

Motivating Bureaucrats through Social Recognition

Evidence from Simultaneous Field Experiments

Varun Gauri
Julian C. Jamison
Nina Mazar
Owen Ozier
Shomikho Raha
Karima Saleh



WORLD BANK GROUP

Development Economics
Development Research Group
June 2018

Abstract

Bureaucratic performance is a crucial determinant of economic growth. Little is known about how to improve it in resource-constrained settings. This study describes a field trial of a social recognition intervention to improve record keeping in clinics in two Nigerian states, replicating the intervention—implemented by a single organization—on bureaucrats performing identical tasks in both

states. Social recognition improved performance in one state but had no effect in the other, highlighting both the potential and the limitations of behavioral interventions. Differences in observables did not explain cross-state differences in impacts, however, illustrating the limitations of observable-based approaches to external validity.

This paper is a product of the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at oozier@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Motivating Bureaucrats through Social Recognition: Evidence from Simultaneous Field Experiments

Varun Gauri, Julian C. Jamison, Nina Mazar, Owen Ozier,
Shomikho Raha, and Karima Saleh¹

Keywords: RCT, external validity, bureaucracy, behavioral insights, nudges, healthcare
JEL Codes: C93, D73, D91, I18, L38, O35

¹ Gauri, Jamison: Development Economics Vice Presidency, The World Bank, vgauri@worldbank.org and jjamison@worldbank.org. Mazar: Boston University, nmazar@bu.edu. Ozier: Human Development Team, Development Research Group, The World Bank, oozier@worldbank.org. Raha: Governance Global Practice, The World Bank, sraha@worldbank.org. Saleh: Health, Nutrition, and Population Global Practice, The World Bank, ksaleh@worldbank.org. The authors gratefully acknowledge the contributions of Obert Pimhidzai, for leading early data collection; and Ghazia Aslam, Iman Hafiz, Temilade Sesan, and Egon Tripodi for subsequent data and analytic work. We also thank David Evans, Anne Fitzpatrick, Rachel Glennerster, Pamela Jakiela, Oyebola Okunogbe, and audiences at UC Berkeley/CEGA and Oxford/CSAE 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 AEARCTR-0002763. 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 on, among other things, 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, and could even have the effect of crowding out other important tasks, cream skimming easier clients to work with, or working in fits and starts (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 test the effectiveness of one such behavioral incentive, a performance-contingent public social recognition, to increase performance on an important task undertaken by health clinic managers in Nigeria: the tracking of income and expenditure streams at primary health care (PHC) facilities to overcome weak accountability structures that can lead to

‘leakage’. Working with the government, we carried out the same intervention in a random subset of clinics within two different states in the country (Ekiti and Niger), allowing us to directly test not only the intervention, but also the generalizability of the results. Furthermore, we exploited survey data at the clinic level to examine which factors influence performance on expenditure tracking as well as the relative success of the intervention across environments.

We contribute to several strands of literature. First, our simple treatment effects add to literatures on incentives within bureaucracies, on preventing corruption, and on the possible scope for behavioral interventions in developing countries. We test whether our inexpensive social recognition intervention motivated individual bureaucrats to improve performance on one very specific and very simple aspect of their jobs: financial resource tracking. We find that the intervention did motivate bureaucrats in Ekiti, the originally better-performing state, but not in Niger.

Second, we contribute to the literature on interventions aimed at health outcomes and 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.

We also contribute to the growing literature on external validity in development economics. Though generalizability is not a challenge specific to randomized trials, this trial does present an opportunity to examine generalizability more directly than is often feasible. We test whether observable characteristics at the individual facility level, such as staff resources and local educational attainment, are 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. Among observable characteristics of facilities, resource tracking was marginally better (even in the absence of the intervention) in places with more educated client (patient) populations, for example. Only one variable (fraction of staff who are female) is predictive of treatment impact in Ekiti; it does not, however, explain the difference in treatment effects between states—if anything, it amplifies it. We interpret this 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 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

Management is at least as important for civil service as for corporations. For instance, rules that insulate bureaucracy from political interests, increasing its autonomy, should enhance effectiveness. In a study of the Nigerian civil service, Rasul and Rogger (2017) support this conjecture, finding that, amid widespread problems in a country where 38% of public sector projects never start, compared to 31% that actually finish, autonomy for bureaucrats is significantly associated with the likelihood of project completion. They also find, perhaps surprisingly, that management practices that control bureaucrats, in an effort to reduce corruption, may make things worse. In particular, a one standard deviation increase in incentives for monitoring corruption is associated with a 14% decline in project completion. This may follow from the fact that high-powered incentives on corruption could have the effect of increasing risk-aversion on the part of individual bureaucrats, and reduce their incentives to act on unmonitored activities, such as inputs to project completion. Bureaucrats are always multi-tasking, and when they know that one aspect of their performance is being watched closely, they may exert more effort on that dimension, at the expense of others.

If observability of bureaucratic action is a fundamental problem, perhaps involving local political leaders and citizens, through decentralization, can solve it. Decentralization might facilitate effectiveness by promoting more transparency and accountability for individual bureaucratic performance. There is evidence for this idea. For instance, data from Brazil show that the adoption of participatory budgeting increases health and sanitation spending, which in turn lowered mortality rates in those municipalities (Gonçalves 2014). On the other hand, decentralization can be associated with a “race to the bottom” as well as with the formation of regional inequalities (Bardhan 2016). Differences in local-level accountability structures can also impair service delivery – Khemani (2005) argues that differences in local-level accountability explained substantially greater non-payment of salaries to primary health care providers in Kogi state than Lagos state, also in Nigeria.

From the point of view of individual motivation, there is a more fundamental problem with decentralization as a solution. A number of tasks that bureaucrats perform, perhaps most,

are important but nevertheless not politically salient. The involvement of local political actors (leaders and citizens) thus may not motivate individual bureaucrats to perform, across a range of tasks, if those tasks are not part of those political actors' agenda. The political agenda is something that local leaders and citizens can shape, but in general the factors that determine compliance with tasks that are not usually taken to be politically salient differ in important ways from the factors affecting compliance with tasks that are salient (Gauri, Staton et al. 2015). In the absence of specific information and top-down support, citizens may not monitor corruption and service delivery on their own (Olken 2007, Björkman and Svensson 2009).

The idea that individuals are motivated by public and peer recognition is not new.² 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). 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 meta-analysis 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 with financial constraints as well. A number of studies on developing countries suggest that non-financial incentives such as 'recognition' are important drivers of employee behavior (Mathauer and Imhoff 2005; Stilwell 2001); for evidence that negative recognition – e.g., peer exposure of non-compliant companies – can also improve behavior, see Chetty et al. (2014) for evidence from Bangladesh.

Specifically regarding the health sector, Mathauer and Imhoff (2006) engaged in detailed semi-structured qualitative interviews with staff from health care facilities from public, private, and NGO facilities in rural areas of Benin. They concluded that health workers highly valued

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

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. The public posting of a prominent number of “stars” to quantify and reward performance, as used in that 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 policy makers, 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 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 shows 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 the available covariates explains 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-context or across-context 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, may drive heterogeneity in effects across settings.

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 health care. 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 health care 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.

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

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 this 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 recognition 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 clinic 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 to measure stocks and recent flows of resources. Basic summary statistics about these facilities are shown in Table 1: most facilities are rural, and are open five days a week, eight or more hours a day. About one-third of facilities have access to electricity, most have running water, but virtually none have a landline or mobile phone specific to the facility.

There are some differences across 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 a Randomized Controlled Trial (RCT). The study leveraged the ongoing resource tracking study of 140 facilities across the two states. Accordingly, we randomly assigned the 140 facilities 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 local government authorities (LGA), meaning that we randomized the facilities to arms within LGA.

In Table 2, we show tests of balance 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 shared.
- B “Social Recognition”: Records were scored weekly by enumerators; scores were then converted to a number of stars between 0 and 5. The number of stars was displayed in a public place on a Certificate of Excellence (for a picture of the certificate, see Appendix Figure A1) for anyone visiting the facility to see. Furthermore, at the end of the experimental period the best-performing facility and its accounting staff (i.e. all staff members that 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.

At the fifth week of the study, an additional incentive program began in the Control group facilities, concluding the experiment. 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 of the accounting staff. 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 the incentive intervention. Each enumerator visited on average 4-5 facilities each week, which could include both sample facilities from the Control group as well as the Social Recognition group.

Facilities in the Social Recognition group were expected to display their Certificate of Excellence in a prominent spot at the facility, so that their weekly performance (or the 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 earn stronger community approbation

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.

(quality-contingent social reward). A key element of the incentive design was the expectation that community members would engage with the certificates to a degree sufficient to provide feedback (mainly in the form of recognition) to facility staff, although the intervention did not directly communicate with community 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 during Weeks 1 – 4. The results are displayed in Table 3. A visualization of those trends in both states is provided in Figure 1 (upper and lower panels, respectively; Week 0 = Baseline performance pre-RCT).

As shown in the Table 3, Social Recognition—having a Certificate of Excellence displaying the facility’s performance in public on a weekly basis—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.⁵

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 3. In the second (and fifth) column, we add the facility’s score in the week prior to intervention as a control variable. This increases precision slightly, but does not substantially change the main coefficient estimate. In the third (and sixth) column, we include the week prior to intervention to estimate a difference-in-differences model; this yields a higher point estimate, but with slightly wider standard errors. Finally, in the fourth (and eighth) column, we include scores from weeks when there were no relevant transactions at the facility, and while this mutes the effect somewhat, the coefficient remains statistically significant. Because the logarithm of the number of outpatients

⁵ 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.

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.

The outcome in Table 3 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 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 2.

As can be seen in Figure 2, 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 relied on the participation of all 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 Ekiti and Niger states saliently differ and that could plausibly interact with the Social Recognition treatment.

The expected direction of these facility-level 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 Ertac Gneezy List Maximiano 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, Kraus et al. 2010). Staff for financial accounting could increase the effect of the social recognition intervention, especially if health care workers care about what that staff member thinks, or having such a staff member could reduce the effect of social recognition because working hard on this could be seen to be suggesting that the financial accounting staff member 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 through 4 report coefficients and p-values from estimation of an equation following the form below:

$$(1) \quad Y_{it} = \beta_{social}Social_{it} + \beta_{direct}X_i + \beta_{interaction}Social_{it} \cdot X_i + \gamma_{LGA} + \epsilon_{it}$$

In each row, we report tests of whether a single covariate, X_i , has a direct effect (β_{direct}), an interaction effect with the social recognition treatment ($\beta_{interaction}$), or both, the main outcome and the checklist score in Ekiti state (where there is a treatment effect to decompose). Estimation also includes local government authority (LGA) fixed effects (γ_{LGA}) because these were the geographic strata within which treatment was randomized. The first row of Table 4, for example, shows 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 negative, but are very small in magnitude and neither is statistically significant).

In the last two columns of Table 4, each row shows a single estimated coefficient and its associated p-value from a specification of the form:

$$(2) \quad Y_{it} = \beta_{social}Social_{it} + \beta_{direct}X_i + \beta_{interaction}Social_{it} \cdot X_i + \gamma_{LGA} + \beta_{state-treatment}Social_{it} \cdot Ekiti_i + \epsilon_{it}$$

The coefficient reported in the table is the estimate of $\beta_{state-treatment}$, 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 3, or roughly 0.14. Estimation of specification (2) above yields an almost unchanged estimate of 0.15 in the first row, 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 does diminish the statistical significance of the difference in 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. In short, our observable attributes at the facility level are unable to explain why the treatment had significant effects in Ekiti but not Niger. This 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 recognition intervention for accounting staff in health facilities over a four-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 substantial 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, albeit with less effect on the existence of substantiating documentation. This is encouraging given the relatively low cost of such an intervention.

The quantitative analyses strongly suggest that contextual factors at the 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). This project attempted to test for contextual factors while keeping the implementer constant. 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 facility 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., pro-sociality, time discounting) and at the firm/ministry level (e.g., institutional or professional incentives including career dynamics, supervision structures, accountability). 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). Our findings are consistent with Kremer and Glennerster's (2011) suggestion that the success of interventions in health systems may depend on institutional context. 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 a number of indicators, Ekiti exhibits higher capacity than Niger, in terms of both bureaucracy and human capital (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. Similarly, among the population, adult literacy rates are much higher (see Table A2 and Figure A2 in the Appendix). This suggests, tentatively, that the social recognition incentive requires higher levels of training and organization on the part of 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 health care workers believed that the bureaucracy had the capacity to use social recognition as a credible input into long-term career incentives. Niger and Ekiti states differ along many 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 (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 1). This suggestive upward trend may have been due to 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 2015, 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 (Gonçalves 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.

Finally, 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. It would also be valuable to know more about the effects of the intervention on the overall performance of bureaucrats in a multi-tasking environment. Future work may be able to speak to these and other questions concerning the establishment of effective organizations in developing countries.

REFERENCES

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.

Arrow KJ (1972) Gifts and Exchanges. *Philosophy and Public Affairs* 1(2): 343–362.

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 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.

Andrews I, Oster E (2017) Weighting for external validity. *mimeo*.

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 J, Ghatak M (2008) Status Incentives. *American Economic Review* 98(2): 206–211.

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, Sandefur J (2013) Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. Working Paper 321, Center for Global Development, Washington, DC.

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, Singhal M (2014) Increasing Tax Compliance through Social Recognition. Policy Brief.

Das A, J Friedman, 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.

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.

Gneezy U, Meier S, Rey-Biel P (2011) When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 25(4): 191–210.

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.

- Kahneman D, Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2): 263–291.
- 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.
- Krizan Z, Windschitl PD (2007) The influence of outcome desirability on optimism. *Psychological Bulletin* 133(1): 95–121.
- Lacetera N, Macis M, Slonim R (2013) Economic rewards to motivate blood donations. *Science* 340(6135): 927–928.
- Larkin I (2011) Paying \$30,000 for a gold star: An empirical investigation into the value of peer recognition to software salespeople. *Mimeo*.
- 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.
- 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*.

Naritomi J (2015) Consumers as Tax Auditors. Working Paper, Harvard University, Cambridge, MA

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.

Peterson SJ, Luthans F (2006) The impact of financial and nonfinancial incentives on business-unit outcomes over time. *Journal of applied Psychology* 91(1): 156.

Piff, P. K., et al. (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, Rogger D (2017) Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *Economic Journal (forthcoming)*

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*.

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.

Stajkovic AD, Luthans F (2001) Differential effects of incentive motivators on work performance. *Academy of Management Journal* 44(3): 580–590.

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

<i>Dependent variable</i>	EKITI			NIGER			P-VALUE
	OBS.	MEAN	S.D.	OBS.	MEAN	S.D.	
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

<i>Dependent variable</i>	BOTH STATES			EKITI			NIGER		
	C	T	P-VALUE	C	T	P-VALUE	C	T	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 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 3: Main effects table

	EKITI				NIGER			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Social recognition	0.124*** (0.044)	0.104*** (0.035)	0.190*** (0.062)	0.110** (0.052)	-0.020 (0.028)	-0.011 (0.023)	0.014 (0.030)	0.054 (0.039)
Baseline score		0.304*** (0.092)				0.261 (0.177)		
Constant	0.742*** (0.033)	0.542*** (0.071)	0.702*** (0.024)	0.735*** (0.021)	0.638*** (0.018)	0.472*** (0.105)	0.619*** (0.012)	0.615*** (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

	Direct effect of covariate		Interaction effect of covariate		State-treatment interaction	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
	(1)	(2)	(3)	(4)	(5)	(6)
Log(monthly outpatients)	0.03	0.4979	-0.03	0.5835	0.15***	0.0093
Outpatient gender ratio	0.02	0.5900	-0.01	0.8627	0.16***	0.0035
Patient educ. (secondary+)	0.10*	0.0633	0.01	0.8281	0.12**	0.0243
Number of days open per week	0.06	0.1730	-0.01	0.9130	0.12*	0.0802
Number of hours open per day	0.07**	0.0435	-0.02	0.7072	0.20**	0.0126
Population density (000 / km^2)	.	.	0.05	0.5103	0.16***	0.0030
Staff for financial accounting	0.04	0.4339	0.07	0.1935	0.07	0.3225
Number of staff listed on roster	0.07*	0.0930	-0.03	0.4866	0.14**	0.0440
Fraction female among staff	0.22**	0.0364	-0.31**	0.0146	0.15**	0.0264

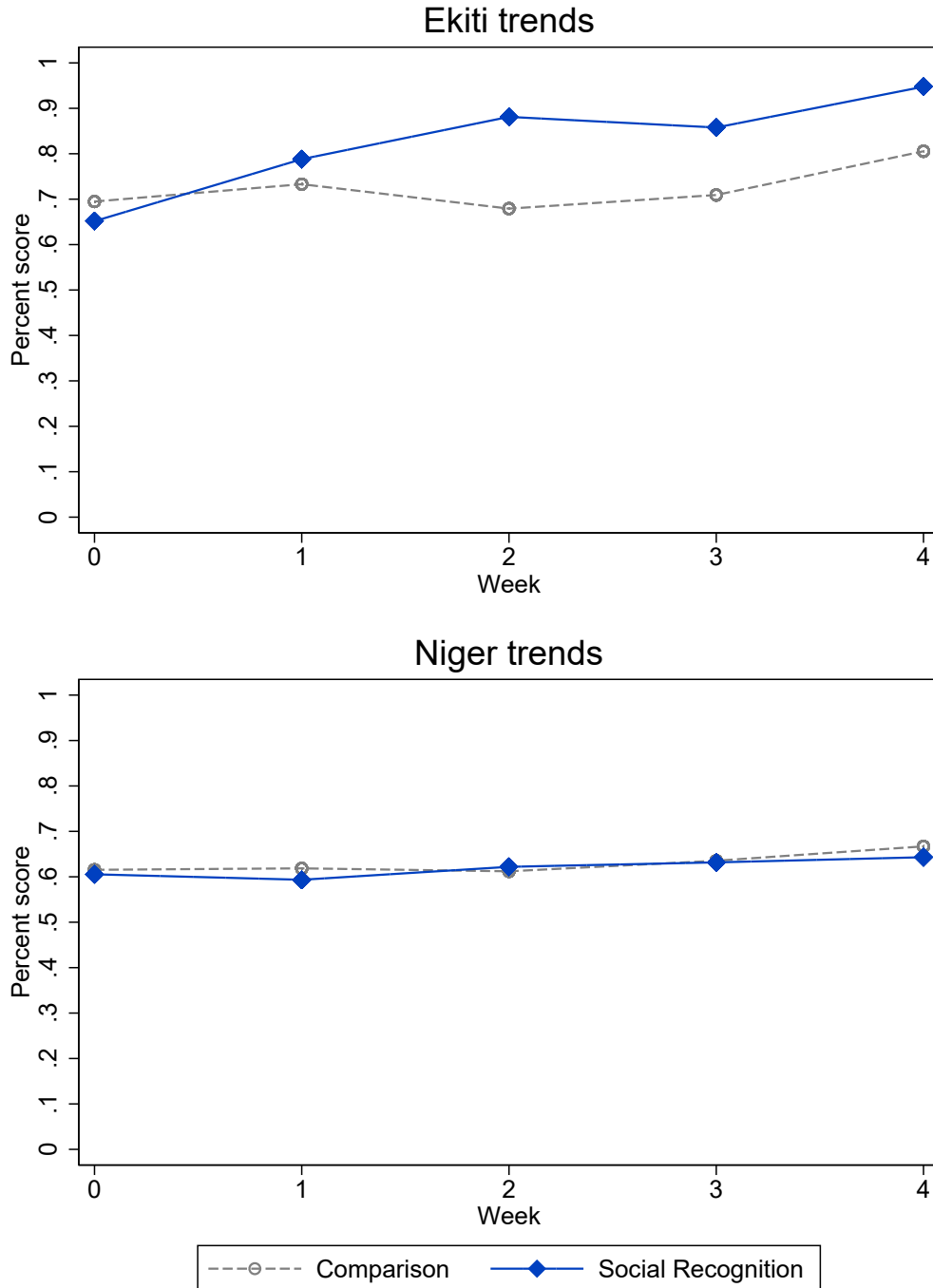
Columns (1) and (3) come from estimating (within Ekiti state only):

$$Y_{it} = \beta_{social}Social_{it} + \beta_{direct}X_i + \beta_{interaction}Social_{it} \cdot X_i + \gamma_{LGA} + \epsilon_{it} .$$

Column (5) comes from estimating (within both Ekiti and Niger states):

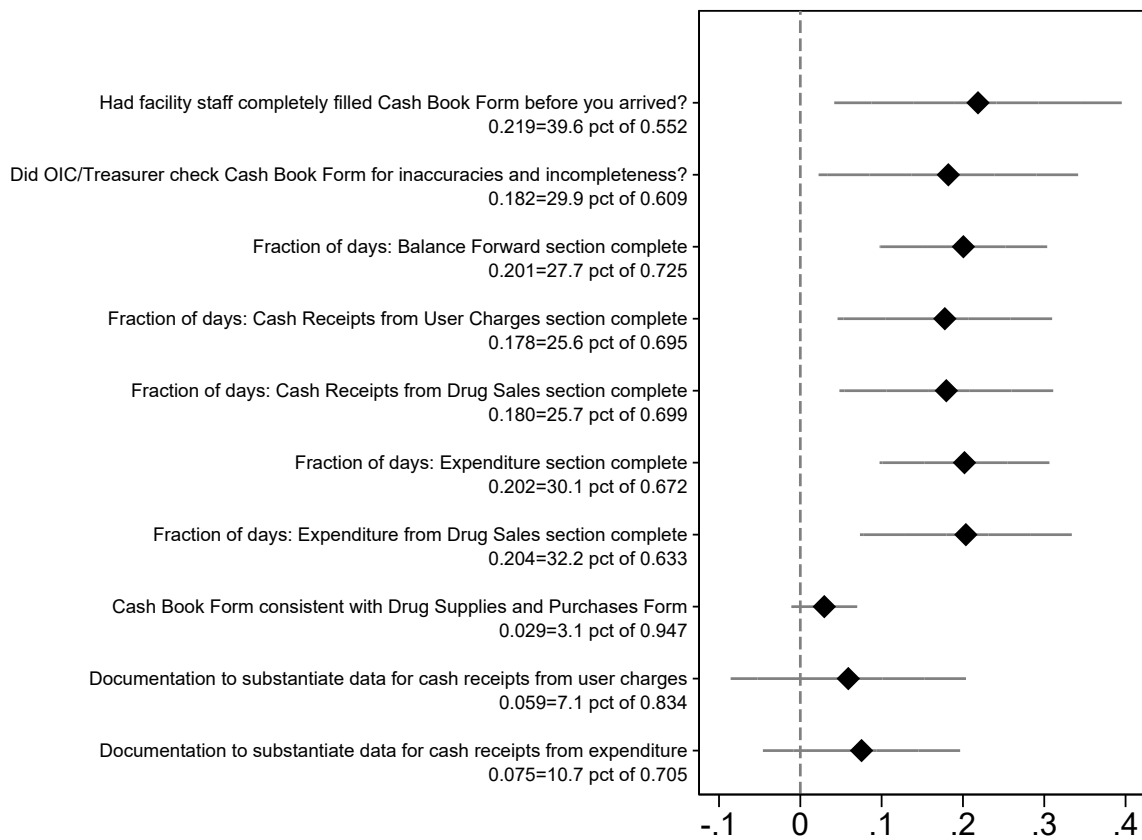
$$Y_{it} = \beta_{social}Social_{it} + \beta_{direct}X_i + \beta_{interaction}Social_{it} \cdot X_i + \beta_{state-treatment}Social_{it} \cdot Ekiti_i + \gamma_{LGA} + \epsilon_{it} .$$

Figure 1: Checklist score over time



In the figures above, average scores on the 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 in each state.

Figure 2: Impacts on individual items on checklist



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.

A Appendix

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 (0.085)	-0.014 (0.039)
Constant	0.716*** (0.060)	0.927*** (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.

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

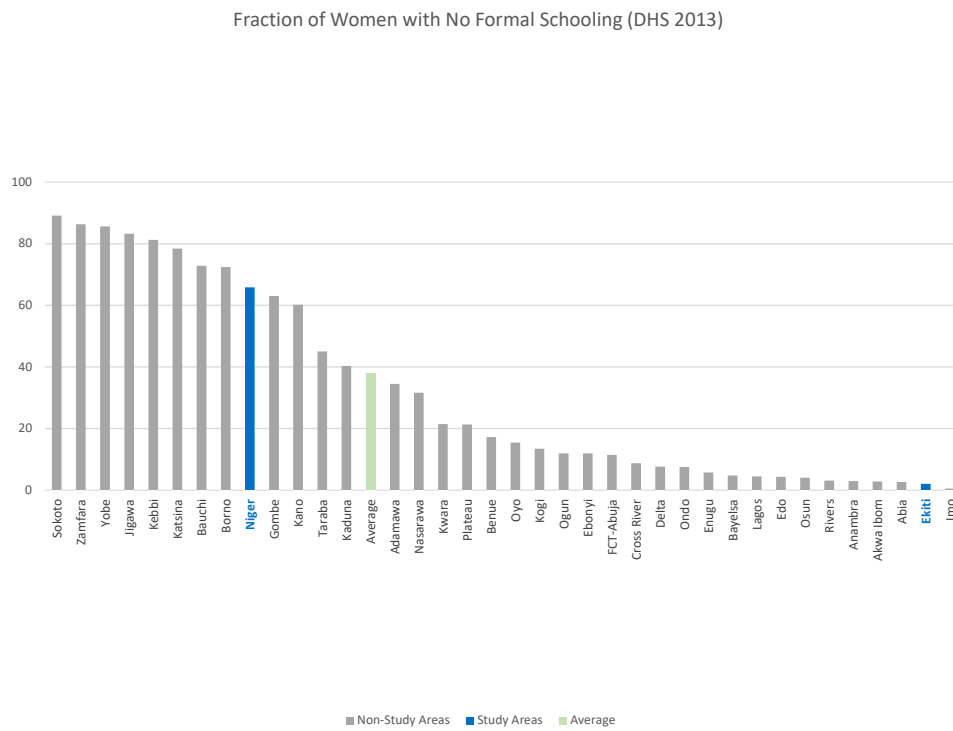
	EKITI		NIGER	
	[1]	[2]	[3]	[4]
Social recognition	0.123** (0.049)	0.098** (0.043)	-0.021 (0.028)	-0.010 (0.025)
Baseline score		0.286*** (0.091)		0.255 (0.185)
Constant	0.682*** (0.136)	0.462*** (0.149)	0.577*** (0.081)	0.457*** (0.110)
R^2	0.239	0.301	0.279	0.325
Observations	180	180	273	269

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.

Figure A1: Appendix Figure: 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

