Pre-analysis plan: ePDS impact evaluation in Jharkhand, India

Karthik Muralidharan* UC San Diego

Paul Niehaus† UC San Diego

Sandip Sukhtankar‡ University of Virginia

April 14, 2017

*UC San Diego, JPAL, NBER, and BREAD. kamurali@ucsd.edu.

†UC San Diego, JPAL, NBER, and BREAD. pniehaus@ucsd.edu.

‡University of Virginia, JPAL, and BREAD. sandip.sukhtankar@virginia.edu.

1 Introduction

How does biometric authentication impact the delivery of in-kind social programs? To contribute to the evidence base on this question we are conducting a randomized controlled trial (RCT) of the rollout of biometric authentication in the Public Distribution System (PDS) in the Indian state of Jharkhand during the period August 2016 - March 2017.

The Public Distribution System is India's primary food security program; it is implemented through a network of government-run Fair Price Shops (FPSs) that offer fixed monthly quantities of staple food grains and commodities such as rice, wheat, sugar, salt, and kerosene at subsidized prices to eligible low-income households.

The reform we study, which we refer to as "ePDS", changed the procedure for disbursing rations to beneficiaries in fair price shops. Prior to the reform beneficiaries established their identity by presenting their rationcard. After the reform, they established their identity by scanning their fingerprints on an electronic Point of Service (ePoS) device linked to the Aadhaar platform, India's national ID database. Beneficiaries thus needed a valid Aadhaar number to collect benefits post-reform. ePOS devices were introduced in study areas in Jharkhand between the months of August and November 2016. At present these devices are operated in three distinct modes - "online", "partially online", and "offline", with different frequencies of data sync required depending on the reliability of connectivity.[1]

Our study population consists of households eligible to receive PDS benefits under the 2013 National Food Security Act (NFSA), which was implemented in Jharkhand in October 2015. Under the NFSA, there are two categories of eligible household entitled to different benefit levels: Priority Households (PH) and Antyodaya Anna Yojana households (AAY). Eligibility was determined in 2015 in two steps: the government first declared eligible all households who met pre-set criteria according to the Socio-economic Caste Census (SECC) which had previously been conducted in 2011, and then invited those not included by this

[1]Offline: 16% of all machines in the state are "offline", deployed in areas with little to no internet connectivity. These machines have beneficiary data stored on them (of beneficiaries tagged to a specific FPS). These do not require beneficiary authentication and essentially function like manual registers. At the beginning of each month, from the 1st to the 4th, dealers have to get access to internet connectivity and upload the previous month's transaction data. If not uploaded, next month's transactions cannot take place. During upload, the machine also downloads next month's beneficiary data (entitlements, stock).

Partial online: 25% of machines in the state are "partially online". When there is internet connectivity, these function like the online machines. When connectivity is low, they function in offline mode. Beneficiary fingerprints are collected but not needed for disbursing ration. This data is to be uploaded between the 1st to the 4th of each month (for the previous month). If not uploaded, next month's transactions cannot take place. During this time, the machine also downloads next month's beneficiary data (entitlements, stock)

Online: Around 59% of the machines across the state are "online" machines which only function when there is internet connectivity. These machines do not have beneficiary data stored on them and require beneficiary biometric authentication. The transaction and authentication takes place in real time (response time from UIDAI server is approximately 22 seconds).

method to petition for inclusion by filling out a self-declaration of income, occupation and assets which was then (ostensibly) verified by village or block-level officials.

This document describes our plan for measuring the impacts of this reform, including the algorithms used for randomization and for sampling, primary and secondary outcomes of interest and our plan for measuring them, other descriptive measures we will calculate, and statistical methods we will use to analyze each measure. This document commits us to undertaking the analysis described in it, but is not intended to be an exhaustive list of all the analysis we (or others) may conduct using the same data.

2 Randomization

Jharkhand has 24 districts; we excluded 7 of these from our study because the government had already begun rolling out ePOS or was planning to roll out a distinct intervention (direct benefit transfer for kerosene, or DBT-K) in these districts. We then selected 10 of the remaining 17 districts by stratified random sampling, stratifying on three variables: an indicator for being above/below the median of district (centroid) latitude, an indicator for being above/below the median of district (centroid) longitude, and an indicator for above/below the median of the first principal component of a number of additional variables. These additional variables were the total population, the population density, the average number of beneficiaries per FPS, the literacy rate, the percentage of households with salaried income, the share of beneficiaries living in rural areas, and the percent of beneficiaries registered under the AAY scheme (see below). We used these 3 binary variables to classify the 17 available districts into 8 (2x2x2) distinct categories. We then sampled half of the districts in each category, rounding down to the nearest integer and using probability proportional to size (PPS) sampling, where we define district size as the number of Fair Price Shops it contains. This yielded 6 sampled districts. To reach a sample of 10 we then sampled 4 more districts from the remaining 11 using PPS sampling.

Our study districts contain 132 blocks (sub-districts). We assigned these blocks to treatment and control at a 2:1 ratio. (While a 1:1 ratio would yield a better-powered experiment, this was not politically feasible as government was eager to roll-out ePOS quickly.) Within each of our 10 study districts, we assigned blocks to treatment status using the following algorithm:

1. We divided blocks into rural and urban samples.

2. Within each grouping, we stratified blocks into groups of three by ordering them on the first principal component of three variables: the average number of unique Aadhaar IDs

per ration card, the average amount of ration claimed per PH household according to administrative records, and the average amount of ration claimed per AAY household. Within each group of 3 blocks we randomly assign 2 to treatment and 1 to control. Some districts also contained residual strata of fewer than 3 blocks; in strata containing two blocks we assigned one block to treatment and one to control, while in strata containing 1 block we assigned the block to treatment with probability $\frac{1}{3}$.

This procedure yielded 87 treatment blocks and 45 control blocks.

3 Sampling

We sampled households for a household-level panel study using the Government of Jharkhand's database of current PDS beneficiaries as the sampling frame, which was created as described above using the Socio-economic Caste Census and an appeals process. Our sample frame includes all households found eligible through either method. It does not include households that may in fact have met eligibility criteria but were mistakenly excluded.

For our baseline survey, we sampled 3,960 ration cards using the follow procedure. We first sampled 3 Fair Price Shops in each of the 132 blocks, for a total of 396 FPSs, using PPS sampling where we defined "size" as the number of ration cards assigned to each shop. Since some FPS draw beneficiaries from multiple villages, we then sampled one of the villages in the catchment area of each FPS, again using PPS with "size" defined as the number of ration cards in that village assigned to that FPS. Finally, we sampled 10 of the ration cards assigned to receive benefits from that FPS and village.

We sampled ration cards using stratified random sampling, with stratification on two variables. The first is the method by which the household became eligible; in our data this is denoted as either "SECC", "New", or "No data". The second is the benefit category to which the cardholder is entitled, which can be either "PH" or "AAY". We sampled ration cards so that the proportions of these six strata would be balanced with their proportions in the overall population of ration cards in our study districts. Specifically, in each village we calculated the proportion of households in each stratum and multiplied by 10 to obtain a (non-integer) target number of households to sample from that stratum. We then rounded these targets up (down) to the nearest integer with probability equal to the decimal portion of this target. For example, if our original target was 1.7 households in a stratum, we sampled 2 households from this stratum with probability 0.7 and 1 household with probability 0.3.

We stored each household's ex ante probability of being included in the sample and will use these to weight estimators to make them representative of the overall frame of ration

card holders. A households sampling probability is product of the probability its FPS is selected, the probability its village is selected within that FPS's catchment area, and the probability that individual household is selected with their village × FPS.

3.1 Determining the analysis sample

Our analysis sample will include all household we are able to locate and survey at endline, and all households we are able to reliably determine do not exist (i.e. "ghost" beneficiaries). We use a conservative procedure to determine that a household is a ghost.[2] For reference, of the 3,960 households we sampled we were able to complete baseline surveys of 3,418 (86.3%), including 6 households that were sampled twice (because they held more than one ration card) and will receive double weight in the analysis. We also identified 108 (2.7%) as ghost households (note that this classification may change after our endline). We began but were unable to complete interviews with 236 (6.0%) households. We were unable to survey 49 (1.2%) households that had migrated, 7 (0.1%) households refused to participate in survey, 70 (1.8%) households we could not reach because of danger from Naxal rebels or from extreme hazards (e.g. crossing dangerous bodies of water), 63 (1.8%) households that were located in a different FPS catchment zone from the FPS listed in our data, and 3 (0.0%) households for various other reasons (no adult present, data corrupted).

4 Primary outcomes of interest

We organize the analysis around three primary dimensions of program performance: value delivered to eligible beneficiaries (V), costs of delivery (C), and leakage (L) (which we think

[2]Specifically, survey enumerators use the following procedure. They first look up household member names based on ration card number on the Department of Food, Public Distribution and Consumer Affairs, Government of Jharkhand (2017) official website. Then they go to the village and ask people where this household is (using household member names pulled up). Many villages are divided into sections that are some distance from each other, called "tolas". They visit every tola of the village looking for the household. If not found, team searches for the household again on another day. If not found, supervisor after consulting with monitor makes a subjective determination about whether it is reasonable (according to geographic spread and population density) that all probable hamlets/habitations in the village/urban area were searched to find the household. If all areas could not be searched, the household is labelled "Not found". If it is determined that all areas had been searched, a separate surveyor/supervisor is assigned the task to contact one or more of these institutions to locate the household: ward member (lowest level elected representative), Anganwadi centre, mukhiya (elected head of panchayat); this person spends half a day dedicatedly searching for this household. If no one has ever heard of any of the household members, we mark this as SS04 in the survey status code section of the instrument (Section H in the household questionnaire). We get confirmation from 2 neighbors and record their names, addresses and mobile numbers in Section H. Additionally we will look at transactions data to see if these ration cards have been transacting at all to confirm that they are in fact ghost beneficiaries. We will attempt to survey at endline all households selected at baseline, including those we were unable to survey at baseline.

of as an additional cost of delivery). We define secondary outcomes (below) to help us interpret and contextualize changes in these primary measures.

4.1 Value delivered

Conceptually, we think of value delivered to beneficiaries by the PDS as the cash-equivalent value they assign to whatever commodities they purchased, net of statutory prices paid. (We treat illegally inflated prices or bribes as a form of corruption, captured below). If markets for these commodities are liquid then an appropriate measure of this is the difference between market and ration shop prices for each commodity multiplied by the quantity purchased:

V1: the sum over commodities of quantity purchased multiplied by the gap between local market value and statutory ration shop price, per household per month

We define local market value of a commodity in a given district here as the median price reported by households in control blocks of that district in the endline survey. Note that in practice a household may purchase in month M the allotment to which it was entitled both in month M but also in earlier months if if had not previously done so. By default we will simply compare disbursements and receipts in month M, ignoring this; if however we find that a substantial share of receipts in a typical month M are from previous month's allotments, we may create adjusted measures that incorporate this information.

If markets are not liquid or there are significant transaction costs, then this measure may not accurately capture the value to recipient households of the commodities they purchased. We therefore supplement it with a measure of stated willingness to accept:

V2: households stated willingness to accept in lieu of benefits received per month

Finally, we will also report impacts on measures of the quantities of the commodities (wheat, rice, sugar, salt, and kerosene) purchased as a simple and direct way to measure delivery.

V3a-e: quantities of each PDS commodity purchased per household per month

4.2 Costs of delivery

Conceptually we think two costs are relevant for assessing the ePDS intervention: the fixed costs of rolling it out, and any effects this rollout may have had on the per-unit costs of delivering commodities to beneficiaries via the PDS. The former is not an outcome, but we estimate it for the purposes of conducting cost-benefit analysis (below). The latter is an outcome, which we estimate as follows.

We consider distribution costs incurred at all steps in the PDS distribution process. This includes (a) costs to the state government of delivering commodities to the Fair Price Shops, (b) operating costs of the dealers running Fair Price Shops, and (c) money and opportunity costs incurred by beneficiaries to collect benefits from Fair Price Shops. Our primary cost outcome is the sum of these costs incurred per beneficiary household per month.

C1: total commodity delivery cost per household per month

We calculate this as the sum of the three components described above. To calculate costs to government of distributing commodities, we plan to use administrative data on disbursement of commodities and costs of government running the scheme. Ideally, we wish to obtain data on (i) the costs of physical commodity distribution, (ii) the ongoing costs of leasing ePOS machines and repairing them (for which the government is responsible if a vendor intentionally damages one) during our study months (January, February, March), and (iii) the cost of IT services provided by the National Informatics Centre in order to support the ePOS rollout. To reliably attribute these data to blocks, we need to obtain data on (i) at the block or FPS level. Data on (ii) and (iii) can be obtained at the district level or above, as we can then attribute these costs to treated blocks (if they are variable with respect to the scale of implementation) or to all blocks (if they are fixed). To then convert these various cost figures to the household level we will apportion them equally among the households in the geographic units within which they were incurred. For example, we will divide the cost of IT services purchased for a district equally among all beneficiary households (or in the case of variable costs, all beneficiary households in treated blocks) within that district.

C1a: state government's administrative costs of commodity delivery to FPS, per household per month

We do not yet know how reliable, comprehensive, and disaggregated the administrative cost data we will obtain from government may be. If in our judgment these data are not sufficiently relevant or reliable, we will not include administrative costs in our measure of total cost, and may instead treat them as a secondary outcome. We do not in any case expect our measure of distribution cost to be comprehensive of all conceivable such costs; for example, we do not expect to obtain data on costs of warehousing commodities and map these to specific blocks. We also intentionally omit the costs of procuring commodities from this measure, since those costs are centrally incurred by the state and there is no reason to expect the ePOS intervention to affect them.

To measure impacts on costs incurred by dealers, we will use data from our survey of dealers. We measure the time and money costs collecting and transporting PDS commodities to

the FPS, and the time cost of retailing those commodities. We monetize time costs by asking dealers how much they could have earned had they spent an hour on some alternative activity. We also measure money costs incurred maintaining or repairing the ePOS devices used (for which vendors are nominally responsible but which dealers sometimes incur anyway). We define total FPS delivery costs as the sum of all these individual items:

C1b: Fair Price Shop owners' costs of retailing PDS commodities, per household per month

Finally, to measure impacts on costs incurred by beneficiary households we will use data from our survey of households. We include both time and money costs, and measure these for each individual trip taken by a beneficiary to attempt to collect benefits.

We measure monetary costs as the transportation cost incurred for the trip to and from the FPS. In the event the beneficiary visited the FPS as part of a multi-purpose trip, we ask respondents to estimate of the amount of money they would have spent if they had only visited the FPS.

We measure time cost as the sum of (i) travel time and (ii) time waiting/transacting at the FPS. In the event the beneficiary visited the FPS as part of a multi-purpose trip, we measure travel time cost (i) as the difference between time actually spent and time they estimate they would have spent had they not visited the FPS. To monetize these time costs, we ask for each how much money the person who performed the relevant task could have earned had they done some other work instead.

We estimate the total cost of each trip to the FPS as the sum of the above, and the total cost incurred by the household in a given month as the sum of the costs of the individual trips.

C1c: total cost incurred by households to collect PDS benefits, per household per month

4.3 Leakage

Conceptually we define leakage as value issued by the government but not received by beneficiaries because of corruption. In practice the main driver of leakage from the PDS is thought to be the diversion of commodities during the distribution process, which may take place at many levels - during transport, from the FPS, etc. In addition to being diverted, commodities may also be temporarily stockpiled at various intermediate distribution centers (e.g. "godowns"); we will obtain data on changes in inventory at these centers if possible. Formally, we define net diversion as

$$diversion = disbursements - receipts - \Delta inventories \tag{1}$$

We plan to measure disbursements and changes in inventories using administrative data obtained from the state government at the block or ideally FPS level. If we are unable to obtain data at this level of disaggregation, we may instead need to use district-level data (which amounts to assuming treatment does not impact disbursements). We plan to measure receipts using household survey data, and as above define receipts as "receipts of allotments for month M" whether purchased in month M or subsequently.

A second contributor to leakage may be over-charging: dealers may sell commodities to beneficiaries but at prices above the statutory ones, or equivalently may demand bribes. We measure the value of overcharges per household per month as the difference between the survey amount of total amount paid for FPS rations and the sum (over ration products) of product the quantity received and the price-per unit the household is entitled to by the government. We do not explicitly ask households to report bribes paid to collect benefits as we found overcharging was the language that households understood and used in our pilots. We then define our primary measures of leakage as the sum of diversion and overpayment:

L1: the value of commodities diverted plus the value of overcharges paid by beneficiaries, per household per month

As above, we define the value of a commodity in a given district as the median value of the unit cost for that commodity reported by households in control blocks of that district in the endline survey.

We will also examine the impact of the intervention on each component of leakage individually:

L1a-d: subcomponents of leakage including value disbursed, value received, value of changes in inventory, and overpayment

Finally, we also examine impacts on the quantities of each individual commodity diverted, using the same approach used to calculate the aggregate value diverted measure (*L1-a*) above:

L2a-e: the quantity of each PDS commodity diverted, per household per month

5 Secondary outcomes of interest

In addition to the primary outcomes above, we also examine impacts on a number of secondary outcomes including measures of food security (F), ration quality (Q), decomposed measures of leakage (L), dealer profits and sustainability (D), as well as outcomes that enable us to test for spillover effects across blocks (S).

5.1 Food security

To measure a household's food security, we measure household consumption of a set of common food items and categories over a 7 day reference period. We then calculate three aggregate measures of nutritional status. First, we calculated the household dietary diversity score (HDDS) using the standard method from Kennedy et al. (2011) by summing indicators for whether a household has consumed any items of a standard food category in the past week. While the most common reference period for computing a HDDS is previous 24 hours, our reference period of previous 7 days is still acceptable practice (see Kennedy et al. (2011)) .

F1: household dietary diversity score

Second, we estimate average daily calories per capita by converting the quantity of each food category consumed in the last week into calorie equivalents using nutrient conversion tables. We then divide this estimate by 7 and the number of household members to obtain average daily calories per capita.

F2: average daily calories per capita

Together these measures are intended to capture the quality and quantity of food consumed. In addition, we calculate a single composite score based on dietary diversity, food frequency, and the relative nutritional importance of different food groups. We calculate this score by counting the frequency with which each food group has been consumed in the last week (capping this count at 7) and then taking a weighted sum of these frequency counts, using nutritional weights from World Food Programme (2008), to obtain an overall food consumption score.

F3: household food consumption score

5.2 Ration quality

To measure the incidence of quality issues (e.g. delivery of rotten grains) we ask households whether they have observed any of the most common quality issues during the last three months: whether wheat was rotten, diseased, or infested with pests; whether rice was rotten, diseased, or infested with pests; whether rice was broken or powdery; whether sugar was brownish in color or powdery. For rice and wheat, which sometimes require cleaning before use because of adulteration, we also measure cleaning time by asking each household how long it takes them to clean a kilogram of each grain purchased from the FPS. We do not ask quality questions about kerosene as we have not identified any equivalently objective

measures of kerosene quality. For each commodity other than kerosene we define an indicator equal to one of the household received a low-quality delivery of that commodity.

Q1a-e: an indicator for receipt of low-quality commodities, per household and commodity

We also ask households whether in their opinion their FPS dealer has adulterated any commodities delivered to them in the last three months.

Q2: an indicator for any perceived adulteration by the FPS dealer

5.3 Leakage decomposition

To better interpret our primary measures of leakage, we also examine the decomposition of commodity leakage measures *L2-6* across stages of the distribution chain. Distribution generally works as follows:

1. The Food Corporation of India procures commodities and delivers them to district-level warehouses.

2. Commodities are then transported to Fair Price Shops, either directly or (in some cases) after being routed through intermediate warehouses or "go-downs" situated at the block level.

3. Beneficiaries purchase rationed commodities from the FPS.

To better understand where in this sequence leakage occurs, we ask FPS dealers to report the amounts of each commodity they received in each of the past three months. Naturally, dealers have an incentive to under-report the amount they receive so that any theft will appear to have taken place higher up in the distribution system, rather than on their watch. We therefore interpret results based on these questions as bounding the decomposition of leakage rather than point identifying it.

Specifically we define upstream leakage as the difference between the amount disbursed to an FPS and the amount received by that FPS. We interpret our measure of this as an upper bound on its true value, given FPS dealer incentives to misreport.

L2a-e': the quantity of each PDS commodity diverted between central disbursement and the FPS, per FPS per month

We then define downstream leakage as the difference between the amounts received by the FPS and those received by households. We interpret our measure of this as a lower bound on its true value, given FPS dealer incentives to misreport.

L2a-e": the quantity of each PDS commodity diverted between receipt at the FPS and distribution to beneficiaries, per FPS per month

In addition to this decomposition, we also ask dealers to report any bribes they have paid in order to continue operating a FPS. To avoid double-counting we do not interpret responses to this question as a component of primary leakage; operating an FPS legally is estimated to be marginally profitable at best so that any bribe costs incurred by dealers would themselves have to be paid out of some other illicit gain. Instead we interpret this question as potentially providing some insight into how primary rents are shared through the corrupt ecosystem.

L3: amount of bribes paid to operate a FPS, per FPS dealer per month

5.4 Dealer profitability

We measure contemporaneous dealer profitability directly, and also elicit dealer intentions to continue operating a FPS as a proxy for anticipated future profitability.

We measure contemporaneous profitability as the difference between monthly revenues and monthly costs. We measure revenue as the sum of revenue from PDS operations (which is the sum of revenue from the sale of ration commodities and monthly commissions received to operate the FPS) and revenue from non-PDS operations (which is revenue from all other goods sold). We measure costs as the sum of costs incurred for rent, electricity, storing ration transporting ration, maintaining shop, labor, and supplies, as well as other miscellaneous costs. We measure each of these items by asking about their average monthly value over the study period. Note that we intentionally estimate the overall profitability of the shop, as opposed to incremental profit from distributing PDS commodities from it, as the two lines of business likely interact with each other.

D1: Average monthly profit, per FPS dealer

To elicit dealer intentions to continue operating a FPS in the future, we ask dealers whether they anticipate they will continue to operate a FPS in two years.

D2: Indicator for dealer's intention to continue operating an FPS in two years

5.5 Outcomes to capture spillover effects

We test for two categories of spillover effect that might lead to bias in our main impact estimates. First, any impacts the intervention might have on local commodity markets

could be transmitted through prices to other nearby markets. We therefore ask FPS dealers to report how much each rationed commodity currently sells for on the private market:

S1a-e: local market price of each commodity as reported by FPS dealers

We also ask households the unit price at which (if at all) they purchased that commodity in the last month, and take the FPS-median price.

S2a-e: median local market price of each commodities reported by households

Second, the intervention could affect control blocks through administrative channels; for example, district officials might have their attention diverted from control blocks in order to focus on solving problems in treated ones, or conversely might pay more attention to managing PDS issues that affect all blocks because of the rollout. To capture such effects we will examine impacts on administrative data on commodities disbursed by block, comparing control blocks in treated districts to blocks in non-study districts.

S3a-e: official disbursements of each commodity, by block and by month

5.6 Beneficiary lists

To understand how the intervention affected "churn" in the list of eligible beneficiaries, and whether it was effective in helping the government to weed out "ghost" households from that list, we compare the list of eligible households obtained prior to our baseline survey with an updated list obtained at around the time of our endline survey and calculate three metrics:

B1: the number of ration cards issued to newly eligible households, by block
B2: the number of ration cards cancelled for previously eligible households, by block
B3: the number of ration cards cancelled for previously eligible households also identified as ghosts in our baseline survey, by block

5.7 Descriptive statistics

5.8 Beneficiary preferences and perceptions

In addition to measuring outcomes of interest to beneficiaries, we also ask recipients in treatment areas their stated preference for the new method of accessing benefits as opposed to the status quo. This has the advantage that it potentially captures dimensions of program performance we may have missed in our outcome set, but the disadvantage that beneficiaries may inaccurately attribute changes in their experience that are in fact unrelated to the intervention to its rollout, or vice versa.

We directly ask beneficiaries in converted blocks within treatment mandals who had been exposed to the old PDS system to express their overall preference between the new system and the old system: "Overall, do you strongly prefer ePoS / weakly prefer ePOS / weakly prefer manual transactions / strongly prefer manual transactions?" We report proportions of households who indicated each preference.

We then ask them to choose any of the following reasons for their preference from a pre-determined list of 5 (9) options for control (treatment) areas. We will use these questions as suggestive evidence of performance, and report proportions of households who indicated each reason.

6 Statistical methods

6.1 Balance

We will test for baseline balance on (i) variables used to stratify the randomization of blocks to treatment and control (which are the average number of unique Aadhaar IDs per ration card, the average amount of ration claimed per PH household, and the average amount of ration claimed per AAY household), and (ii) primary outcomes. In any cases where the variables needed to compute an outcome measure were not collected at baseline, we will either test for balance on the closest reasonable approximation or omit that variable.

We will test for balance using the specification

$$Y_{hfbs} = \alpha + \beta Treated_{bs} + \delta Stratum_s + \epsilon_{hfbs} \tag{2}$$

for outcomes that are defined at the household level, and analogously for outcomes defined at higher units of aggregation, clustering standard errors at the block level when the unit of analysis is smaller than the block.

6.2 Treatment effect dynamics

Our endline survey covers PDS beneficiary experiences in January - March 2017, enabling us to construct a retrospective panel for many outcomes. It is a priori unclear whether to expect effects on outcomes in these months to be the same, or whether to expect heterogeneity (because of ongoing implementation of the intervention or because of other external factors, e.g. seasonality). We will therefore estimate effects on such outcomes for (i) the last month only, (ii) the last three months pooled, and (iii) the last three months including a linear trend. If, examining all of our primary outcomes, we find that we regularly reject the null

that the linear trend is zero in (iii), we will interpret (ii) as a hypothesis test but privilege results from (i) for point estimation; otherwise we will treat point estimates from (ii) as preferred.

6.3 Estimation

We will report ITT estimates, which compare average outcomes in treatment and control areas. Our primary outcome measures are defined at the household-month level, which is the unit at which we will analyze them. We will analyze other outcomes defined at higher units of aggregation (e.g. the FPS) at the level at which they are defined, using analogous specifications. All regression specifications will include stratum fixed effects (with a single fixed effect for all strata of size 1) and will be estimated using inverse sampling probabilities as weights:

$$Y^t_{hfbs} = \alpha + \beta Treated_{bs} + \delta_s + \epsilon^t_{hfbs} \tag{3}$$

For outcomes which are measurable (or approximately measurable) from baseline data, we will also estimate specifications that condition on the lagged value of the outcome to improve precision:

$$Y^t_{hfbs} = \alpha + \beta Treated_{bs} + \gamma Y^0_{hfbs} + \delta_s + \epsilon^t_{hfbs} \tag{4}$$

For these regressions we will omit from the sample households for whom baseline values were not observed.

To test for geographic spillovers, on outcomes such as market price, we estimate specifications of the form

$$Y^t_{hfbs} = \alpha + \beta Treated_{bs} + \gamma \overline{T}^R_{fbs} + \delta_s + \epsilon^t_{hfbs} \tag{5}$$

where $\overline{T}^R_{fbs}$ is the fraction of Fair Price Shops located within R kilometers of f and in a different block which are treated, for a range of values of R. If we find evidence of significant effects, we may also report adjusted estimates of impacts on our primary outcomes using the methodology in Muralidharan et al. (2017).

To test for administrative spillovers across blocks, we estimate a difference-in-difference model in which control blocks of study districts are" treated" and blocks in non-study districts act as the control group:

$$Y^t_{bs} = \alpha + \beta StudyDistrict_{bs} * Post^t + \delta_b + \pi^t + \epsilon^t_{bs} \tag{6}$$

6.4 Inference

We will conduct inference using the above models using standard errors clustered at the block level.

We address the risk of mistaken inference due to multiple hypothesis testing in two ways. First, we pre-specify a small number of primary outcomes (above). Second, we will report standard errors adjusted to control the false discovery rate (FDR) within clearly defined families of outcomes such as the sub-components of a primary outcome or outcomes defined across each of the five PDS commodities. Specifically, we will report FDR-adjusted p-values for the following families of outcomes:

- *V3a-e*
- *C1a-c*
- *L1a-d*
- *L2a-e*
- *Q1a-e*
- *L2a-e'*
- *L2a-e"*
- *S1a-e*
- *S2a-e*
- *S3a-e*

(Note that we also specify below how we will address multiple hypotheses testing when examining heterogeneous impacts.)

6.5 Attrition

To address the possibility of bias due to attrition from our sample (specifically, households in our original sample whom we are unable to survey at endline) we will report aggregate attrition rates by study arm, and will test for differential attrition from different arms by regressing an indicator for missingness on a treatment indicator and its interaction with baseline values of primary outcome variables. If we consistently find evidence of significant differential attrition using an F-test of these interactions across covariates and outcomes, we will report Lee bounds for treatment effects as a robustness check.

6.6 Heterogeneity analysis

We will test for heterogeneous effects along three categories of variation: measures relevant for policy-making, measures likely to capture differences in the quality of ePOS implementation, and the baseline values of outcomes. Specifically, we will test for heterogeneity impact with respect to the following variables:

1. Policy-relevant measures

    - An indicator for whether a household is above/below median value of predicted income given assets and household characteristics at baseline.
    - An indicator for whether a household belongs to a scheduled caste or tribe.
    - An indicator for whether the average years of education of the two most schooled household members is above / below media.

2. Measures expected to predict implementation quality

    - An indicator for whether the FPS is above or below the median predicted subjective household rating of their FPS given distance, unsuccessful trips, collection time, indicator of whether household was overcharged for rations, and relationship with dealer (length of relationship, same caste (jati), same religion).
    - An indicator for being located in an urban area
    - Indicators for the ePOS operating model being used by the FPS - one of online, partially online, offline.
    - An indicator for being above or below median cellular network signal strength, as measured in decibels above a reference level of one milliwatt (dBm). If we are able to obtain information on the minimum number of decibels needed for a ePOS machine to function effectively, however, we will use this as the threshold rather than the median to define the indicator.

3. Baseline values of the outcome variable

For each predictor of heterogeneous impact we will regress our primary outcomes on treatment and its interaction with that predictor. We will report standard errors corrected to control the false discovery rate within each of the three main groupings 1-3 above.

6.7 Cost-effectiveness analysis

To assess the cost-effectiveness of the intervention, we will calculate its impact on the overall flow costs and benefits of the intervention. If these are negative, we conclude the intervention cannot have been cost-effective; if positive, we then divide the flow gain by the fixed costs of rolling out the intervention to estimate the effective return on investment.

Specifically, we define the total change in flow value as the change in value received by beneficiaries (*V1* or *V2*) minus the change in delivery costs (*C1*) minus the change in estimated leakage (*L1*) multiplied by a range of relative Pareto weights in $[0, 1]$ to capture the different possible values policy-maker makers might attach to transfers to the corrupt.

If the sum above is positive, we divide it by the estimated fixed costs of setting up the intervention based on government records, including the cost of training government officials and dealers and the cost of infrastructure (organizational and physical) built to support the ePDS system. We will also capture costs incurred by beneficiaries in order to participate in the new system, including costs incurred over the last three months to get new Aadhaar cards made, to get Aadhaar numbers seeded to ration cards, and to get biometric information corrected/updated on ration cards.

7 Comprehensive list of data sources

This is a comprehensive list of relevant data sources that we have obtained or are aiming to obtain for this study.

- *Administrative-NFSA Beneficiary List*
  We used the Government of Jharkhand's database of current PDS beneficiaries as the study's sampling frame, which was created as described above using the Socio-economic Caste Census and an appeals process.

- *Baseline household and dealer survey*
  We conducted a baseline survey (June 17, 2016 - July 25, 2016) of dealers at sampled FPS and beneficiary households who possessed sampled rationcards.

- *Endline household and dealer survey data*
  We will collect endline survey data of dealers at sampled FPS and beneficiary households who possess sampled rationcards.

- *Administrative transaction data at beneficiary × commodity level*
  We aim to obtain official monthly transaction data for the distribution of each ration

commodity to each rationcard for every rationcard in the state of Jharkhand for for as many months as possible before, during and after our study months (ideally January 2016 - June 2017).

- *Administrative inventory of each ration commodity at district warehouses and FPSs*
  We aim to obtain official data on the inventory of each commodity at district warehouses and FPS for as many months as possible before, during and after our study months (ideally January 2016-June 2017).

- *Administrative spending on PDS (at block × month level or lower)*
  We aim to obtain data on administrative costs of government running the scheme on (i) the costs of physical commodity distribution, (ii) the ongoing costs of leasing ePOS machines and repairing them (for which the government is responsible if a vendor intentionally damages one) during our study months (January, February, March), and (iii) the cost of IT services provided by the National Informatics Centre in order to support the ePOS rollout. To reliably attribute these data to blocks, we need to aim to obtain at the block or FPS level.

- *Administrative official disbursement for each ration commodity to district warehouses and FPS*
  We aim to collect official data on official disbursement for each ration commodity to district warehouses and FPS for as many months as possible before, during and after our study months (ideally January 2016 - June 2017). We will use this data in our leakage analysis.

- *Administrative rural vs. urban classifications*
  We use administrative classifications of each block as rural or urban. We will use this binary classifier to assess differences in primary outcomes by rural and urban areas.

- *Administrative data on type of ePOS machine implemented (online, partially online, offline)*
  We aim to obtain administrative data on the rollout of ePOS machines by type in each FPS. We will use this data to assess differences in primary outcomes by the type of ePOS machines implemented.

- *Network connectivity at FPS shops*
  We will have survey enumerators check for service/data connection for major cellular providers at each FPS during the endline dealer survey. To obtain a standard measure of signal strength, they use SIM cards from the most popular telecom providers to check

signal strength of 2G network (closest approximation to GPRS network we can get) in decibels above a reference level of one milliwatt (dBm) and ASU (arbitrary signal unit) using their phones or survey tablets.

References

Department of Food, Public Distribution and Consumer Affairs, Government of Jharkhand, "Jharkhand Government PDS System SECC Cardholders," April 2017.

Kennedy, Gina, Terri Ballard, and Marie Claude Dop, "Guidelines for measuring household and individual dietary diversity," Technical Report, Food and Agriculture Organization of the United Nations 2011.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, "General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India," *Working Paper*, 2017.

World Food Programme, "Food consumption analysis: Calculation and use of the food consumption score in food security analysis," 2008.