ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gauri, Varun; Jamison, Julian C.; Mazar, Nina; Ozier, Owen

Working Paper

Motivating Bureaucrats through Social Recognition: External Validity — A Tale of Two States

IZA Discussion Papers, No. 12251

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Gauri, Varun; Jamison, Julian C.; Mazar, Nina; Ozier, Owen (2019) : Motivating Bureaucrats through Social Recognition: External Validity — A Tale of Two States, IZA Discussion Papers, No. 12251, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/196749

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12251

Motivating Bureaucrats through Social Recognition: External Validity — A Tale of Two States

Varun Gauri
Julian C. Jamison
Nina Mazar
Owen Ozier

MARCH 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12251

Motivating Bureaucrats through Social Recognition: External Validity — A Tale of Two States

Varun Gauri
eMBeD and The World Bank

Julian C. Jamison
eMBeD and University of Exeter

Nina Mazar
Boston University

Owen Ozier
BREAD, IZA and The World Bank

MARCH 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Motivating Bureaucrats through Social Recognition: External Validity — A Tale of Two States*

Bureaucratic performance is a crucial determinant of economic growth, but little real-world evidence exists on how to improve it, especially in resource-constrained settings. We conducted a field experiment of a social recognition intervention to improve record keeping in health facilities in two Nigerian states, replicating the intervention - implemented by a single organization - on bureaucrats performing identical tasks. Social recognition improved performance in one state but had no effect in the other, highlighting both the potential benefits and also the sometimes-limited generalizability of behavioral interventions. Furthermore, differences in facility-level observables did not explain cross-state differences in impacts, suggesting that it may often be difficult to predict external validity.

JEL Classification: C93, D73, D91, I18, L38, O35

Keywords: RCT, external validity, bureaucracy, behavioral insights, nudges, healthcare

Corresponding author:
Owen Ozier
The World Bank
1818 H St NW
Washington, DC 20433
USA

E-mail: oozier@worldbank.org

* The authors gratefully acknowledge the contributions of Ghazia Aslam, Iman Hafiz, Obert Pimhidzai, Shomikho Raha, Karima Saleh, Temilade Sesan, and Egon Tripodi for research support. We also thank David Evans, Anne Fitzpatrick, Rachel Glennerster, Pamela Jakiela, Oyebola Okunogbe, and audiences at UC Berkeley/CEGA, Oxford/CSAE, and the Paris School of Economics for helpful comments on earlier drafts. This field trial of social recognition was part of a larger study undertaken by Saleh, from the HNP, World Bank, described in Gauthier, Pimhidzai and Saleh. 2018. "Resource Tracking in Primary Health Care in Nigeria: Case Study from Niger and Ekiti states," Vol I and II, World Bank, Washington, D.C. This study is registered with the AEA RCT registry (socialscienceregistry.org), number AE-ARCTR-0002763. A previous version of this paper was circulated as World Bank Policy Research Working Paper number 8473, June, 2018. The findings and views expressed in this paper do not necessarily represent those of the World Bank, its Executive Directors, or the governments of the countries they represent.

INTRODUCTION

Public spending on health and education can occupy as much as one third of developing country government budgets, but long-standing concerns surround the effectiveness of this expenditure (World Bank 2004, Reinikka and Svensson 2004). Effective bureaucracies are crucial for economic growth and poverty reduction (Evans and Rauch 1999), but what makes individual bureaucrats more productive remains an open question for research. To motivate individual bureaucrats, governments typically use a combination of meritocratic recruitment, professional standards, laws and civil service codes, wages and bonuses, and long-term career rewards. The most suitable mix of incentives likely depends, among other things, on aspects of the bureaucratic task at hand, especially task complexity and the extent to which discretion on the part of the bureaucrat is important, as well as on the information available to political principals and on the observability of bureaucratic performance (Aghion and Tirole 1997). As a result, there have been waves of management approaches and philosophies that emphasize certain incentive schemes, often at the expense of others.

For specific, observable tasks, for example, some public organizations have introduced bonus pay (see Basinga, et al., 2011; Muralidharan and Sundararaman 2011; et cetera). Monetary awards, however, can be expensive to implement, particularly for developing country governments. Further, they could even crowd out other important tasks, or cause cream skimming of easier-to-work-with clients (Prendergast 1999). Recently, behavioral economists have begun to examine the use of less expensive, targeted interventions when bureaucrats need to achieve specific tasks. Sometimes these take the form of reminders and other kinds of choice architecture or "nudges," which have been shown to be cost-effective in achieving various policy goals in the US and UK (Benartzi, et al., 2017). Social recognition can be considered as one type of nudge, whether in the form of certificates for "employee of the month," public ceremonies, or other kinds of appreciation (e.g., Ashraf, Bandiera, and Jack 2014).

In this paper we present a field experiment that we conducted to test the effectiveness of one such behavioral incentive, a performance-contingent social recognition intervention to increase performance on an important task undertaken by health facility staff in Nigeria: the tracking of income and expenditure resource streams at primary health care (PHC) facilities (clinics, posts, and health centres). Specifically, the four weeks long intervention consisted of posting PHC facilities' weekly performance through a number of "stars" on a prominently

displayed "Certificate of Excellence" within the facility. In addition, the best performing facility was promised a public award ceremony with the Permanent Secretary of Health.

The overall goal of our trial was to overcome weak accountability structures that can lead to 'leakage', and this particular instantiation matters enough that there is a larger collaboration between the government of Nigeria and the World Bank focused on it. Working with the government, we rigorously evaluated a behavioral science approach by carrying out the same intervention in a random subset of facilities within two different states in the country (Ekiti and Niger), allowing us to directly test not only the intervention but, more importantly, also the generalizability of the results. Furthermore, we exploited survey data at the facility level to examine not only which factors influence performance on expenditure tracking but also the relative success of the intervention across environments.

In a nutshell, we found that our relatively inexpensive social recognition intervention motivated bureaucrats in one of the two states yet showed no effect in the other. Note that the dependent variable – the task of tracking resources at the facility –was quite basic, so a successful intervention may be unsurprising, but that makes the lack of success in the other state harder to explain. None of our observable variables could explain the difference in treatment effects between states.

We contribute to several strands of literature. First, in addition to examining a social recognition intervention, our top-level treatment effects add to literatures on incentives within bureaucracies (including preventing corruption) and on the possible scope for behavioral interventions in developing countries. As a secondary contribution along these lines, we contribute to the literature on interventions aimed at health systems in developing countries. The recent work of Dunsch, Evans, Eze-Ajoku, and Macis (2017), for example, relates closely to ours, in trying to improve practices within health systems in Nigeria. In alignment with the findings in that paper, we find larger effects among the simpler sub-tasks within the broader set of resource tracking measures we capture.

Finally, we contribute to the growing literature on external validity in development economics and behavioral sciences. Though generalizability is not a challenge specific to randomized trials, this trial presented an opportunity to examine generalizability more directly than is often feasible. We tested whether observable characteristics at the facility level (e.g., geographic distance between facilities, staff resources, educational attainment of the local

population), were predictive not only of levels and treatment effect heterogeneity within a single state, but also of the difference in treatment effects across states, holding constant the implementing organization. Given the lack of explanatory power of our variables, we interpret our findings to imply that social recognition can be an effective and relatively low-cost tool to motivate bureaucrats to complete a specific task such as record keeping, but that its effects are dependent on social milieu and institutional structures in ways that are difficult to measure. Our contribution is thus not only an examination of behavioral incentive schemes in an important and novel policy context, but also a next step in the assessment of external validity in field experiments.

Existing literature

The idea that individuals are motivated by public and peer recognition is not new.[2] Social recognition and peer appreciation can have internal effects on motivation, by highlighting what Cassar and Meier (2018) call "feelings of relatedness" in job design, or external effects on motivation, through the attribution of recognition to potential career trajectories. The notion that individuals crave status has long been studied and has more recently been formalized (Moldovanu, Sela, and Shi 2007; Besley and Ghatak 2008; Kosfeld and Neckermann 2011) and reviewed (Ashraf and Bandiera 2018). Psychologists sometimes worry that monetary rewards can indirectly crowd out the valuable intrinsic motivation of agents, whereas recognition is comparatively less likely to do so (Ryan and Deci 2000). It is even possible that recognition may enhance such intrinsic motivation, for instance by making the positive attributes of the effort more salient. Even without such a mechanism, recognition and other non-pecuniary incentives can enhance reputations, spur competitive behavior, and/or simply be valued in their own right.

There is consistent empirical evidence to show that employees value recognition (for a metaanalysis see Stajkovic and Luthans 1997; also, see Larkin 2011). For example, Nelson (2001) reviewed studies that identified 'appreciation' and 'recognition' as being more important than traditional incentives such as 'good wages' and 'job security' or 'promotion opportunities.' This observation remains true in developing countries, where financial constraints are especially binding. A number of studies on developing countries suggest that non-financial incentives such

[2] See, for example, Hobbes (1651) *Leviathan*, Chapter 17: "men are continually in competition for honour and dignity" – cited in Hirschman (1973).

as 'recognition' are important drivers of employee behavior (Mathauer and Imhoff 2006; Stilwell 2001); for evidence that negative recognition – e.g., peer exposure of non-compliant companies – can also improve behavior, see Chetty et al. (2014) in Bangladesh.

Specifically regarding the health sector, Mathauer and Imhoff (2006) engaged in detailed semi-structured qualitative interviews with staff from healthcare facilities from public, private, and non-governmental organization (NGO) facilities in rural areas of Benin. They concluded that health workers highly valued recognition and appreciation from superiors and colleagues as well as patients. Furthermore, this valuation did not differ by the type of institution (private, public, or NGO).

Finally, and perhaps most directly relevantly, in a field experiment run in Zambia in collaboration with a public health organization, Ashraf, Bandiera, and Jack (2014) randomized 800 community agents hired to sell condoms in urban compounds into four monetary and nonmonetary performance contingent reward treatments. Agents who were assigned to the nonmonetary reward treatment—namely, stars for performance plus a public ceremony for top performers—sold twice as many condoms as agents who were offered a modest financial margin on each pack sold. Such a study could help in public service contexts where management and performance could benefit from new techniques of motivation due to, for example, limited financial resources.

In their review of the social incentives literature, Ashraf and Bandiera (2018) conclude that the positive effect of social incentives on performance is large, ranging between 7 and 16 percent, but that its sign depends on the social group implicated in the social incentives as well as the nature of the agents' social preferences that underpin them (for the latter see also recent work by Butera et al. [2019] about the role of the shape of the social recognition utility function for quantifying the social efficiency of using social recognition in comparison to financial incentives for behavior change). Specifically, in a context like ours, where the health facility's accounting performance was to be publicly displayed to a vertical social group (i.e. non-peers; all visitors and other types of health facility employees), social incentives typically increase performance. In addition, relevant to our low-capacity context, in a study of the Nigerian civil service, Rasul and Rogger (2018) find that 38% of public sector projects never start, compared to 31% that actually finish, and Khemani (2006) finds substantial levels of non-payment of salaries to primary health care providers in Nigeria, as well as state-level variation in the same. Therefore, the public

posting of a prominent number of stars to quantify and reward performance including a public ceremony, as used in Ashraf et al.'s (2014) experiment, served as the inspiration for the specific social recognition intervention in the present study.

External validity

The first time an intervention is shown to be effective (or not) in any given setting has been of substantially greater interest to academic audiences—and has been accompanied by substantially greater rewards for academic authors—than is any subsequent attempt to replicate the finding (Galiani, Gertler, and Romero 2018). For policymakers, however, information about when and where the intervention works, and for whom, is just as important as the initial finding. Indeed, this point applies more broadly than just to field experiments or program evaluation per se. A prominent recent example of different treatment effects in otherwise similar settings is that of the intergenerational effects of incarceration, examined in Sweden and in Norway (see Dobbie et al. 2018 and Bhuller et al. 2018, respectively).

In our context, as in most situations, there are different possibilities regarding the exact mechanism underlying social recognition, some of which can be informed by sufficiently rich data on covariates. We have a more complete set of accompanying administrative data than is typical, though it is not complete (especially in terms of what would be needed for many behavioral models), and we use that to empirically ascertain three things: first, whether these covariates are predictive of the levels of the outcome in the absence of the intervention; second, whether these covariates are predictive of the impact of the intervention; and third, most crucially, whether the estimation of heterogeneous treatment effects allows us to predict the treatment effect across state boundaries. We are able to perform this last step, which appears relatively uncommon in field experiments, because we were able to implement the same intervention in the same manner and by the same organization in two different settings (in our case states within a federal country).

Hence, in engaging the recent literature on external validity, this study addresses two key issues. First, in relation to the work of Bold, et al. (2013), who show that different implementers have different effects in the same country, and Das, Friedman, and Kandpal (2018), who show that implementers and populations both explain differences in effects, we are able to show that

even for a single implementer in a single country, very different treatment effects are obtained in different states. One need not turn to implementer quality as an explanation, when the implementer is the same. Second, in relation to the work of Bates and Glennerster (2017) and Andrews and Oster (2017), we unsurprisingly show that there are multiple covariates that predict the outcome (quality of record-keeping). However, despite that relative wealth of predictive covariates, we find a scarcity of covariates that are predictive of the treatment effect. Further, we find that none of those predictive covariates predict both the treatment effect and the difference in treatment effects across states. This suggests that even when covariates that could be *logically* connected to the outcome indeed prove to be *statistically predictive* of the outcome in question, these covariates are not necessarily predictive of either within-setting or across-setting treatment effect heterogeneity. Thus, they may not be sufficient to translate treatment effects from one setting to another. Some other structural variables, potentially more difficult to quantify, such as the nature of the individuals' underlying social preferences (Ashraf and Bandiera 2018; Butera et al. 2019), may drive heterogeneity in effects across settings.

EXPERIMENTAL BACKGROUND

In Nigeria as elsewhere, record-keeping is generally viewed as a mundane task, with no broader pro-community mission acting as a motivating factor to undertake the task. There is no existing incentive for better resource tracking. In fact, there may even be perverse incentives to further highlight a situation of scarcity – with the assumption that under-reporting of resources could lead to more assistance especially from development partners. In addition, there is a structural weakness to enforce accountability. Very rarely does a fear of supervision exist that makes facility officers ensure records are maintained, to protect themselves from corruption allegations.

The importance of record keeping, in the Nigerian context, stems from several factors. First and foremost is the fact that the government of Nigeria has struggled to reliably track how much it spends on primary healthcare. This is partly because classifications of expenditure in the health sector tend to be too broad to capture such granular but vital information, but also because the resources (cash and in-kind) that flow to primary healthcare in the country are heavily fragmented, with no system in place to aggregate them into central databases. This means that it

is important to devise alternative ways of tracking income and expenditure in the sector. In an effort to help Nigeria with that development, the World Bank introduced an innovative continuous Public Expenditure Tracking Survey (PETS) and Resource Tracking (RT) exercise in two states (Ekiti and Niger), which recorded and aggregated information on resource flows in real time, rather than following the more established practice of retrospectively generating such data. In both states, the intervention and the monitoring were conducted by the same local firm.[3]

The need is especially great at the primary facility level, for which no entity above the facility level – including local government and state authorities – has reliable information on income and expenditure streams. One of the central issues identified by previous studies is the weakness of accountability structures at the facility level – a gap that a World Bank-financed intervention is attempting to address. The main contribution of the PETS/RT has been to design and introduce relatively simple-to-fill-out forms (spreadsheets; other resource tracking tools) to track various income and expenditure streams. This includes, for example, creation of a standardized cash book form (for the recording of cash receipts from user charges e.g., consultation fees, laboratory tests, and drug sales and expenditures such as due to purchases of drugs, equipment, materials, or supplies, wages, and facility maintenance), accompanied by training for staff at the facility level on how to fill out these forms on a daily basis.

However, absent sufficient motivation, the introduction of resource tracking forms (and instruction in their use) is not likely to change bureaucratic behavior at the facility level. In an important sense, the larger agenda regarding transparency and anti-corruption in countries such as Nigeria depends, in part, on the motivation of bureaucrats to complete their record keeping tasks. Indeed, accountability and the reduction of 'leakage' is another explicit factor in the importance of careful tracking of monetary inflows and outflows among civil servants. Without full tracking of cash flows, it is relatively easy to divert funds illicitly. Of course, filling out a form does not in itself guarantee accuracy, and dedicated agents could still manage to enrich themselves, but being forced to explicitly report fraudulently adds an additional barrier (both psychological and formal; Mazar and Hawkins 2015).

In the face of this challenge, the literature discussed above suggests that targeted interventions, such as social recognition, can be effective motivators of behavioral change for public officials as well as for consumers and citizens. Based on these insights, the clear need in

[3] The local firm carrying out this work was Hanovia Medical Ltd., Nigeria.

the Nigerian setting, and a growing body of evidence showing that "nudges" to influence behavioral choices can be used to address problems in public service delivery (World Bank 2015; Thaler and Sunstein 2008; Ashraf 2013; see also Ashraf et al. 2016), we designed a randomized control trial (RCT) in which we tested in the field the effectiveness of a social incentive intervention to improve performance of record-keeping at medical facilities in Nigeria. In particular, we extended the PETS/RT exercise in the two states Ekiti and Niger, focusing on the quality with which facility staff filled out the cash book form. The feasibility of the possible implementation of the intervention beyond the study period was an important consideration in its choice and design.

DATA AND EXPERIMENTAL DESIGN

In 2015, the World Bank and the Government of Nigeria embarked on a collaboration to verify that resources – funds, equipment, medication, and so forth – are received at public health facilities in the same quantities that they are sent by state and central government ministries. This exercise involved regularly visiting a set of 140 facilities across two of Nigeria's 36 states, Ekiti and Niger (see Figure 1 and Appendix Figure A5) to measure stocks and recent flows of resources. Basic summary statistics about these facilities are shown in Table 1. Most of the facilities are rural, and are open five days a week, eight or more hours a day. About one third of the facilities have access to electricity, most have running water, but virtually none has a landline or mobile phone specific to the facility.

There are some differences across the two states in these characteristics, also shown in Table 1. Facilities in Ekiti are larger, open more days per week, and open more hours per day than those in Niger. A comparable fraction of facilities in each state are categorized as rural, however, and while neither state's facilities are very likely to have access to electricity, both states' facilities are likely to have access to water. The two states differ in many other ways; Ekiti is more densely populated than Niger, for example: the former has a population density comparable to that of Rwanda; the latter, lower than that of Kenya.

In light of the motivation and goals described above, we devised a behavioral intervention that could be tested in the two states: social recognition for good performance. From January to

March, 2016, we tested the intervention in an RCT. The study leveraged the ongoing resource tracking study across the two states. Because the underlying expenditure-tracking survey was happening regardless, our intervention had a very low marginal cost. In other environments it might be necessary to leverage existing visits of different types, perhaps involving a lower frequency of monitoring.

Thus, we randomly assigned a total of 140 facilities in the two states to two arms per state for the purposes of this RCT, to find out whether the program could improve record-keeping. To maximize the study's statistical power, we assigned half of the facilities to each arm in each state. To ensure that geography would not confound analysis, we stratified this randomization by state, and within state, by local government authorities (LGA); meaning that we randomized the facilities to arms within each LGA. Nigerian LGAs are the lowest of three tiers of Nigerian government, situated below the federal and state levels (Khemani 2006). We stratify on 16 LGAs in Ekiti and 23 in Niger, so this was stricter than stratifying by state.

In Table 2, we show balance tests for the observables that, *ex ante*, we hypothesized might be responsible for heterogeneity in treatment effects. We do this for the entire sample, and separately by state. Most covariates are balanced at conventional statistical levels, but the logarithm of monthly outpatients is imbalanced in Ekiti, and the presence of designated staff for financial accounting is imbalanced in Niger.

The treatment "arm" provides social recognition as an incentive based on scoring the facilities' record-keeping qualities. The record in question was the cash book form, a simple excel balance sheet that tracked the cash receipts from user charges as well as expenditures, by category, on a daily basis. The quality of those records could be assessed by enumerators who visited the facilities on a weekly basis. The precise design of the evaluation was as follows. For four weeks of the incentives study, the two study arms being compared were

A *Control.* Records were scored weekly by enumerators; however, scores were not publicly shared.

B *Social Recognition*. Records were scored weekly by enumerators; scores were then converted to a number of stars between 0 and 5. The stars were displayed in a public place on a "Certificate of Excellence" for anyone visiting the facility to see (for a picture of the certificate, see Appendix Figure A1; for examples of their display, see Appendix Figure A4). Furthermore, the social recognition treatment

was comparative – at the end of the experimental period the best-performing facility and all staff members who participated in completing the cash book form were commended and posed for photographs and an honorable handshake with the Permanent Secretary of Health in a special ceremony, and this was explained in advance.

At the fifth week of the study, an additional incentive program began in the Control group facilities, concluding this experiment. This change was not announced in advance. The randomized design permits us to attribute any differences in performance over these four weeks (post-baseline) to the Social Recognition scheme.

Data

To measure performance, we gathered a weekly dataset on the quality of the cash book record-keeping. The scoring "checklist," which assessed the quality of the records, focused on whether the cash book had been filled out at all prior to arrival; how complete its various sections were; and how consistent it was with other documentation (including other forms and paper receipts); Based on the answers to these questions, the checklist produced a weekly score for each facility.[4]

Implementation Details

Enumerators already making weekly visits to facilities for the larger PETS/RT project were trained on our social incentive intervention. Each enumerator visited on average four or five facilities each week, which could include sample facilities from the Control group, the Social Recognition group, or both.

Facilities in the Social Recognition group were expected to display their Certificate of Excellence in a prominent spot at the facility (see Appendix A4 for sample photographs), so that their weekly performance (or lack of it) would be visible to all. The assumption here was that staff would be motivated to work better at filling out the cash book form because of the desire to

[4] For details on the exact questions, see Appendix A. Note that one of the questions, question 8 links the main "cash book" form to a different form which records dates and quantities for any drug shipments that the facility receives, whether locally purchased or regionally distributed; this link allows an enumerator to check both forms (cash book and drug purchases) for consistency on at least the arrival of drugs in the facility on a given day. The distribution of scores is shown in Appendix Figure A3.

earn stronger community approbation (quality-contingent social reward), and/or recognition from their community (e.g., visitors, patients, other facility staff members).

RESULTS

We begin with results from the RCT, reporting separately in Ekiti and Niger. Conditional on having any nonzero transactions (one of the auxiliary questions on the checklist; note that most items on the checklist cannot be meaningfully completed if no relevant transactions took place over the previous week), we compare the checklist scores in the Social Recognition arm to those in the Control arm.[5] For completeness, we first display the results pooled across the two states in Table 3A, then in Table 3B our main focus: disaggregated by state. A visualization of those trends in both states is provided in Figure 2 (upper and lower panels, respectively; "Week 0" indicates baseline performance prior to the start of the experiment).

As shown in Table 3B,[6], Social Recognition—having a Certificate of Excellence displaying the facility's performance in public on a weekly basis + public ceremony for top performer at the conclusion of the trial —improved the checklist score by 12.4 percentage points in Ekiti State: a roughly 18-percent increase over baseline. The p-value for this effect is 0.005. In Niger State, however, there was no detectable effect of the Social Recognition intervention.[7]

To be sure that the results are not sensitive to our choice of the simplest possible specification, we provide three alternative empirical specifications in Table 3B. The specification in the first and fifth columns is straightforward:

$$(1) \qquad Y_{it} = \beta^{l}_{social} Social_{it} + \gamma^{l}_{LGA} + \epsilon^{l}_{it}$$

The outcome, Y_{it}, is the score on the record-keeping checklist; the main right-hand-side variable is $Social_{it}$, an indicator for whether facility i is exposed (by the randomization) to the social recognition intervention in period t. For facilities assigned to treatment, this is 1 for periods 1-4,

[5] Nonzero transactions are present in 80 percent of facility-weeks. The fraction of weeks (out of four) that a given facility has nonzero transactions varies from zero to one; the distribution is shown in Appendix Figure A6.
[6] Note that the pooled results presented in Table 3A (odd numbered columns) vary based on the specification. We interpret this as inconclusive and do not claim that there is an overall effect of social recognition across both states.
[7] We have just one main outcome, but two states in which we test the intervention's effects on it. With an unadjusted p-value of 0.005, our finding is robust to the Dunn/Bonferroni multiple test correction (in its simplest form, multiplying the p-value by the number of tests), which yields an adjusted p-value of approximately 0.01.

and 0 only in baseline period zero (which we do not use in our initial specification); for facilities assigned to comparison, this is 0 throughout. Because treatment was randomized within "Local Government Authority," or "LGA," we include fixed effects for these strata. (We also use this specification in the first column of Table 3A, including an interaction between $Social_{it}$ and an indicator for Ekiti state, $Ekiti_i$, in the second column of Table 3A.) In the second (and sixth) column of Table 3B, we also add the facility's score in the week prior to intervention (i.e. baseline performance in period 0) as a control variable. Doing so increases precision slightly, but does not substantially change the main coefficient estimate. (Analogously in Table 3A, columns 3 and 4.) In the third (and seventh) column of Table 3B, we include data from the week prior to intervention (baseline period zero) to estimate a difference-in-differences model:

$$(2) \qquad Y_{it} = \beta^2_{social} Social_{it} + \gamma^2_i + \epsilon^2_{it}$$

As is shown in Equation 2, we include facility fixed effects rather than simply LGA fixed effects. This yields a higher point estimate, but with slightly wider standard errors. (Analogously in Table 3A, columns 5 and 6.) Finally, in the fourth (and eighth) column of Table 3B (and analogously in the last columns of Table 3A), we use the difference-in-difference specification of Equation 2, but also include scores from weeks when there were no relevant transactions at the facility. While this mutes the effect somewhat, the coefficient remains statistically significant.[8] Except for the baseline week score included in columns 2 and 6, no baseline covariate is included in any other specification in Table 3B. Because the logarithm of the number of outpatients was unbalanced across treatment arms in Ekiti, we include it as a control in regressions in Appendix Table A3. The pattern of results, in terms of magnitude as well as statistical significance, remains unchanged. Table 3A, corroborating the patterns in Table 3B (through only slightly different specifications, as described above), shows an interaction between the Ekiti indicator and the social recognition treatment that is positive and significant in three of four specifications ($p<0.01$ in two cases; $p<0.05$ in one case).

The outcome throughout Tables 3A and 3B is a scoring scale that runs from a minimum percentage score of zero to a maximum of one. While the "score" on the checklist is the basis for

[8] Note that R^2 is lower in columns 1, 2, 5, and 6, and higher in columns 3, 4, 7, and 8. This is because, as shown in Equation 2, facility fixed effects absorb a large amount of variation.

the Social Recognition incentive, on its own it may not translate or compare meaningfully to other contexts; underlying behaviors may translate better. To clarify this, the impact in Ekiti can be broken down into effects on each of the ten key components of the score, as shown in Figure 3.

As can be seen in Figure 3, Social Recognition motivated staff in Ekiti to complete all sections of the forms, to do so prior to enumerator arrival, and to check the forms for their accuracy and completeness. These changes on the first seven checklist items are all near 20 percentage points, which in some cases is a very large fraction of the Control group value: only half of the Control group facilities, for example, had the form checked by the treasurer or officer-in-charge *prior* to the enumerator's weekly visit.

The Social Recognition intervention, however effective at encouraging staff to complete the cash book form, did not change the likelihood that documentation was available to substantiate the sections of the form that pertained to cash receipts. It also did not change the likelihood that the different forms relating to drug acquisition agreed with one another, though the Control group already performed relatively well on this particular component of record keeping. This may have occurred because while record keeping officers were themselves able to enhance completeness and accuracy of the records, documentation of cash receipts and drugs also relied on the participation of other health workers in the facilities.

The bulk of the analysis shown here is conditional on nonzero transaction data: that is, that there is any information to record in the cash book form, and thus for which the checklist can assess completeness and accuracy. Any changes in the rates of nonzero transaction data that are induced by the intervention could pose a potential threat to validity. In Appendix Table A1, we test for changes in the rates of nonzero transaction data on the cash book form in both states, and find no significant difference between experimental study arms.

Heterogeneity

We explore the hypothesis that the difference in treatment effect is driven by facility-level differences rather than a different managerial environment at the state level. In Table 4, we consider a range of dimensions from Table 1 along which our contexts in Ekiti and Niger saliently differ and that could plausibly interact with the Social Recognition treatment. The expected direction of their effects may depend on various psychological and behavioral

mechanisms and assumptions. For example, busier staff could be more responsive to the treatment if the recognition is salient as time scarcity makes staff focus on short term outcomes (Mullainathan and Shafir 2013), or they could be less responsive if the payoff from the intervention is construed as to be career benefits whose temporal distance leads them to be heavily discounted. Having more women on staff or as patients could decrease the effect of the Social Recognition intervention if women compete less for this recognition than men do (Anderson et al. 2013), or it could make the treatment more effective because women may be more pro-social than men (Eckel and Grossman 1998). More educated patients could either make the treatment more effective because such patients are more likely to understand what the Certificate of Excellence stands for and react to it, or could make the treatment less effective if more socioeconomically advantaged populations—perhaps both patients and providers—prove less likely to respond to the certificate because they are less pro-social (Piff, et al. 2010). Having staff exclusively dedicated to accounting at a PHC facility could increase the effect of the social recognition intervention, especially if healthcare workers care about what that staff member thinks, or having such staff members could reduce the effect of social recognition because working hard on this could be seen as suggesting that the accounting staff was previously ineffective. And the number of staff listed on the roster could increase the effectiveness of social recognition – a larger number of peers could increase the pressure on staff to perform and attain recognition – but it could also conceivably reduce the effect of social recognition because it could make it easier for responsibility to become diffuse through a bystander effect (Chekroun and Brauer 2002).

Accordingly, in Table 4 we show results from a series of tests. In each row of the table, Columns 1 and 2 report coefficients and p-values from estimation of an equation of the form:

$$(3) \qquad Y_{it} = \beta^3_{social} Social_{it} + \beta^3_{direct} X_i + \beta^3_{interaction} Social_{it} \cdot X_i + \gamma^3_{LGA} + \epsilon^3_{it}$$

In each row, we report tests of whether a single covariate, X_i, has a direct effect, predicting the levels of the main outcome (β^3_{direct}), or an interaction with the social recognition treatment, predicting heterogeneity in treatment effects ($\beta^3_{interaction}$), taking data from both states pooled. The estimation also includes local government authority (LGA) fixed effects (γ_{LGA}) because these were the geographic strata within which treatment was randomized. Columns 3 and 4 provide

analogous reporting but with the sample restricted to Ekiti only. For clarity, we rewrite the equation above, replacing the superscript "3" with a superscript "4" for estimation within Ekiti, although the specification is the same:

$$(4) \qquad Y_{it} = \beta^4_{social} Social_{it} + \beta^4_{direct} X_i + \beta^4_{interaction} Social_{it} \cdot X_i + \gamma^4_{LGA} + \epsilon^4_{it}$$

Finally, Column 5 reports the state-treatment interaction term after inclusion of the covariate, X_i,, and its interaction with the social recognition treatment. This column estimates the following equation with data from both states:

$$(5) \qquad Y_{it} = \beta^5_{social} Social_{it} + \beta^5_{direct} X_i + \beta^5_{interaction} Social_{it} \cdot X_i + \beta^5_{state\text{-}treatment} Social_{it} \cdot Ekiti_i + \gamma^5_{LGA} + \epsilon^5_{it}$$

The coefficient reported in Column 5 of the table is the estimate of $\beta^5_{state-treatment}$, which quantifies the difference between social recognition treatment effects in Ekiti and Niger. Without any inclusion of covariates, this would be the difference between coefficients reported in Columns 1 and 5 of Table 3B, or roughly 0.14.

Taking the first row of Table 4, for example, Columns 1 and 2 show that the logarithm of the number of patients seen monthly (the covariate used as X_i in this row) is not predictive of a facility's level of performance in maintaining records, nor is it predictive of the effect of the social recognition intervention (both the direct association and the interaction are estimated to be very small in magnitude and neither is statistically significant). Columns 3 and 4 confirm that this is also true when restricting attention to Ekiti. Column 5 shows an estimate of 0.15 in the first row (almost unchanged from 0.14), showing that the inclusion of this covariate did not explain the difference in treatment effects across states.

As can be seen in columns 1 and 2, though two of the covariates are associated with performance (i.e. checklist score) at the 5 percent level, and two more are significant at the 10 percent level, only one—fraction female among staff—has a statistically significant association with variation in treatment effects (columns 3 and 4), and it does not explain any of the difference in treatment effects between Ekiti and Niger (columns 5 and 6). The two covariates whose inclusion diminishes the statistical significance of the difference between the states,

number of days open per week and staff for financial accounting, have neither a statistically significant association with performance nor are predictive of treatment effect heterogeneity in Ekiti, and neither drives the point estimate of the difference in treatment effects between states to zero. Finally, we neither observe heterogeneity in treatment effect with population density (which is significantly higher in Ekiti), nor with respect to distance to the nearest other health facility in that state, either of which could have affected communication between facilities and communities thereby influencing the effectiveness of our treatment. To corroborate this, we use the LASSO (as implemented by Ahrens, et al., 2018) to select variables in a pooled regression of the outcome on possible (standardized) covariates and interactions with the social recognition treatment, collapsing the dataset so that there is only one observation per facility. Results, shown in Appendix Table A5, always select the interaction of the indicator for Ekiti state and the social recognition treatment, but never the interaction of any other covariate with the social recognition treatment.

In short, our observable attributes at the facility level are unable to explain why the treatment had significant effects in Ekiti but not in Niger. This finding may suggest the importance of institutional, managerial, and perhaps even behavioral variables that are not routinely collected in surveys.

CONCLUSION

In this paper we provide quantitative experimental evidence regarding the effect of an incentive designed to encourage bureaucrats to perform better at work. Specifically, we introduced a public social incentive intervention for accounting staff in health facilities over a 4-week period in two Nigerian states, Ekiti and Niger, and tested its effect on how well they tracked their resources. As has been shown, despite being part of the same country and despite parallel program implementation through the same organization, the two states exhibited substantially divergent results.

The social recognition intervention had a nontrivial and statistically significant positive impact on performance in Ekiti but essentially no effect in Niger. The positive outcome in Ekiti was seen across multiple sub-dimensions, especially regarding completeness of the target cash book form, with less effect on the existence of substantiating documentation. Though the

intervention we study was not itself intended to be an exact replication of any other intervention, it closely resembles the format of other social recognition interventions (in this case, combining visual "stars" with a public ceremony for recognition, similarly to Ashraf, Bandiera, and Jack 2014). Our study results may be seen as corroborating evidence that social recognition can have substantial—though heterogeneous—effects.[9]

That we find a difference in effects across states both adds a layer of complication, and connects this work to a wider literature on external validity. Our quantitative analyses strongly suggest that contextual factors at state, community and facility levels may play a key role in determining the viability and effectiveness of our interventions in particular facilities. Recent work shows that the same intervention can have different effects in the hands of different implementers – a regularity across low-, middle-, and high-income countries (Bold et al. 2013; Banerjee, Karlan, and Zinman 2015; Allcott 2015). Others have shown that the same intervention can have different effects at different times (Rosenzweig and Udry 2016). The present project attempted to test for contextual factors while keeping the implementer and timing constant.[10] Ours is also not the first study to find different treatment effects in different locations: Kremer, Miguel, and Thornton (2009), for example, found strong effects of a merit scholarship program in one Kenyan district but not in another. However, there were immediate signs of differential program behavior in their case: for example, of the 127 facilities they study, spread across two districts, five immediately quit the study following a lightning strike in one district while only one facility left the study in the other district. We saw no such obvious outward sign of differential program reception: none of the 140 facilities dropped out of our study in either state.

---

[9] A similar finding across multiple studies makes it more likely that the finding is not a false positive. This is pointed out by Maniadis, Tufano, and List (2014) in relation to recent laboratory experiments in economics. Relatedly, it is the general intent of meta-analysis and systematic review to aggregate patterns across studies, improving precision and understanding; this of course can only be done once there are multiple studies to aggregate (Hunt, 1997; Petticrew and Roberts, 2008). The present study would only qualify as a replication in the third and weakest sense described by Levitt and List (2009), however, since it addresses the same general hypothesis as others have previously, but with a different experimental design, context, and so on.

[10] Note that the intervention is carried out by the same implementer (certificates were scored and posted by the same non-government firm) in the two states, but that the health centers themselves are managed by each state government's independent health agency.

Our analysis found that the reason for the heterogeneous effects between the two states is not explained by facility-level variables of the kind usually collected in field surveys, including measures of human capital and facility-level productivity. Explaining heterogeneity and assessing external validity of social recognition interventions may require the collection of new kinds of data both at the individual staff level (e.g., social preferences, time discounting) and at the firm/ministry level (e.g., institutional or professional incentives including career dynamics, supervision structures, accountability). Consistent with Kremer and Glennerster's (2011) suggestion that the success of interventions in health systems may depend on institutional context, we speculate that the availability of institutional, psychological, and cultural variables might help pin down the mechanism through which social recognition may operate in a given context (e.g., comparison to peers, principal-agent concerns, a gift relationship between employers and employees, the concerns of patients and customers, and reputation effects).[11] Future work could help distinguish these potential pathways as a function of the population and environment.

What we can say is that the RCT took place in two states that are very different in terms of human capital and bureaucratic organization. Across several indicators, Ekiti exhibits higher capacity than Niger, in terms of both aspects (Table 1). In addition, figures from the 2013 Demographic and Health Survey (DHS), such as data on birth registration rates, suggest that there is more bureaucratic capacity in Ekiti. This suggests, tentatively, that the social recognition incentive requires higher levels of training and organization on the part of higher-level public sector health officials in order to be effective. It may be that social recognition was motivating for officials in Ekiti, but not Niger, because in Ekiti healthcare workers believed that the bureaucracy had the capacity to use social recognition as a credible input into long-term career incentives. Indeed, qualitative interviews conducted alongside our experiment suggested that Ekiti staff rotate more often than Niger staff. Social recognition might be more valuable in contexts in which staff are pursuing new positions.

Niger and Ekiti states also differ along many other dimensions; in addition to the many demographic differences enumerated in this paper, the drug procurement systems for the public health systems in the two states depend on different fractions of public funding; selling drugs on

[11] Qualitative work that accompanied our experiment suggests that the social recognition certificates did not grab the attention of many patients, and that pride or shame in front of peers may have been a more active mechanism.

credit, which makes bookkeeping more complex, also appears to be more widespread in Niger. (Gauthier, Pimhidzai, and Saleh 2018, Chapter 5). As in many developing countries, the bureaucratic and managerial environment for service providers in Nigeria is highly heterogeneous, with variation between states and even within a single state (Rogger 2017). That an intervention was successful in one state, but not another, speaks to the importance of considering this variation when translating successful findings to a new context—even when the implementers in both cases are the same.

It is noteworthy that the control group in Ekiti increased performance over the study period (see Figure 2). This suggestive upward trend may have been due to unrelated external factors, training (i.e. learning about how to fill out the forms correctly), spillovers from the Social Recognition arm, or the mere fact of being monitored. The last would be consistent with work finding that doctors improve their performance when peers simply inquire about their work (Brocke, Lange, and Leonard 2016) and suggests that even light-touch interventions, such as phone calls or texts from peers, could be effective for improving bureaucratic performance in developing countries. Consistent with the transparency and decentralization literature (Goncalves 2014, Bardhan 2016, Gauri et al 2015, Olken 2007, Bjorkman and Svensson 2009), this also suggests that observation from the public may matter. Indeed, the broader role of community engagement to improve performance is ripe for further exploration.

In Ekiti, the Social Recognition intervention significantly improved the quality with which facilities filled out the cash book form and sustained the improved performance over the intervention period. Naturally, this analysis cannot tell us anything about long-term effects of similar approaches, and in particular whether there is potentially either adaptation to the recognition or, on the flipside, habit formation, regarding the desired activity.

Finally, it would be valuable to know more about the effects of our intervention on the overall performance of the PHC staff in a multi-task environment. For example, our measured increase in record keeping performance may have come at the cost of other tasks the staff was also responsible for (Holmstrom and Milgrom 1991). While such an analysis may be empirically difficult to operationalize when one doesn't know where to look for potential spill-over effects, we explored the effect on two other performance indicators that seemed potentially likely: patient wait times and treatment times; we did not find evidence for these kinds of externalities (see Appendix Table A4).

Future work may further explore the external validity of interventions in organizational contexts and light touch ways of improving bureaucracies in developing countries.

REFERENCES

Ahrens, A., Hansen, C.B., & M.E. Schaffer. 2018. lasso2: Program for lasso, square-root lasso, elastic net, ridge, adaptive lasso and post-estimation OLS. http://ideas.repec.org/c/boc/bocode/s458458.html Revised 24 Jan 2019

Aghion, P. and J. Tirole (1997). "Formal and real authority in organizations." *Journal of Political Economy* 105(1): 1-29.

Allcott H. (2015). Site Selection Bias in Program Evaluation. *Quarterly Journal of Economics*. 130(3): 1117-1165.

Andersen S, Ertac S, Gneezy U, List JA, Maximiano S (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*. 95(4): 1438–43.

Andrews. I., and E. Oster (2017). "Weighting for External Validity." NBER Working Paper No. 23826, National Bureau of Economic Research, Cambridge, MA.

Ashraf N (2013) Rx: Human Nature: How Behavioral Economics Is Promoting Better Health Around the World. *Harvard Business Review* 91(4): 119–125.

Ashraf N, Bandiera (2018). Social Incentives in Organizations *Annual Review of Economics* 10:439-463.

Ashraf N, Bandiera O, Jack K (2014) No Margin, No Mission? A Field Experiment for Public Services Delivery. *Journal of Public Economics* 120: 1–17.

Ashraf N, Bandiera O, Lee SS (2016) Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery. Unpublished paper, London School of Economics, London.

Banerjee A, Karlan D, Zinman J (2015) Six Randomized Evaluations of Microcredit: Introduction and Further Steps. *American Economic Journal: Applied Economics* 7(1): 1-21.

Bardhan, P. (2016). "State and development: The need for a reappraisal of the current literature." *Journal of Economic Literature* 54(3): 862-892.

Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet*, 377(9775), 1421-1428.

Bates MA, Glennerster R (2017). The generalizability puzzle. *Stanford Social Innovation Review* Summer 2017: 50-54.

Benartzi, S, Beshears J, Milkman KL, Sunstein CR, Thaler RH, Shankar M, Tucker-Ray W, Congdon W, Galing S (2017). Should Governments Invest More in Nudging? *Psychological Science* 28(8): 1041-1055.

Besley, T., & Ghatak, M. (2008). Status incentives. *American Economic Review*, *98*(2), 206-11.

Brock, J. M., Lange, A., & Leonard, K. L. (2016). Generosity and prosocial behavior in healthcare provision evidence from the laboratory and field. *Journal of Human Resources*, 51(1), 133-162.

Bhuller, M, Dahl, GB, Loke, KV, and Mogstad, M. (2018). Incarceration spillovers in criminal and family networks. NBER Working Paper 24878.

Björkman, M. and J. Svensson (2009). "Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda." *The Quarterly Journal of Economics* 124(2): 735-769.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and J. Sandefur (2013) Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. Working Paper 321, Center for Global Development, Washington, DC.

Butera, L., Metcalfe, R., Morrison, W., & Taubinsky, D. (2019) The Deadweight Loss of Social Recognition, *NBER Working Paper No. 25637*.

Cassar, L. and S. Meier (2018). Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives* 32 (3): 215-238.

Chekroun, P. and M. Brauer (2002). "The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations." *European Journal of Social Psychology* 32(6): 853-867.

Chetty, R., Mobarak, M., and M. Singhal (2014) Increasing Tax Compliance through Social Recognition. Policy Brief.

Das, A., Friedman, J., and E. Kandpal (2018). Does involvement of local NGOs enhance public service delivery? Cautionary evidence from a malaria-prevention program in India. *Health Economics.* 27:172-188.

Dobbie, W., Goldon, J, and Yang, CS. (2018). The effects of pretrial detention on conviction, future crime, and employment: evidence from randomly assigned judges." American Econmic Review 108:2 (201-40).

Dunsch,FA, DK Evans, E Eze-Ajoku, and M Macis (2017). Management, Supervision, and Health Care: A Field Experiment. *IZA Discussion Paper no. 10967*

Eckel CC, Grossman PJ (1998). Are women less selfish than men? Evidence from dictator experiments. *The Economic Journal*. 108(448): 726–35.

Evans, P. and J. E. Rauch (1999). "Bureaucracy and growth: A cross-national analysis of the effects of 'Weberian' state structures on economic growth." *American Sociological Review*: 64(5): 748-765.

Galiani, S., P. Gertler, and M. Romero (2018). "How to make replication the norm." *Nature* 554(7693): 417-419.

Gauri, V., et al. (2015). "The Costa Rican Supreme Court's Compliance Monitoring System." *The Journal of Politics* 77(3): 774-786.

Gauthier, B., O. Pimhidzai, and K. Saleh (2018). "Resource Tracking in Primary Health Care in Nigeria: Case Study from Niger and Ekiti states," Vol I and II, World Bank, Washington, D.C.

Gonçalves S (2014). The Effects of Participatory Budgeting on Municipal Expenditures and Infant Mortality in Brazil. *World Development*, 53(1), 94-110.

Hirschman AO (1973). An alternative explanation of contemporary harriednes. *The Quarterly Journal of Economics*, 87(4), 634-637.

Holmstrom, B. and Milgrom, P. (1991) Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics, and Organization, 7, 24-52. http://dx.doi.org/10.1093/jleo/7.special_issue.24

Hunt, M. (1997). *How science takes stock: The story of meta-analysis.* Russell Sage Foundation.

Khemani, S. (2006). Local government accountability for health service delivery in Nigeria. *Journal of African Economies*, 15(2), 285-312.

Kosfeld, M. and S. Neckermann (2011). Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics* 3: 86-99.

Kremer, M., & Glennerster, R. (2011). Improving health in developing countries: evidence from randomized evaluations. In *Handbook of Health Economics* (Vol. 2, pp. 201-315). Elsevier.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics, 91*(3): 437-456.

Larkin I (2011) Paying $30,000 for a gold star: An empirical investigation into the value of peer recognition to software salespeople. *Mimeo*.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1-18.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, *104*(1), 277-90.

Mathauer I, Imhoff I (2006) Health worker motivation in Africa: the role of non-financial incentives and human resource management tools. *Human Resources for Health, 4*(1): 24.

Mazar, N and S. A. Hawkins (2015), Choice Architecture in Conflicts of Interest: Defaults as Physical and Psychological Barriers to (Dis)honesty, *Journal of Experimental Social Psychology, 59 (July), 113-117.*

Moldovanu B, Sela A, Shi X (2007) Contests for Status. *Journal of Political Economy, 115*(2): 338–63.

Mullainathan, S. and E. Shafir (2013). Scarcity: Why Having Too Little Means So Much, Macmillan.

Muralidharan K, Sundararaman V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39-77.

National Population Commission and ICF International (2013) *Nigeria Demographic and Health Survey*.

Nelson RB (2001) Factors that encourage or inhibit the use of non-monetary recognition by U.S. managers. Dissertation, Claremont Graduate University, available at http://novascotia.ca/psc/pdf/employeeCentre/recognition/toolkit/step1/BobsPhDDissertation.pdf. Accessed Oct 20, 2016.

Olken, B. A. (2007). "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of Political Economy* 115(2): 200-249.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide.* John Wiley & Sons.

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. *Journal of personality and social psychology*, 99(5): 771.

Prendergast, C. (1999). "The provision of incentives in firms." *Journal of Economic Literature* 37(1): 7-63.

Reinikka R, Svensson J (2004) Local capture: evidence from a central government transfer program in Uganda. *Quarterly Journal of Economics* 119(2): 679-705.

Rasul, I., and Rogger, D. (2018) Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *Economic Journal* 128(608): 413-46.

Rogger D (2017) Who Serves the Poor? Surveying Civil Servants in the Developing World. Policy Research Working Paper 8051, World Bank Group, Washington, DC.

Rosenzweig M, Udry C (2016) External validity in a stochastic world. *NBER Working Paper No. 22449,* National Bureau of Economic Research, Cambridge, MA.

Ryan RM, Deci EL (2000) Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25(2000): 54–67.

Stajkovic, AD, Luthans F (1997) A meta-analysis of the effects of organizational behavior modification on task performance, 1975–95. *Academy of Management Journal* 40(5): 1122–1149.

Stilwell B (2001) Health worker motivation in Zimbabwe. Unpublished paper/internal report for the Department of Organization of Health Care Delivery.

Thaler R, Sunstein CR (2009) *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin Books, New York.

World Bank (2004) Making Services Work for Poor People. *World Development Report,* Washington, D.C.

World Bank (2015) Mind, Society and Behavior. *World Development Report,* Washington, D.C.

Table 1: Summary Statistics by State

	Ekiti			Niger			
Dependent variable	Obs.	Mean	S.D.	Obs.	Mean	S.D.	p-value
Rural (indicator)	65	0.86	0.35	75	0.88	0.33	0.748
Total No of Beds	65	6.75	4.44	75	3.11	3.55	0.000***
Electricity availability	65	0.37	0.44	75	0.27	0.38	0.170
Water availability	65	0.82	0.36	75	0.89	0.30	0.214
Toilet availability	65	0.72	0.44	75	0.52	0.49	0.011**
Log(monthly outpatients)	64	4.53	0.94	75	4.99	1.19	0.012**
Outpatient gender ratio	63	3.98	1.73	75	3.39	1.56	0.039**
Patient education (secondary+)	61	0.70	0.21	69	0.43	0.33	0.000***
Number of days open per week	65	6.28	0.96	75	5.28	0.69	0.000***
Number of hours open per day	65	14.33	7.18	75	6.97	2.49	0.000***
Population density (per km^2)	65	471.28	216.36	75	270.80	744.14	0.028**
Staff for financial accounting	65	0.58	0.50	75	0.12	0.33	0.000***
Number of staff listed on roster	65	10.74	7.26	75	4.72	5.55	0.000***
Fraction female among staff	65	0.89	0.16	75	0.47	0.36	0.000***

In the table above, for each variable, the number of observations, mean for the variable, and standard deviation of the variable are reported separately for facilities in Ekiti and Niger state. The last column reports the p-value for the test of equality of the means in the two states. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. The table is broken into four sections. The first section reports basic characteristics of the facilities; electricity, water, and toilets are coded 1=available, 0=unavailable, and 0.5=not available "right now." The second, third, and fourth sections report variables against which we may be interested in testing for treatment effect heterogeneity. The first variable is the natural logarithm of the average number of outpatients per month seen in a seven-month period, October 2015 - April 2016. Outpatient gender ratio is reported as female/male. Patient education is calculated as the average over patient exit surveys conducted by the research team at the facility. Population density is calculated at the Local Government Authority (LGA) level from Nigeria's 2006 census.

Table 2: Tests of covariate balance across treatment arms

Dependent variable	Both states			Ekiti			Niger		
	C N=71	T N=69	p-value	C N=33	T N=32	p-value	C N=38	T N=37	p-value
Log(monthly outpatients)	4.74	4.81	0.7184	4.20	4.89	0.0027***	5.22	4.75	0.0950*
Outpatient gender ratio	3.41	3.92	0.0734*	3.65	4.32	0.1236	3.21	3.58	0.3209
Patient education (secondary+)	0.57	0.54	0.5697	0.72	0.68	0.4578	0.45	0.40	0.5405
Number of days open per week	5.70	5.79	0.5847	6.12	6.44	0.1859	5.34	5.22	0.4588
Number of hours open per day	9.98	10.88	0.4066	13.64	15.05	0.4317	6.80	7.18	0.5273
Staff for financial accounting	0.35	0.32	0.7240	0.52	0.66	0.2549	0.21	0.03	0.0140**
Number of staff listed on roster	7.82	7.26	0.6482	10.30	11.19	0.6282	5.66	3.78	0.1493
Fraction female among staff	0.71	0.62	0.1506	0.91	0.88	0.3580	0.54	0.40	0.1090

The table above reports differences between treatment and control arms for a set of covariates described in the first column. There are nine numeric columns: the first three report the means for the comparison group (C) and Social Recognition treatment group (T), followed by a p-value from a test of the equality of these means. The next three report the analogous means and test for Ekiti state; the last three report these for Niger state. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level

Table 3A: Main effects, pooled across states.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Social recognition	0.032	-0.020	0.034	-0.009	0.091**	0.014	0.081**	0.054
	(0.025)	(0.027)	(0.022)	(0.025)	(0.035)	(0.030)	(0.032)	(0.039)
Social recognition × Ekiti		0.144***		0.114***		0.176**		0.055
		(0.051)		(0.025)		(0.030)		(0.039)
Baseline score			0.318***	0.290***				
			(0.087)	(0.084)				
Constant	0.681***	0.679***	0.477***	0.493***	0.650***	0.653***	0.671***	0.671***
	(0.017)	(0.017)	(0.057)	(0.054)	(0.014)	(0.012)	(0.013)	(0.013)
R^2	0.326	0.349	0.396	0.410	0.625	0.635	0.578	0.579
F test		4.325		4.646		4.986		3.191
p-value from F test		0.0152		0.0113		0.0082		0.0441
Observations	459	459	451	451	558	558	696	696

The table above reports estimated treatment effects. All columns pool Ekiti and Niger states for analysis. Odd-numbered columns present pooled analysis only; even-numbered columns interact treatment with an indicator for Ekiti state, separating the effect in Niger from that in Ekiti. Columns 1 and 2 include all treated weeks (weeks 1 through 4) in which there is non-zero transaction data at a facility; columns 3 and 4 include the checklist score in the baseline "week zero" as a control variable (conceptually, the "ANCOVA" specification); columns 5 and 6 show the difference-in-differences specification, in which "week zero" data are included as separate observations for each facility, and in which the treatment indicator only switches on in week 1; and columns 7 and 8 show the difference-in-differences specification but also include observations in weeks during which the facility had no transactions to score on the checklist. Columns 1 through 4 include Local Government Authority (stratum) fixed effects; columns 5 through 8 do not. All standard errors are clustered at the level of the facility. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3B: Main effects.

	Ekiti				Niger			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Social recognition	0.124***	0.104***	0.190***	0.110**	-0.020	-0.011	0.014	0.054
	(0.044)	(0.035)	(0.062)	(0.052)	(0.028)	(0.023)	(0.030)	(0.039)
Baseline score		0.304***				0.261		
		(0.092)				(0.177)		
Constant	0.742***	0.542***	0.702***	0.735***	0.638***	0.472***	0.619***	0.615***
	(0.033)	(0.071)	(0.024)	(0.021)	(0.018)	(0.105)	(0.012)	(0.015)
R^2	0.221	0.289	0.574	0.545	0.257	0.324	0.620	0.519
Observations	182	182	228	323	277	269	330	373

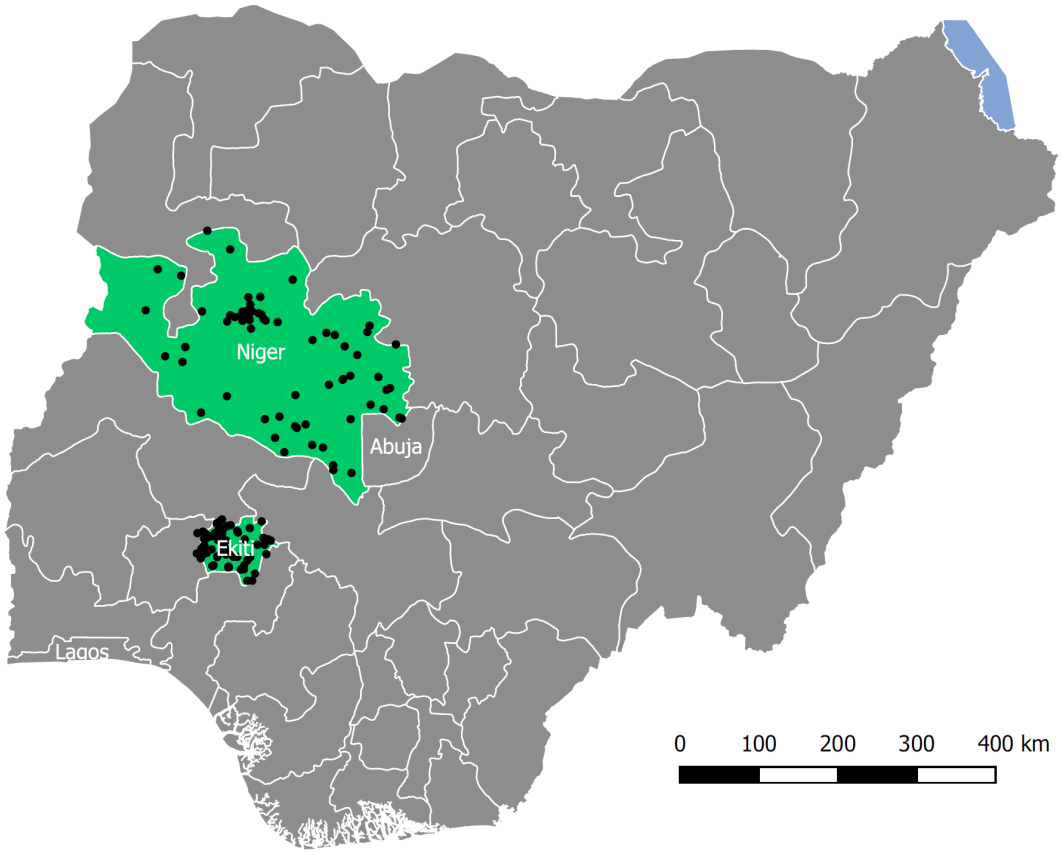
The table above reports estimated treatment effects. Columns 1 through 4 estimate treatment effects in Ekiti state; columns 5 through 8 do so in Niger state. Columns 1 and 5 include all treated weeks (weeks 1 through 4) in which there is non-zero transaction data at a facility; columns 2 and 6 include the checklist score in the baseline "week zero" as a control variable (conceptually, the "ANCOVA" specification); columns 3 and 7 shows the difference-in-differences specification, in which "week zero" data are included as separate observations for each facility, and in which the treatment indicator only switches on in week 1; and columns 4 and 8 show the difference-in-differences specification but also include observations in weeks during which the facility had no transactions to score on the checklist. Columns 1, 2, 5, and 6 include Local Government Authority (stratum) fixed effects. All standard errors are clustered at the level of the facility. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 4: Interactions

	Both states				Ekiti				Both states	
	β^1_{direct} **Direct effect of covariate** **(1)**		$\beta^1_{interaction}$ **Interacted covariate** **(2)**		β^2_{direct} **Direct effect of covariate** **(3)**		$\beta^2_{interaction}$ **Interacted covariate** **(4)**		$\beta^3_{state-treatment}$ **State-treatment interaction** **(5)**	
Patients:										
Log(patients)	0.04	$p = 0.151$	-0.02	$p = 0.520$	0.03	$p = 0.498$	-0.03	$p = 0.584$	0.15***	$p = 0.009$
Gender ratio	0.00	$p = 0.881$	0.03	$p = 0.363$	0.02	$p = 0.590$	-0.01	$p = 0.863$	0.16***	$p = 0.004$
Education	0.02	$p = 0.385$	0.04*	$p = 0.078$	0.10*	$p = 0.063$	0.01	$p = 0.828$	0.12**	$p = 0.024$
Pop. density	.	.	-0.00	$p = 0.917$	.	.	0.05	$p = 0.510$	0.15***	$p = 0.005$
Operational characteristics:										
Days open	0.05*	$p = 0.057$	0.03	$p = 0.245$	0.06	$p = 0.173$	-0.01	$p = 0.913$	0.12*	$p = 0.080$
Hours open	0.03	$p = 0.263$	0.03	$p = 0.301$	0.07**	$p = 0.043$	-0.02	$p = 0.707$	0.20**	$p = 0.013$
Baseline score	0.08***	$p = 0.004$	0.00	$p = 0.940$	0.08***	$p = 0.001$	-0.03	$p = 0.375$	0.12***	$p = 0.006$
KM to neighbor	0.06**	$p = 0.040$	-0.04*	$p = 0.050$	-0.09	$p = 0.600$	0.12	$p = 0.588$	0.15***	$p = 0.008$
Staff:										
Accounting	0.05**	$p = 0.013$	0.06**	$p = 0.024$	0.04	$p = 0.434$	0.07	$p = 0.194$	0.07	$p = 0.323$
Total staff	0.02	$p = 0.558$	0.03	$p = 0.311$	0.07*	$p = 0.093$	-0.03	$p = 0.487$	0.14**	$p = 0.044$
Fraction female	-0.00	$p = 0.919$	0.04**	$p = 0.035$	0.22**	$p = 0.036$	-0.31**	$p = 0.015$	0.15**	$p = 0.026$

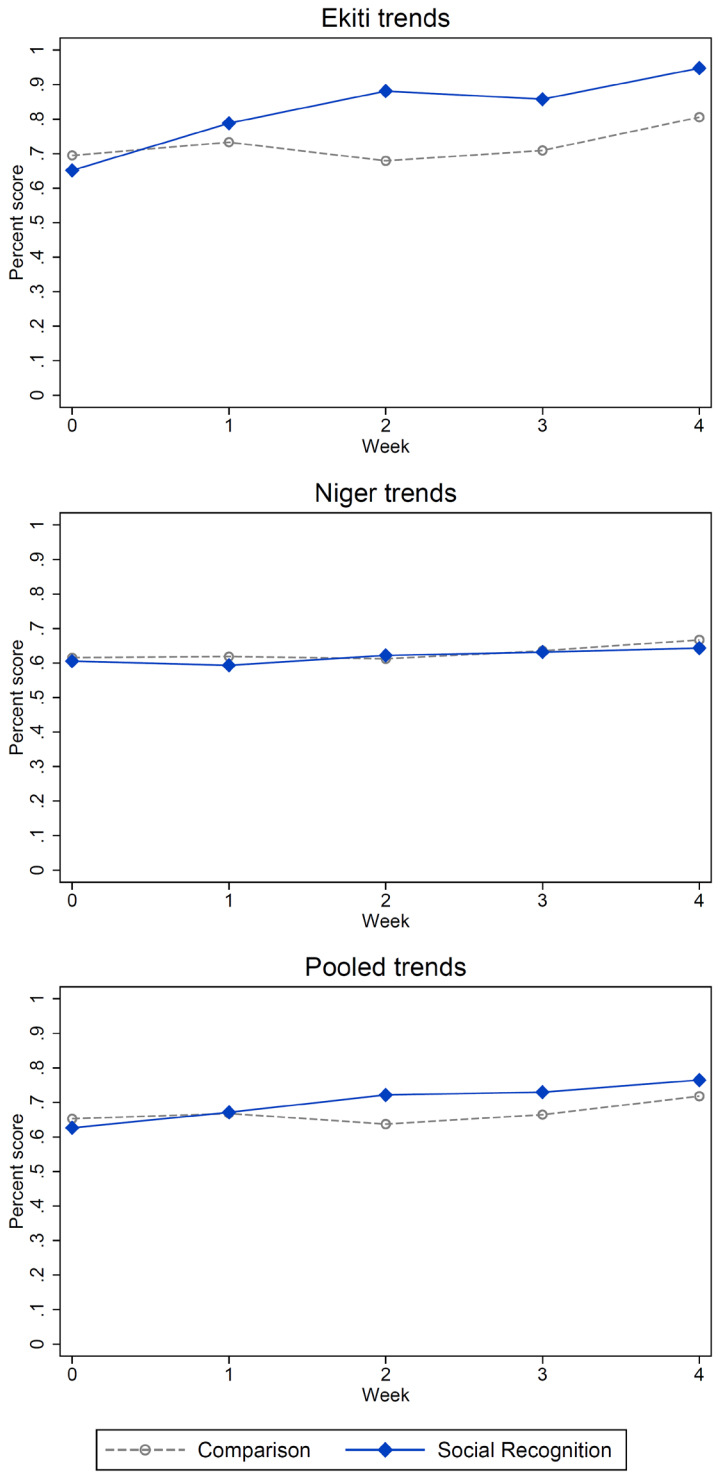
Columns (1) and (2) come from estimating Equation (3), within Ekiti state only; Columns (3) and (4) come from estimating the same equation, but within both Ekiti and Niger states pooled, shown in Equation (4). Column (5) comes from estimating a specification with one more term, shown in Equation (5), again within both Ekiti and Niger states pooled. Details on variable definitions: Log(patients) is the logarithm of the monthly number of outpatients; the "patient gender ratio" is calculated among outpatients; "Patient education" is the fraction of surveyed patients who had completed at least secondary school; "Density" refers to population density; "Staff for accounting" is an indicator for whether each facility has a specific staff member tasked with financial accounting.

Figure 1: Map of Nigeria with Study Sites



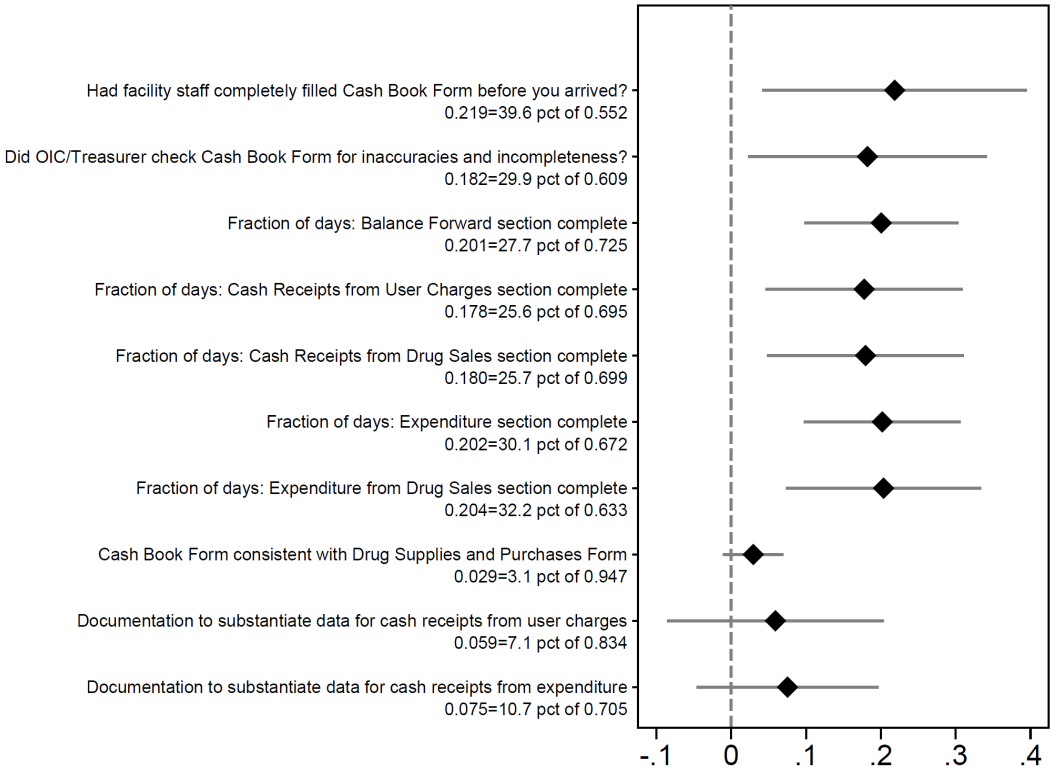
In the map of Nigeria above, the two study states, Ekiti and Niger, are shaded green; black dots in these areas mark the health facilities included in the study. The federal capital, Abuja, and the largest city, Lagos, are both labeled.

Figure 2: Record-keeping checklist score over time



In the figures above, average scores on the record-keeping checklist (where 1.0 is the maximum score) are plotted over time, conditional on having nonzero transactions to report on the checklist in each week. The upper panels show each state separately; the last panel pools both states.

Figure 3: Impacts on individual items on checklist in Ekiti



Each point above represents a coefficient estimate; each line above represents a 95-percent confidence interval. Quantities plotted above are impact estimates on specific checklist items in Ekiti state, conditional on having nonzero transactions to report on the checklist.

Appendix A: SCORING CHECKLIST QUESTIONS ASSESSING QUALITY OF RECORDS

Introductory question

1. Is the cash book form kept in the facility?

Auxiliary questions

1. Number of days in this week the facility is open
2. Are all entries on cash book form zero?

Main questions:

1. Prior to arrival:

Had the facility staff completely filled the cash book form before you arrived?

2. Treasurer/Officer-In-Charge (OIC) check:

Did Facility OIC/Treasurer check cash book form for inaccuracies and incompleteness?

3-7: Completeness on each day. For how many days

3. is the 'Balance forward' section of cash book form complete?
4. is the 'Cash receipts from user charges' section of cash book form complete?
5. is the 'Cash receipts from drug sales' section of cash book form complete?
6. is the 'Expenditure' section of cash book form complete?
7. is the 'Expenditure from drug sales' section of cash book form complete?

8. Consistency Cross-Check:

Are the drug purchases and sales records appearing on cash book form consistent with the drug purchases and sales records appearing on the drug supplies and purchases form?

9-10. Documentation. Can the staff provide receipts/invoices or other documentation

9. to substantiate the data for cash receipts from user charges?
10. to substantiate the data for cash receipts from expenditure?

Table A1: Appendix table: Effects on likelihood of nonzero transactions, by state

	Ekiti [1]	Niger [2]
Social recognition	-0.034	-0.014
	(0.085)	(0.039)
Constant	0.716***	0.927***
	(0.060)	(0.032)
R^2	0.146	0.130
Observations	253	290

The table above reports estimated treatment effects on the likelihood of having any nonzero transactions. Column 1 estimates effects in Ekiti state; column 2 does so in Niger state. Both columns include Local Government Authority (stratum) fixed effects. All standard errors are clustered at the level of the facility. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A2: Appendix table: differences between states, selected DHS 2013 statistics

	Ekiti [1]	Niger [2]
Percent of children whose births are registered	50.5	14.1
Percent of children with a birth certificate	29.2	5.6
Percent of women unable to read	7.5	68.5
Percent of women with no formal schooling	2.0	65.8
Percent of women with post-primary schooling	85.2	24.7
Median years of schooling among women	11.2	0.0
Percent of men unable to read	4.0	34.3
Percent of men with no formal schooling	1.0	31.1
Percent of men with post-primary schooling	90.5	57.6
Median years of schooling among men	11.5	8.8

All figures above come from the 2013 Demographic and Health Survey Final Report for Nigeria (National Population Commission and ICF International 2013).

Table A3: Appendix table: Main effects with log monthly patients as control

	Ekiti		Niger	
	[1]	[2]	[3]	[4]
Social recognition	0.123**	0.098**	-0.021	-0.010
	(0.049)	(0.043)	(0.028)	(0.025)
Baseline score		0.286***		0.255
		(0.091)		(0.185)
Constant	0.682***	0.462***	0.577***	0.457***
	(0.136)	(0.149)	(0.081)	(0.110)
R^2	0.239	0.301	0.279	0.325
Observations	180	180	273	269

The table above reports estimated treatment effects. Columns 1 and 2 estimate treatment effects in Ekiti state; columns 3 and 4 do so in Niger state. These are identical to columns 1, 2, 5, and 6 in the main effects table in the paper, except for the inclusion of the log of monthly patients (an imbalanced covariate) as a control variable. Columns 1 and 3 include all treated weeks (weeks 1 through 4) in which there is non-zero transaction data at a facility; columns 2 and 4 include the checklist score in the baseline "week zero" as a control variable (conceptually, the "ANCOVA" specification). All columns include Local Government Authority (stratum) fixed effects. All standard errors are clustered at the level of the facility. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A4: Appendix table showing treatment effects on possible spillover outcomes

	Wait time	Treatment duration
	[1]	[2]
Social recognition	0.584	-0.532
	(0.767)	(1.256)
Social recognition × Ekiti	-0.168	-0.502
	(2.322)	(2.050)
R^2	0.217	0.199

Wait times and treatment durations, taken from exit surveys of patients, are measured in minutes. All standard errors are clustered at the level of the facility. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

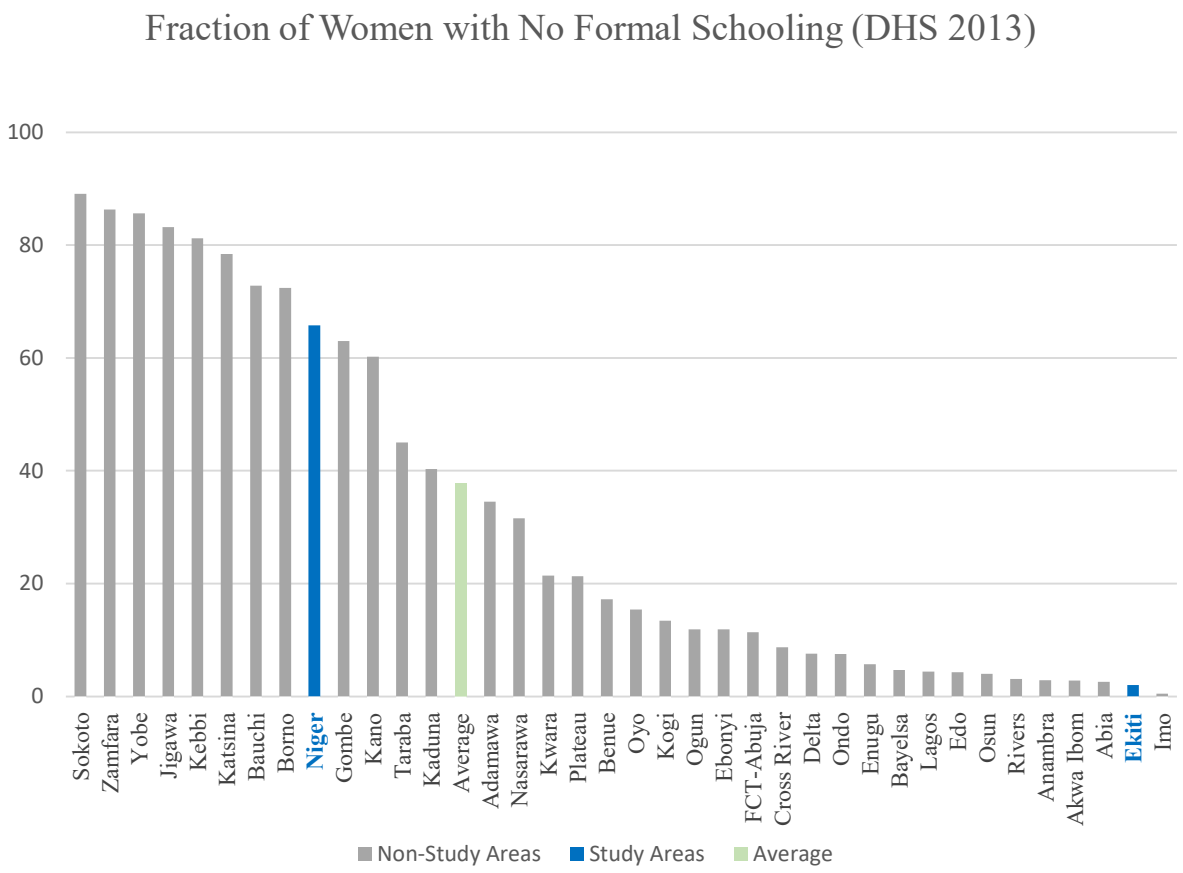
Table A5: LASSO-selected specifications

	[1] b/se/p	[2] b/se/p
Ekiti indicator	0.053 (0.044) 0.225	0.087** (0.037) 0.02
Ekiti indicator x Social recognition	0.095** (0.047) 0.045	0.125*** (0.045) 0.006
Social recognition		
Days open per week	0.025 (0.019) 0.184	
Days open per week x Social recognition		
Accounting staff	0.029 (0.018) 0.115	
Accounting staff x Social recognition		
Baseline score		0.046*** (0.015) 0.003
Baseline score x Social recognition		
Monthly outpatients		
Monthly outpatients x Social recognition		
Outpatients gender ratio		
Outpatients gender ratio x Social recognition		
Outpatient education		
Outpatient education x Social recognition		
Hours open		
Hours open x Social recognition		
Density		
Density x Social recognition		
KM to nearest		
KM to nearest x Social recognition		
Staff count		
Staff count x Social recognition		
Female staff fraction		
Female staff fraction x Social recognition		
Electricity		
Electricity x Social recognition		
Water		
Water x Social recognition		
Toilets		
Toilets x Social recognition		
Mobile phone		
Mobile phone x Social recognition		
Radio		
Radio x Social recognition		
Constant	0.647*** (0.023) p<0.001	0.632*** (0.020) p<0.001
N	129	127

Column 1, above, shows results from OLS regressions using LASSO-selected variables when all variables (even those sometimes missing) are included for all observations. Column 2 shows analogous results where we exclude outpatient education and its interaction (which are missing for ten observations), and we restrict analysis to a set of 127 observations for which no other variable is missing.

Figure A1: Appendix Figure of Certificate

Figure A2: Appendix Figure on DHS Differences



The bars above show the relative positions of Ekiti and Niger in the distribution of states in Nigeria according to the educational attainment of women.

Figure A3: Appendix Figure on Score Distribution

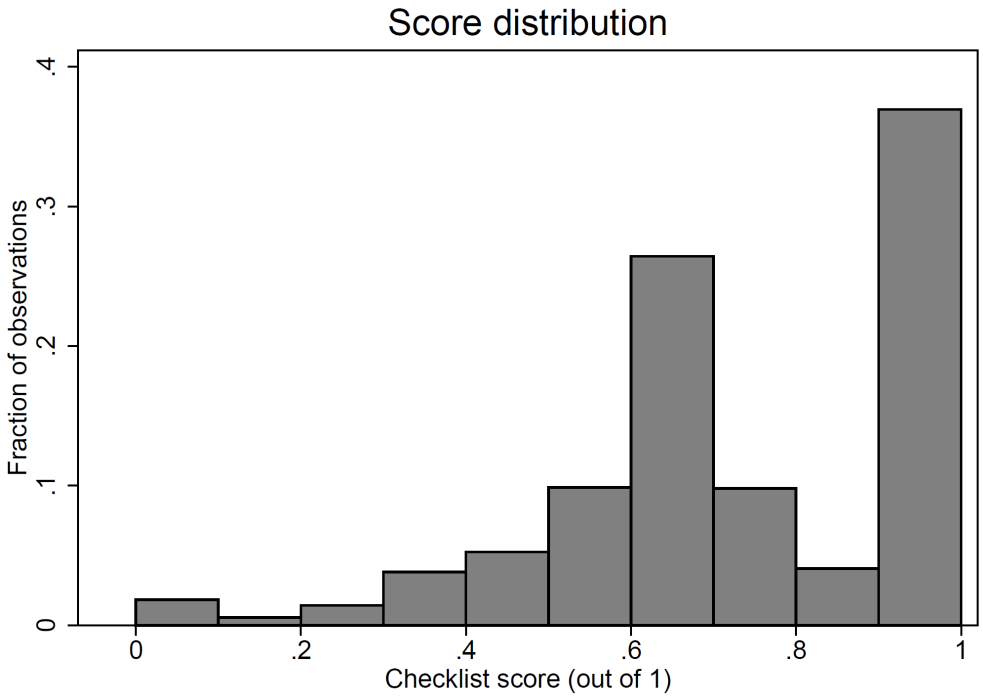


Figure A4: Appendix Figure on Display of Certificates

Figure A5: Geography of Sites by Study Arm

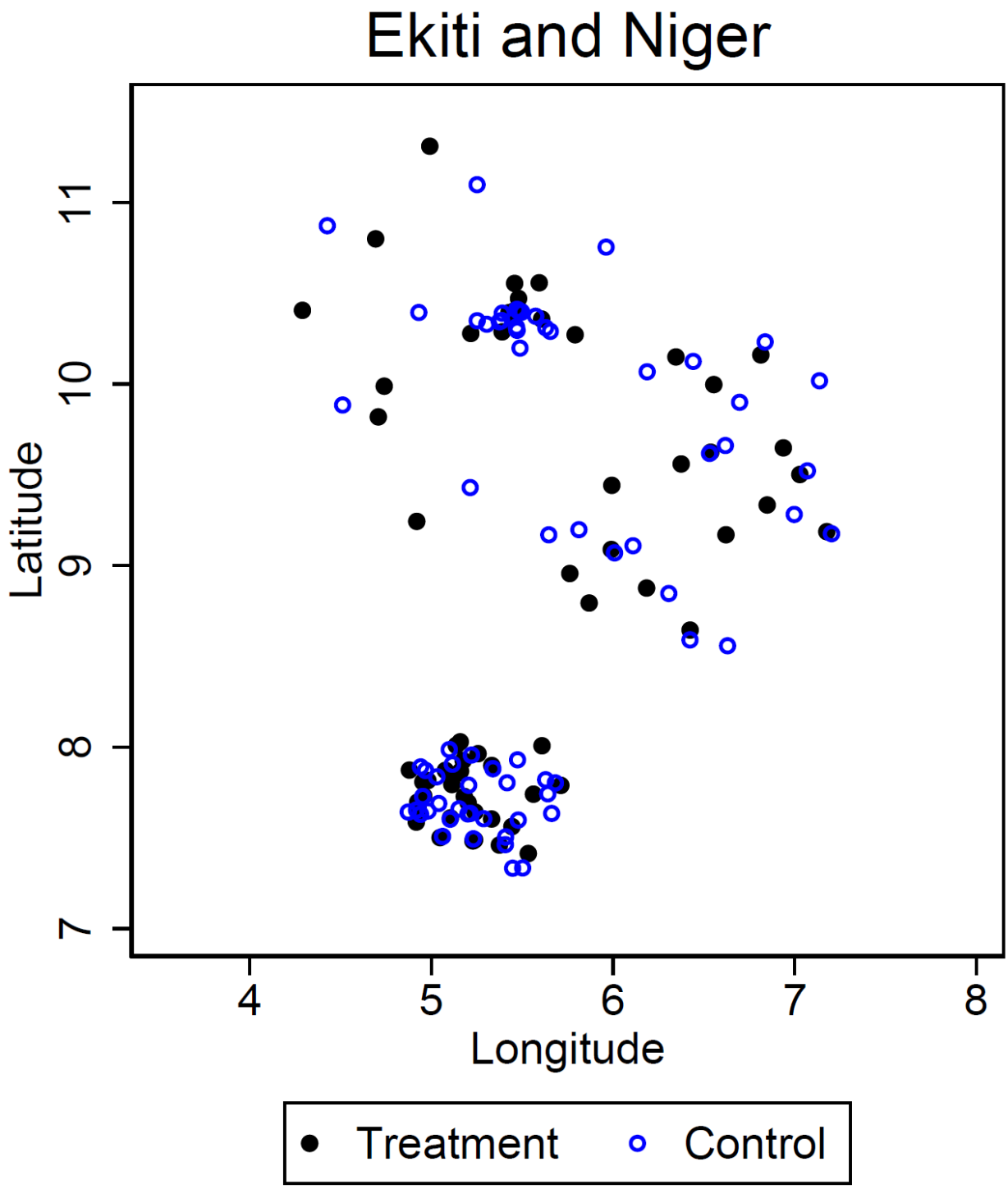
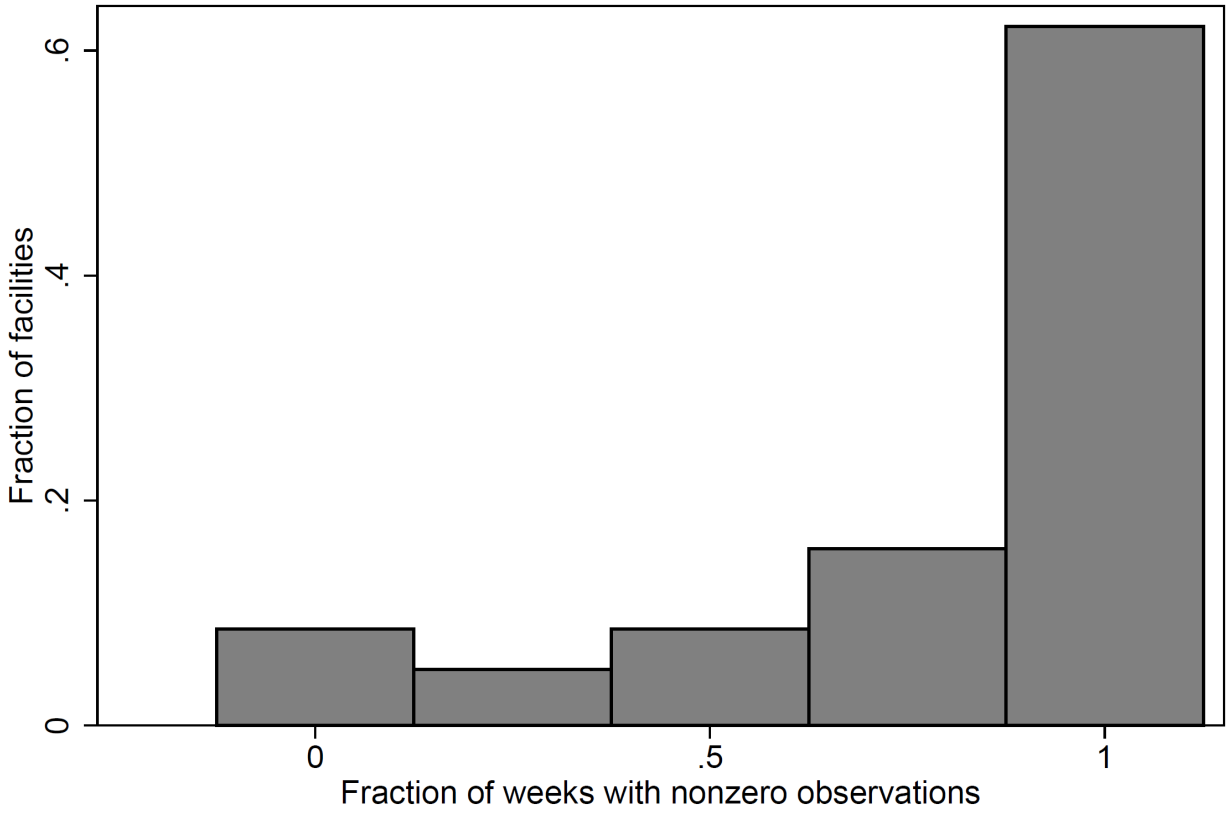



Figure A6: Nonzero transaction distribution