier, this work can prove invaluable to carrying out a valid and impactful evaluation study. Unlike many researchers who appropriately draw on their connections to national-level policymakers or international organizations with sway domestically, this project began from the bottom-up: the lead researcher developed the relationship with the leadership team of a state-level Regional Health Authority and persuaded them of the value of our pilot study. Without their buy-in and support in 2012, it is unlikely that the rest of the study would have been possible. In the pilot study, the lead researcher utilized a

Table 1 Overall research design for assessing the effects of decentralization reform on health services delivery and population health outcomes in Honduras

Data source	Sample	Time period	Components
<i>Pilot study—natural experiment and difference-in-differences</i>			
Administrative data	50 health centers One state	2005–2012	Health services and reform timing
Contextual data	17 municipalities One state	2005–2012	Demographic, institutional, social, and political characteristics
Health center survey	50 health centers One state ~200 health workers	2012	Attitudes, beliefs, perceptions, and behaviors
Interviews and participant observation	2 months in Intibucá 10 formal interviews	2012	Attitudes, beliefs, perceptions, reform experiences, and other comments
<i>Nationwide study—propensity score matching and difference-in-differences</i>			
Administrative data	298 municipalities 18 states	2006–2017	Health services and reform timing
Contextual data	298 municipalities 18 states	2006–2017	Demographic, institutional, social, and political characteristics
Household survey	65 municipalities 14 states ~9000 households	2016 [DHS 2005– 2006 and 2011–2012]	Health services, health outcomes, demographics, perceptions, social and political characteristics
Health center survey	65 municipalities 14 states ~300 health centers ~600 health workers	2016	Attitudes, beliefs, perceptions, and behaviors of health workers; conditions, supplies, and resources of facilities
Inter-organizational network survey	65 municipalities 14 states ~800 organizations	2016	Organizational and relational factors
Behavioral games and experiments	31 municipalities 14 states ~230 health workers	2017	Behaviors, relationships, perceptions, and attitudes

Table 1 (continued)

Data source	Sample	Time period	Components
Focus groups	31 municipalities 14 states ~230 health workers	2017	Attitudes, beliefs, perceptions, reform experiences, and other comments
Knowledge-to-action survey experiment	Nationwide ~250 health workers	2018	Factors influencing uptake and use of information in policymaking and practice
Qualitative case studies	18 municipalities 10 states	2019	Confirmatory evidence wrt identified relationships and mechanisms; exploratory evidence wrt additional important factors

sub-regional design that leveraged an existing natural experiment within a largely homogenous state, collected existing administrative and contextual data, conducted an original survey of about 200 health workers in the state, and engaged in extended qualitative data collection within the Regional Health Authority and across different decentralized managing organizations. While the substantive findings of the pilot study are important in and of themselves (as presented in Zarychta 2018), the reliance on administrative records and one original survey of health workers limits that work in two important ways: it necessarily focuses on health services over population health outcomes, and is restricted in its ability to consider moderating conditions as the strength of the identification strategy comes at the expense of variation on those conditions. In short, the pilot study calls out the need for a scaled-up version of this research at the national level.

In terms of the goal of designing a nationwide study, the pilot study benefited us in two additional and critical ways: it gave us experience with the numerous practical considerations that inevitably come up in field research abroad and it gave us legitimacy vis-à-vis the national level policy makers in the MOH whose support would be needed for a nationwide study, as well as the frontline service providers who would be participants in our study and potential users of the findings from this work. Both of these factors made it more likely for us to gain the buy-in of these key gatekeepers and constituencies for a scaled-up, nationwide evaluation study, and for us to be able to successfully carry out the extensive and diverse data collection efforts we developed for that nationwide study. While the barriers to developing these types of partnerships can be substantial, and relying on top-down relationships may work as a shortcut in some cases, in our experience the bottom-up approach improved the quality of our substantive research and created the foundation needed for that work to be meaningfully used within the policy process.

Second, it is critical to confirm evaluation results using multiple data sources, particularly when relying on quasi-experimental designs and administrative data. All quasi-experimental research designs and all data collection strategies have limitations, whether that is difference-in-differences, instrumental variables, regression discontinuity, or matching, or whether that is administrative data, surveys, behavioral games, or experiments. Given the various factors, some observable and others less so, that often shape the targeting of governance reforms, and the limited availability of high-quality data in less developed countries, it is paramount that researchers draw on multiple, distinct information sources prior to finalizing the conclusions of any evaluation in this type of setting. In our case, we were intentional in the nationwide study to compile any and all existing data to understand assignment to the decentralization treatment, and to design data collection efforts so that we would have complementary sources of information for our key outcomes on health services delivery and population health.

In terms of assignment to treatment, we first conducted a set of interviews with policy-makers in the MOH who were present and involved in the initial development and rollout of the reform to better understand the factors they were considering at that time, as well as the information they had available to them. Based on that, the lead researcher spent months over multiple years requesting, compiling, and integrating data on all demographic, social, and political factors that could have influenced reform targeting to be used in the propensity score matching analysis to choose our sample of municipalities for original data collection in the nationwide study. While a small portion of these data was available online, the vast majority required in-person visits to ministries, institutes, and centers in the capital city of Honduras, as well as extended email and phone follow-up, to collect. Through this work, and following the Imbens and Rubin (2015) approach to propensity score matching, we were able to identify a reasonably well-balanced sample of 65 municipalities (42

decentralized, 23 centrally-administered) where we conducted original household, health center, and inter-organizational network surveys. The details of this matching analysis, assessments of balance, and the identification of a reduced sample of 50 municipalities as a robustness check are included in the appendix.

With respect to triangulating findings across different types of data, we rely on both administrative records and original survey data for the main health services and population health outcomes in our research, which are based on the MOH's own priorities as discussed in the appendix. Administrative data play a large and appropriate role in policy evaluation and health services and outcomes research (Normand et al. 2007; Blume et al. 2014), but especially in less-developed country settings, the production of these data may be subject to some of the same weaknesses in state capacity that motivate governance reform efforts in the first place. The lead researcher here conducted extended participant observation in the statistics office of one Regional Health Authority, the unit that receives and checks reports from the primary care health centers before sending them on to the central offices of the MOH, as well as multiple interviews with staff in the central information management unit that compiles and enters those reports into its databases. These efforts helped us understand the nature of the reporting process within the Honduran health system, and thereby identifying a set of indicators that were consistently reported over time and developing our own quality control checks for the administrative data. Furthermore, we took advantage of two conveniently timed prior applications of the Demographic and Health Surveys (DHS) in Honduras: 2005–2006, just before the initial rollout of the reform in 2007, and 2011–2012, about 2–3 years after the early rollout period. Given this, we conducted our own household survey as a comparable, somewhat shortened, version of the DHS applied in the nationwide sample of 65 municipalities selected through matching. This gave us three periods of health services indicators that complement the administrative data, and a set of population health indicators that are expected to be linked to improvements in services delivery as displayed in Fig. 1. Identifying this existing data source and building on it was a major advantage of our work as it allowed us to move from a limited cross-sectional design to having three periods of high-quality data, before and after the reform, over 10 years.

Third, it is important to consider hypothesized mechanisms and moderating conditions at the outset of the research design process, and to make efforts to tailor data collection instruments accordingly. In addition to conceptualizing health services delivery as a mediating factor between governance reform and population health outcomes, and beyond the standard contextual characteristics we account for, we focused particularly on the role of behavior in our research design. Specifically, we hypothesized that the attitudes, beliefs, and behaviors of health workers, including administrators and public officials active in the health sector, constitute a key link between governance reform and health services delivery. And, we considered the attitudes, beliefs, and behaviors of community members to similarly constitute an important channel connecting health services delivery to population health outcomes. Based on this, we developed traditional surveys and behavioral games and experiments to help characterize aspects of these causal mechanisms, all nested within the previously described quasi-experimental research design. These data collection elements consisted of a health center survey (about 300 health centers and 600 health workers across 65 municipalities), an inter-organizational network survey (more than 800 organizations across 65 municipalities), a set of behavioral games and experiments (day-long workshops convened in a sub-sample of 31 municipalities with about 230 health workers), as well as supplemental modules and a list experiment in the household survey (about 9000 households across 65 municipalities).

The behavioral games and experiments were a particularly useful approach for characterizing these types of social mechanisms that have only recently begun to be incorporated in public administration and health services research (Grimmelikhuijsen et al. 2017; Moynihan 2018). In a sub-sample of 31 municipalities, chosen based on the previous matching approach with restrictions due to the composition of participants from different levels of government needed for the games and resource constraints (also described in the appendix), we convened daylong workshops with systematic samples of 5–8 doctors, nurses, administrators, and municipal officials. Per our earlier discussion, partnership with the MOH was instrumental in supporting the recruitment and scheduling work our field staff did to carry out these workshops. Once convened, these health workers engaged in a series of structured and incentivized activities based on different behavioral mechanisms potentially linking governance reform to health services delivery: for example, cooperation through a public goods game, communication and coordination in a puzzle game, organizational trust through a modified investment game, effort in a framed effort task, and self-sacrifice as a key aspect of public service motivation in a modified dictator game. Additionally, we engaged these public officials in a guided focus group discussion where they reflected on their experiences in the preceding activities and were able to communicate their views on if and how the reform works. We concluded the day with a capacity-building session on teamwork and leadership in health services administration, a topic linked to the workshop activities and also of interest to our MOH collaborators. While the other survey efforts are detailed in the appendix, we highlight the behavioral games as a novel approach that allowed us to gain improved measures of behavioral mechanisms that can be difficult to elicit in traditional surveys, and because these were structured in a way that allowed us to engage with and hear the views of a substantial number of health workers, administrators, and public officials charged with serving the health needs of communities across Honduras.

5 Conclusion

We have presented a quasi-experimental, multi-method, and participatory approach for evaluating the effects of public sector governance reforms, and illustrated that approach using examples from our ongoing research on health services delivery in Honduras. Moreover, we have leveraged a natural experiment in decentralization, conducted extensive fieldwork, and collected substantial original data at the local and national levels. Together, these efforts allowed us to highlight our research design choices and associated strategies for addressing several key methodological and practical challenges that confront policy evaluation research—from planning, to implementation, and through analysis.

While informative, this case presentation has limitations both in terms of the design presented and in terms of serving as a guide for other research. First, we prioritize internal validity of the causal identification strategy by focusing on one region in the pilot study and one country in the nationwide study, which is done at some expense to generalizability. Second, while we have tried to highlight important lessons that would apply to research design choices in other circumstances, and we argue our approach and illustrations can serve as valuable guidance for other contexts, individual researchers will ultimately need to weigh tradeoffs in their own settings and may come to different conclusions given the constraints they face. Nonetheless, it is our hope that through the detailed and transparent discussion both in the main paper and in the supplemental

appendix, those researchers will be better informed when making their own tradeoffs. Third, as this paper's goal is to present and illustrate a methodological approach and research design, we are unable to discuss a full set of health services results and thus cannot speak to the broader health sector decentralization literature in this paper.

From the examples presented here, we conclude by drawing five main lessons concerning the evaluation of health sector decentralization specifically, and public sector governance reforms in developing countries more generally. We do not view this as an exact template to follow step-by-step, but rather a set of considerations and best practices drawn from our experiences and informed by the policy analysis literature, which researchers may apply or adapt to the specific contexts in which they are planning to work.

- (1) *Collaboration with the implementing agency is incredibly important in studying the effects of policy reforms in developing countries.* Without participation and buy-in from the Regional Health Authority and MOH our research would not have been possible. Our relationships with staff at one Regional Health Authority are what led to us to discover sub-regional variation in decentralization. Furthermore, this collaboration allowed us to design a study that was legitimate in the eyes of administrators, thereby increasing cooperation from staff in the local health centers. The commitment of the MOH to this evaluation coupled with external funding gave us access to multiple data sources that were not publically available while also allowing us independence, both of which were necessary for reaching policy-relevant conclusions. Most importantly, this early buy-in has continued to give us unparalleled access to decision-makers where we expect our results to have resonance.
- (2) *Leveraging subnational variation in policy implementation can be a useful strategy for evaluating governance reforms.* Given the many challenges for identifying the effects of changes in governance structure, and the limitations of utilizing experiments or randomized trials in this context, being both creative and strategic in research design is critical for reaching sound inferences. Moreover, as we illustrated here, it can be worthwhile for researchers to search out subnational and local variation in implementation to study the effects of policy reforms in developing countries, as well as to pilot test design choices and data collection strategies before scaling up research to regional, national, or cross-national levels.
- (3) *In line with Rubin (2001, 2008), research design is more important than analysis for objective causal inference with respect to governance reforms.* In short, no amount of statistical analysis will enable one to extract valid treatment effects if the initial research design is flawed. Therefore, it is critical to think about counterfactuals, potential exogenous sources of variation, confounding variables, and overlap in characteristics between treated and control groups prior to data collection. Local knowledge, like the kind we gained from qualitative interviews with administrators, is crucial for characterizing how units were assigned to treatment. Furthermore, longitudinal data and extensive fieldwork are almost certainly prerequisites for evaluating the effects of governance reforms in developing countries.
- (4) *Context-sensitive approaches and multi-method data collection remain important for isolating and understanding causal mechanisms.* Within any theory of social change, specific mechanisms will be hypothesized to generate observable implications, and quantitative data may often be best suited for testing those hypotheses. Understanding if the underlying mechanisms actually operate as specified, however, requires some

combination of the types of original survey work, semi-structured interviews, participant observation, and behavioral games we illustrated here. Merging these multiple sources of information within a single evaluation study increases the likelihood of reaching sound conclusions.

- (5) *There is necessarily a tradeoff between causal identification at the subnational level and the ability to generalize.* In this paper we have presented an approach for context-sensitive causal inference about health sector decentralization. Our approach trades off some generalizability for valid estimates of treatment effects, particularly at the pilot study phase. Nonetheless, we suggest that using insights gained from this type of local, cost-effective research is a productive way to identify promising evaluation strategies that can be utilized in national or cross-national research with greater emphasis on generalization, as we did in our nationwide study. Furthermore, focusing on moderators and mechanisms also helps address concerns about generalizability, as both emphasize the conditions and pathways that need to be in place for an identified relationship to exist in other settings or at different scales.

The systematic approach we have detailed here addresses many of the persistent methodological, analytical, and practical challenges in policy evaluation, and is relevant to a broad range of research on public sector governance and health services and outcomes in developing countries. Additional research identifying scientifically sound, practical, and cost-effective strategies for evaluating governance reforms will be valuable in furthering the substantive goal of these interventions: improving the livelihoods of vulnerable and marginalized communities around the world. By better identifying what works and under what conditions, policymakers will be able to more effectively target scarce public resources. Toward these ends, we believe that future impact evaluation research should focus on better elaborating strategies for linking causal inference with causal mechanisms, such as using diverse quantitative and qualitative data to test ancillary hypotheses derived from well-defined theories of social change, as well as additional research on how best to incorporate survey experiments and behavioral games as components of broader research designs.

Funding This project was completed with financial support from the National Science Foundation (Award Numbers DGE-1144083 & SMA-1328688), Social Science Research Council, University of Colorado Boulder, and University of Chicago. We are especially grateful for the support and assistance we received from staff at the Ministry of Health in Honduras and the Regional Health Authority of Intibucá. All errors and omissions are our own.

Compliance with ethical standards

Conflict of interest The authors declare they have no conflicts of interest.

Ethical approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Approval for the research was received from the University of Colorado Boulder Institutional Review Board (Protocol #12-0318).

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Andersson, K.: Understanding decentralized forest governance: an application of the institutional analysis and development framework. *Sustain. Sci. Pract. Policy* **2**(1), 25–35 (2006)
- Bamberger, M.: Innovations in the use of mixed methods in real-world evaluation. *J. Dev. Eff.* **7**(3), 317–326 (2015)
- Baron, R.M., Kenny, D.A.: The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**(6), 1173–1182 (1986)
- Bennett, A.: Process tracing: a Bayesian approach. In: Box-Steffensmeier, J., Brady, H., Collier, D. (eds.) *Oxford Handbook of Political Methodology*. Oxford University Press, Oxford (2008)
- Blume, G., Scott, T., Pirog, M.: Empirical innovations in policy analysis. *Policy Stud. J.* **42**(S1), S33–S50 (2014)
- Brownson, R.C., et al.: Getting the word out: new approaches for disseminating public health science. *J. Public Health Manag. Pract.* **24**(2), 102 (2018)
- Brady, H.: Data-set observations versus causal-process observations: the 2000 U.S. presidential election. In: Brady, H., Collier, D. (eds.) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd edn. Rowman & Littlefield Publishers, Lanham (2010)
- Cash, D., et al.: *Saliency, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making*. Social Science Research Network, Rochester. SSRN Scholarly Paper. <https://papers.ssrn.com/abstract=372280> (October 9, 2018) (2002)
- Cash, D.W., et al.: Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci.* **100**(14), 8086–8091 (2003)
- Clark, W.C., van Kerkhoff, L., Lebel, L., Gallopin, G.C.: Crafting usable knowledge for sustainable development. *Proc. Natl. Acad. Sci.* **113**(17), 4570–4578 (2016)
- Collier, D.: Understanding process tracing. *PS Polit. Sci. Polit.* **44**(4), 823–830 (2011)
- Dixit, A.: Evaluating recipes for development success. *World Bank Res. Obs.* **22**(2), 131–157 (2007). <https://doi.org/10.1093/wbro/lkm005>
- Dunning, T.: *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press, New York (2012)
- Fu, A.Z., Dow, W.H., Liu, G.G.: Propensity score and difference-in-difference methods: a study of second-generation antidepressant use in patients with bipolar disorder. *Health Serv. Outcomes Res. Methodol.* **7**(1–2), 23–38 (2007)
- Grimmelikhuisen, S., Jilke, S., Olsen, A.L., Tummers, L.: Behavioral public administration: combining insights from public administration and psychology. *Public Adm. Rev.* **77**(1), 45–56 (2017)
- Habyarimana, J., Humphreys, M., Posner, D.N., Weinstein, J.M.: *Coethnicity: Diversity and the Dilemmas of Collective Action*. Russell Sage Foundation, New York (2009)
- Imai, K., Keele, L., Tingley, D., Yamamoto, T.: Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *Am. Polit. Sci. Rev.* **105**(04), 765–789 (2011)
- Imbens, G.W., Rubin, D.B.: *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, New York (2015)
- Imbens, G.W., Wooldridge, J.M.: Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* **47**(1), 5–86 (2009)
- Jung, H., Pirog, M.A.: What works best and when: accounting for multiple sources of pure selection bias in program evaluations. *J. Policy Anal. Manag.* **33**, 752–777 (2014). <https://doi.org/10.1002/pam.21764>
- Kapiszewski, D., MacLean, L., Read, B.: *Field Research in Political Science*. Cambridge University Press, New York (2014)
- King, G., et al.: A ‘politically robust’ experimental design for public policy evaluation, with application to the Mexican universal health insurance program. *J. Policy Anal. Manag.* **26**(3), 479–506 (2007)
- Kreif, N., Grieve, R., Radice, R., Sekhon, J.S.: Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. *Health Serv. Outcomes Res. Methodol.* **13**(2), 174–202 (2013)
- Lindner, S., John McConnell, K.: Difference-in-differences and matching on outcomes: a tale of two unobservables. *Health Serv. Outcomes Res. Methodol.* (2018). <https://doi.org/10.1007/s10742-018-0189-0>
- Matson, P., Clark, W.C., Andersson, K.: *Pursuing Sustainability: A Guide to the Science and Practice*, 1st edn. Princeton University Press, Princeton (2016)
- Ministry of Health (MOH), Government of Honduras: *Marco Conceptual Político Y Estratégico de La Reforma Del Sector de Salud* (2009)
- Ministry of Health (MOH), Government of Honduras: *Plan Nacional de Salud 2010–2014* (2010)

- Morgan, S.L., Winship, C.: Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press, New York (2007)
- Moynihan, D.: A great schism approaching? Towards a micro and macro public administration. *J. Behav. Public Adm.* **1**(1) (2018). <https://doi.org/10.30636/jbpa.11.15>
- Normand, S.-L.T., Wang, Y., Krumholz, H.M.: Assessing surrogacy of data sources for institutional comparisons. *Health Serv. Outcomes Res. Methodol.* **7**(1), 79–96 (2007)
- O'Neill, K.: Decentralization as an electoral strategy. *Comp. Polit. Stud.* **36**(9), 1068–1091 (2003)
- O'Neill, S., et al.: Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Serv. Outcomes Res. Methodol.* **16**(1–2), 1–21 (2016)
- Oakerson, R.J., Parks, R.B.: The study of local public economies: multi-organizational, multi-level institutional analysis and development. *Policy Stud. J.* **39**(1), 147–167 (2011)
- Ostrom, E.: Understanding Institutional Diversity. Princeton University Press, Princeton (2005)
- Ostrom, V., Tiebout, C.M., Warren, R.: The organization of government in metropolitan areas: a theoretical inquiry. *Am. Polit. Sci. Rev.* **55**(4), 831–842 (1961)
- Pearl, J.: Causal inference in the health sciences: a conceptual introduction. *Health Serv. Outcomes Res. Methodol.* **2**(3–4), 189–220 (2001)
- Posner, M.A., et al.: Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of mammography on stage of diagnosis. *Health Serv. Outcomes Res. Methodol.* **2**(3–4), 279–290 (2001)
- Ricks, J.I., Liu, A.H.: Process-tracing research designs: a practical guide. *Polit. Sci. Polit.* **51**, 842–846 (2018)
- Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1), 41–55 (1983)
- Rubin, D.B.: Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv. Outcomes Res. Methodol.* **2**(3–4), 169–188 (2001)
- Rubin, D.B.: For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2**(3), 808–840 (2008)
- Tannahill, A., Kelly, M.P.: Layers of complexity in interpreting evidence on effectiveness. *Public Health* **127**(2), 164–170 (2013)
- Sekhon, J.S.: Opiates for the matches: matching methods for causal inference. *Annu. Rev. Polit. Sci.* **12**(1), 487–508 (2009)
- Stokes, S.C., Dunning, T., Nazareno, M., Brusco, V.: *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge University Press, New York (2013)
- White, H.: An introduction to the use of randomised control trials to evaluate development interventions. *J. Dev. Effect.* **5**(1), 30–49 (2013)
- Wright, G.D., Andersson, K.P., Gibson, C.C., Evans, T.P.: Decentralization can help reduce deforestation when user groups engage with local government. *Proc. Natl. Acad. Sci.* **113**(52), 14958–14963 (2016)
- Zarychta, A.: *Making Social Services Work Better for the Poor: Evidence from a Natural Experiment with Health Sector Decentralization in Honduras*. Working Paper (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.