Short Run Effects of Skill Training for the Unemployed Youth in India*

Rashmi Barua, Pratibha Joshi and Swati Singh[1]

July, 2022

Abstract

Using administrative data and multiple methodologies, including parametric and non-parametric approaches, we evaluate a skilling and placement initiative of the Government of India. We study potential job seekers' employment outcomes, among those who have registered for this skill training program in four major states of India. The results for the baseline and follow up survey data for the treatment and comparison groups show that there is a large positive and significant impact of training on wage employment in the short run. The results are robust to employing recently developed non-parametric econometric methodologies. Further, while there are no differences by gender in the effect of training on wage employment, trained women are also more likely to engage in self-employment. One of the key reasons that hinders youth employment in developing countries is large skill gap due to lack of formal education. Our findings suggest that policy-makers, particularly in developing countries, can consider skill training programs as an essential policy tool for increasing employability, particularly among women.

---

* We are grateful to the Ministry of Rural Development, GoI, for providing access to Kaushal Panjee MIS information on skills training registrants. We also thank Azim Premji University for the financial support that made this study possible. Special thanks to the project team at LEAD at Krea University for their support with data collection and research activities and Abhiroop Mukhopadhyay for comments and suggestions. All errors are our own.

[1] Rashmi Barua, Centre for International Trade & Development, Jawaharlal Nehru University, New Delhi, barua.bhowmik@gmail.com; Pratibha Joshi, pratibha.joshi25@gmail.com ; Swati Singh, Centre for International Trade & Development, Jawaharlal Nehru University, swatisingh7march@gmail.com

1. Introduction

Nearly 30 million Indians between the age of 20 to 29, 85 percent of the unemployed, were actively seeking employment in 2021 (Centre for Monitoring Indian Economy (CMIE)). Alarmingly, a large population of the unemployed are educated. Youth unemployment is a significant worry for any economy as early-life unemployment is directly associated with social conflict, financial hardships, psychological effects and detrimental effects on long run labour market outcomes (Daly and Delaney 2013; Fougère et. al. 2009; Kahn 2010). Just like India, many emerging and low-income nations face difficulties integrating young people into the job market. Fuelled by the global financial crisis of 2007–2008, population ageing, high unemployment rates, and automation, policy makers have resorted to active labor market programs (ALMP) to promote a smooth entry into the job market (McKenzie, 2017). The aim of ALPMs is to introduce innovative labour market interventions to create employment opportunities and improve job matching to fill the gap between labour supply and demand.

Some recent ALMP interventions include the provision of microloans, capital infusions, and cash grants for entrepreneurship. However, some scholars argue that these policies are gender-biased. Financial injections such as cash loans/grants given to women may be seized by their husbands and other household members, resulting in inefficiencies (Maitra and Mani, 2017). Alternative ALMP policies include vocational/skills training provision, wage subsidies, job search assistance, and other employment and social protection interventions such as public works programs. Vocational training programs provide sector and skills specific training, especially to less educated unemployed youth, which might be helpful for developing countries due to the limited formal job opportunities available to the less educated. These interventions result in human capital accumulation specific to the individual and cannot be confiscated by others. Further, skill certification by these programs acts as a signalling mechanism to employers, thereby improving job matching (Hirshleifer et al., 2016).

Extant literature on the evaluation of public programs has evolved in the debate around the success of these training programs. While some studies find these interventions effective, McKenzie (2017) and Heckman et. al. (1999) argues that many of these policies/programs are ineffective and have an insignificant impact on employment outcomes, contrary to what policymakers believe. In a meta-analysis of 97 studies conducted between 1997 and 2007, Card et. al. (2010) find that classroom and on-the-job training programs are not effective in the short run, but have positive relative impacts after two years (medium term). Moreover, they find that interventions targeted towards the youth are generally ineffective and there are no significant differences by gender. Contrary to this, Bergemann and Van Den Berg (2008) find that training programs positively affect labour market outcome of women, particularly in regions where female labour force participation is low.

Theoretically, labour supply response of unemployed women to ALMPs should be higher in labour markets with search frictions and low female labour force participation (Killingsworth and Heckman; 1986). In an intertemporal framework, an unemployed woman chooses between the decision to work or to bear children (and/or choose to be a housewife) while the spouse is the primary income earner. In this scenario, a job training program that upgrades skills and thereby increases the labour market opportunities may increase participation. Lower is the female labour force participation rate in a region, the more responsive would the female labour supply be to change in wages. A second reason for higher responsiveness of females to training programs is that it may reduce the noise in measures of female productivity leading to lower statistical discrimination by employers (Bergemann and van den Berg, 2008).

In 2014, the Ministry of Rural Development (MoRD), Government of India, launched the "Deen Dayal Upadhyay Grameen Kaushalya Yojana" (DDU-GKY), a free-of-cost, wage employment training program that offers skill training to rural poor youth, intending to place them into gainful wage employment. The primary objective of this paper is to study the effectiveness of the DDU-GKY program in enabling livelihoods for trainees. A first step for any rural youth seeking to upgrade skills is to register online in the Kaushal Panjee (KP) portal. Our primary data comes from administrative data from this portal. We obtained individual level information on more than 600,000 registered candidates from four states of India. The second source of data comes from phone surveys that were conducted for a random sample of approximately 12,000 registrants from the KP portal 6 months to two years after registration.

Investigating the effect of skill training on employment is challenging due to the selection bias associated with trainees. Skill trainees are not randomly chosen and generally tend to be positively selected from the general population as they are intrinsically motivated to change their employment situation. Though we do not have experimental data to infer causality directly, we deal with the endogeneity concerning trainee selection using a battery of methods.

First, to difference out unobserved characteristics, we use as a control group the subset of non-trainees who self-report their interest in completing the training within six months from the time of survey[2]. This control group is arguably similar to the treatment group since unobservable characteristics such as the desire for work, alternative employment options, and talent would not differ much. We then use a canonical Differences-in-Differences (DID) model and compare pre- and post-training employment between the trainees and the control group using a combination of administrative and survey data. We control for a rich set of explanatory variables including very detailed measures of risk and social network, two important omitted variables that are usually not available in survey data.

Second, we employ two nonparametric approaches: propensity score matching (PSM) and minimum bias estimator (MBE) (Heckman et al., 1999; Millimet and Tchernis, 2013). PSM assumes that conditional independence assumption (CIA) holds which implies that potential outcomes are independent of treatment and thus revolves the issue of self-selection. Further, we used MBE to reduce the bias that might arise due to the failure of the CIA assumption.

The canonical DID estimates suggest that those who are trained under DDU-GKY are 13 to 17 percent more likely to be wage employed depending upon the control group used. The two different non-parametric approaches confirm these results. Several studies suggest that differences in results between the experimental and non-experimental impact estimates of ALMP are small and statistically insignificant (Card et. al. 2010). Therefore, we cautiously conclude that regardless of the methodology employed, DDU-GKY training is associated with a significant increase in the probability of being wage employed a few months after the completion of training.

Next, we examine the effect of skill training on employment levels by the gender of the trainee. Results suggest no heterogenous effect by gender in the probability of getting wage employed. However, women (unlike men) are also more likely to engage in self-employment after undergoing skill-training. This is an important result for public policy

[2] There is no information on whether they joined training after six months or not.

as it suggests that skill training may be helpful in empowering women to engage in a business activity and/or helps in reducing information asymmetries.

The rest of the paper is structured into the following sections. The section 2 conducts a literature review focusing on findings for similar programs used in other developing countries. The specific DDUGKY program details are presented in Section 3, along with information on the intake sample and the survey process. The definition and methodology used for the analysis is discussed in Section 4. Our main results and robustness checks of the impact of training are described in Section 5. We conclude in Section 6.

2. Literature Review

Active labour market policies (ALMP) towards livelihoods range from interventions such as vocational/skills training provision, cash grants for entrepreneurship, wage subsidies, job search assistance, and other employment and social protection interventions such as public works program. There are three main categories of traditional ALMPs. To begin with, supply-side policies work towards increasing the employability of people (often jobless and youth) through vocational/skill training programs. Then there are demand-side policies that aim to boost labour demand by subsidizing labour costs for businesses through employment grants and incentives. Lastly, search and matching aid programs try to reduce labour market barriers caused by supply and demand mismatches. McKenzie (2017) reviews recent experimental studies on ALMP and finds that vocational skills training, wage subsidy programs, and job search assistance, on average, increase employment modestly. However, the author highlights the potential of supply-side interventions to reduce sectoral and spatial frictions that help workers access different labour markets. While there is an extensive literature on evaluations of ALMPs, below we discuss the literature on randomized evaluations of supply-side policy interventions across developing countries. Heckman, LaLonde and Smith (1999) conducts a more complete review of the literature.

The extensive literature on randomised evaluations of youth training programmes (such as the Job Training Program Act (JTPA) and Job Corps) in the United States has found little impact on adult earnings (Heckman et al., 1999; Schochet et al., 2003). The scenario may differ in developing countries compared to developed nations, as developing and low-income countries may yield higher returns on training programmes due to low levels of skills of the population to start with. In developing countries, opportunities in the formal sector are limited to the more educated worker (Attanasio et al., 2011), while the youth population consists of a large section of dropouts from schools/colleges or completed secondary level of education. As a result, training programmes may be more successful wherever skill shortage is likely to be the most significant barrier to employment (Betcherman et al., 2004). Attanasio et al. (2011) evaluated a randomized experiment of the vocational training program on employment and earning levels of young unemployed men and women in Colombia. They concluded that the program had a significant impact on female wage earnings and employment prospects but had no effect on men. Friedlander et al. (1997) and Maitra and Mani (2017) find that skill training programs are effective for women in promoting employment. At the same time, other studies that investigate gender equality in the impact of randomised training programmes have found no significant effect on employment and earnings for either gender (Card et al., 2011; Hirshleifer et al., 2016).

Some studies look into differences in the treatment impact based on the type of training delivered. Hirshleifer et al. (2016) distinguish the impact of training provided by government and private institutions. They find that training provided by private providers is more effective than training provided by government training institutes. However, the benefits of privately provided training are marginal and fade with time. Alfonsi et al. (2020) observed the impact of firm-provided training vs. vocational training on unemployed youth. After four years of training, both the groups had better wages, employment rates, and labour market outcomes compared to a control group of non-trainees. However, the gains to the vocational training group were more long-lasting than the firm-provided training group.

3. Institutional Context and Data

Government of India's youth skills training scheme, DDU-GKY is designed to bridge the skill-gap among India's rural youth. DDU-GKY is a free of cost wage employment training program which offers job training to rural poor youth[3] who are 15 to 35 years of age (the age limit is 45 for minority groups), with the aim of subsequently placing

[3] Eligibility for DDUGKY scheme is any of the following: Age 15- 35 (extended to age 45 for Minority groups and women), and *any* of the following documents: Below Poverty Line (BPL) card, BPL ration card, Rashtriya Swasthya Bima Yojana (RSBY) cards i.e. poor households who benefit from a government run health insurance program, family member being part of an Self Help Group (SHG), family member participating in the rural employment guarantee program (MGNREGS).

them into gainful employment.[4] Trainings are 3 to 12 months long, and are implemented by Project Implementation Agencies (PIAs) who bring together the trainers, experts, sectoral knowledge and infrastructure required for skill development. Training is free of cost for attendees, and takes place in a PIA set training centre. The training centre can be residential with food and lodging provided or non-residential if the training centre is close to the candidate's location.

A first step for any rural youth seeking to upgrade skills is to register online in the Kaushal Panjee (KP) portal.[5] Since its inception in October 2017, rural poor youth from 690 districts across all states of India have registered for employment training on the KP portal. Training Partners, PIAs, and Banks can use KP to contact registered candidates for training or jobs.

Apart from their contact information, registrants are asked to provide detailed information about their educational attainment, family, income, and the sectors and trades of their interest for subsequent matching with PIAs. Further, they are asked if they are willing to move to a different state or district for work, and if they prefer self-employment or wage employment. This provides us with preferences for occupation before a candidate has undertaken any training, and is assumed to be looking for livelihood opportunities. KP portal also contains information on their employment status before training.

3.1. Kaushal Panjee MIS

Our data comes from administrative data from the Kaushal Panjee portal. We obtained individual level information on 646,679 registered candidates from four states i.e., Gujarat, Madhya Pradesh, Odisha and Tamil Nadu (Map A.2 in appendix). These candidates registered on the KP Portal from October 1, 2017, the date of its inception, to 5th February 2019. These states were chosen on the basis of two criteria. First, all 4 states had, as of 5th February 2019, more than 15,000 registrants, a sample size that was sufficient for our follow up phone surveys where we expected low pick-up rates

[4] In addition to ddugky, the Rural Self-Employment Training Institutes (RSETI) is an initiative of Ministry of Rural Development (MoRD) to impart training and skill upgradation of rural youth geared towards entrepreneurship development. RSETIs are managed by banks with active co-operation from the Government of India and State Governments.

[5] Their enrolment relies on two key channels. First, direct outreach to eligible youth by State Rural Livelihoods Mission (SRLM) field staff and Project Implementation Agencies (PIAs) who register interested candidates for training. Second, indirect outreach through information dissemination via key stakeholders, word of mouth, advertising in newspapers and radio.

and high attrition. Second, we chose two states with high in-migration, Tamil Nadu and Gujarat, and two states with high out-migration, Odisha and Madhya Pradesh, because jobs and placements frequently require relocation to a different state or district. The majority (94%) of the KP sample was unemployed at the point of online registration and almost no one had ever migrated before for work opportunities[6]. Thus, between the time of KP registration and our phone survey, we are able to observe the changes in skill training status and employment outcomes of a random sample of registrants.

3.2. The Phone Survey

We used mobile number data from KP to conduct phone surveys as conducting door to door survey of candidates would be very expensive, and as this sample is anticipated to be on the move as they seek out livelihood opportunities. Studies are increasingly looking at collecting data through phone surveys, for instance, Muralidharan et al. (2018) used phone survey to improve service delivery of direct-benefit-transfer to farmers in Telangana. The details of survey collection are presented in the Appendix (see figure A.1 and A.2). Data from the phone survey helps us check the probability of reaching the intended registrant using contact information from KP data. Table A.1 is a sampling cascade and summarizes the process of how the final sample was reached. Surveyors attempted to call 57,261 registrants. Of these attempts, 12,144 calls materialized into full-length surveys, which lasted for 18 minutes on average. Of the successfully surveyed individuals, the identity of 10,151 respondents (83.6 percent) matched with that of the individuals for whom we had KP information. The names of the remaining 1,992 individuals reached using KP mobile numbers led us to respondents whose names did not match with the names listed in KP for the same mobile number.[7]

The phone survey data cross validates questions that are asked in the KP form already such as candidates' gender, age, disability, employment preferences. The survey also provides new information about candidates which has not been collected in the KP form. This includes detailed information on respondents' employment history and preferences, risk appetite, social networks, and awareness and outreach of DDU-GKY programs.

[6] 99.94 % of the KP sample reported that their present district of residence is same as their permanent district as per domicile.
[7] Though we include these observations in the study, our results are robust to dropping them.

The combination of the KP and phone survey information gives us a unique dataset, which allows us to view respondents' information, preferences and employment outcomes in two time periods. For the respondents who we can correctly identify in both the surveys, we are able to see how time varying factors evolve between time period 0 (data from KP portal) and time period 1 (phone survey). Most importantly, we are able to identify the individuals who took DDU-GKY training during the period between registering in the KP portal and our phone surveys.

3.3. Data Description

Table 1 shows the summary statistics from the KP registration data for the sample reached by us. Among the KP registrants, 32% preferred wage employment, 24% prefer to be self-employed. While 38% reported that they were willing to move to another state for work as opposed to 58% stating their willingness to move intra state[8]. Note that unemployment rate in the KP sample is very high, only 6% of the sample reported that they were employed at the time of registering into the KP portal (712 respondents out of the sample of 12144 respondents). Descriptive statistics for four additional variables which are used as control variables in certain specifications are shown, namely, whether the respondent belonged to a minority group[9], whether the household is Below Poverty Line (BPL), whether any household member worked as a MGNREGA[10] worker and belonged to a Self-Help Group (SHG).

Table 2 summarizes the background characteristics from the phone survey for the four states. The average age is 25 years and only 38% of the sample are women with gender balance in Tamil Nadu (50%) and an unbalanced sample in Madhya Pradesh (20%). 40% of the sample is married and the average household size is four. Odisha has the lowest percentage of self-employed in the family (12%) while for other states, the probability that either parent is self-employed is approximately 30%. The education levels of the respondents are high, only 8% of the sample has not completed 9th grade while 42% of the sample has completed high school. Thus, combining information on individual and household characteristics from the KP and phone survey, we find that

[8] There were two questions on migration preferences in the KP form which were also replicated in our phone surveys: "Are you willing to migrate to another state for work?" and, "Are you willing to migrate to another district within your state for work?"
[9] There are six religious groups that are recognized as minorities in India, namely, Muslims, Christians, Sikhs, Buddhists, Parsis and Jains.
[10] Mahatma Gandhi National Rural Employment Guarantee Act provides employment opportunity to the rural poor for up to 100 days in the financial year. The nature of the work under this scheme is unskilled Labour work.

the sample is representative of rural educated youth who were actively seeking employment and interested in upgrading their skills. Compared to the KP data, where almost 96% respondents were unemployed, the unemployment rate had reduced to 39% at the time of the survey. Among those who are employed, 21% are wage earners while 8% are in self-employment. The remaining workers are either out of the labour force (21%) or working as agricultural labour (2%) and casual labourers (8%). While Tamil Nadu has the lowest rate of unemployment in the sample, it has the highest rate of wage employed (34%) and lowest rate of self-employed (5%). Unemployment rate, at 45%, is highest in Gujarat.

3.4. Measures of Risk and Social Networks

To estimate the effect of DDU-GKY on employment outcomes, apart from socio-economic variables, we control for risk attitude and social networks because employment outcomes are affected by individuals' risk preferences and network characteristics.

Individual risk attitudes are one of the primary determinants of self-employment (Hartog et al., 2002; Brown et al., 2011). Willingness to take risk is positively linked to the likelihood of self-employment. We measure risk attitude in two different ways. The first measure of risk is a self-reported general risk assessment based on the question: *"How do you see yourself: are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks?"* Responses are given on a 10 point Likert scale where 1 means not at all willing to take risks and 10 means very completely willing to take risks. The second measure of risk is based on a lottery game with hypothetical money. We asked respondents to play a coin toss game with a fixed winning amount (Rs. 500 or approximately USD 7) if the flip of the coin showed heads, but we gradually increased the losing amount if the coin toss yielded tails. The losing amounts were Rs. 25, 75, 150, 250, and 350. If a respondent did not want to play the game at all, we assigned them the lowest risk score (value of 1). If the respondent played the game only for the loss amount of Rs. 25 but refused to play for higher values, we assigned them a risk score of 1. Similarly, if the respondent decided to play even with a loss of Rs. 350, their risk appetite was highest (with a score of 7).

The self-reported and lottery-based risk scores present some interesting findings. Risk attitudes elicited through the two methods differ from each other. The Spearman correlation coefficient indicates a very low correlation of 0.16 between the two measures[11]. There are also differences in the two risk measures across states. For example, while respondents in Tamil Nadu have the lowest risk-taking abilities according to the lottery-based measure, the self-reported risk measures suggest they have a very high risk score. Table 2 shows that Madhya Pradesh has the highest average risk lottery measure (3.30) and Tamil Nadu has the lowest (2.04). Interestingly, Madhya Pradesh has the largest percentage of youth who are self-employed and the lowest percentage of youth who are wage workers, whereas Tamil Nadu has the lowest percentage of youth who are self-employed and the highest percentage of youth who are wage workers. These figures, albeit correlational, are in line with the risk attitudes and self-employment theory (Hartog et al., 2002; Brown et al., 2011).

The theory of social network and self-employment suggests that one's social network provides support financially and emotionally, reducing the cost of employment (Allen, 2000). An effective social network consists of considerable size and close family/friends who encourage entrepreneurial activities. Social networks work as a control and support mechanism for migrants' employment, earnings, and other labour market outcomes (Munshi, 2003; Beaman, 2012). To proxy for networks, we use two variables. The first, the number of blood relatives in the district, is an exogenous proxy for kinship networks. Second, we use the number of friends met in the past week as a proxy for the regularity of contact between the individual and network members.

As shown in Table 2, measures of social networks are smallest for Odisha and largest for Tamil Nadu. On average, youth from Odisha and Tamil Nadu have 29 and 39 blood relatives living in the same district, respectively. Similarly, in Odisha, respondents meet six friends per week, while in Tamil Nadu, they interact with nine friends per week. Interestingly, Odisha also has the highest number of youth who prefer wage employment, and Tamil Nadu has the highest number of youth who prefer self-employment. Though these numbers are correlational, it is consistent with the theory of networks and self-employment.

[11] Lönnqvist et al. (2015) and Mamerow et al. (2016) also found very low correlation between self-reported risk measure and lottery-based risk measure. This is in contrast to some studies that suggest that self-reported risk measures are a good proxy for lottery-based risk measures (Dustmann et al., 2020; Hardeweg et al., 2013).

4. Empirical Strategies

We study whether a wage employment skilling program is an effective policy instrument to affect short term employment outcomes. We first compare the trainees and non-trainees who registered in the KP portal but have not received the training yet. Analysing the effect of skill training on employment is challenging due to the selection bias associated with trainees. Skill trainees are not randomly chosen. Trainees tend to be positively selected from the general population as they are more motivated to increase the probability of getting a job and change their employment situation. This is evident from Appendix Table A.2 which shows the t-statistics of differences in mean between DDU-GKY trainees and non-trainees. First, we observe in post-treatment phone survey data that, relative to non-trainees, trainees are more likely to be wage employed and less likely to be unemployed (with no statistically significant difference among self-employed individuals). Second, this correlational evidence is not necessarily causal as trainees are also significantly different in observed characteristics from non-trainee KP registrants. They tend to be younger, are less likely to be married, belong to smaller family sizes, are more likely to own land and are more likely to speak multiple languages. Moreover, trainees have bigger social networks as measured by the number of blood relatives in the district and are more risk loving.

Though we do not have experimental data to infer causality directly, we attempt to deal with the endogeneity with respect to trainee selection by using an arguably exogenous control group. The phone survey asked the registered candidate who had not yet started training if they would start training within the next six months[12]. We treat these individuals as the control group since we expect their unobserved characteristics, such as motivation, tastes and preferences, to be similar to the treatment group. While the difference in the mean of observable characteristics between the trainees and the control group may still exists, conditional on those characteristics, the two groups are assumed to be similar. Thus, using the phone survey data, we can estimate a simple probit regression to see the effect of DDU-GKY training on employment outcomes as:

$$Y_{ids} = \alpha_0 + \alpha_1 Training_{ids} + X'_{ids}\alpha_3 + D + \epsilon_{ids} \quad (1)$$

[12] There is no information on whether or not the reasoned control group received training six months later.

Where the Y_{ids} measures the probability of being wage employed in the phone survey of an individual i from district d and state s. $Training$ is a binary variable that equals 1 if the respondent has undergone DDUGKY training and 0 if the respondent did not undertake training but planned to do so within the next 6 months.

X vector includes age, age-squared, gender and own education. We control for family background characteristics including mother's education, an indicator for self-employment status of either parent, assets measured by land and smartphone ownership, minority status, whether the household is Below Poverty Line (BPL), whether any household member worked as a NREGA worker and belonged to a Self-Help Group (SHG). We control for an individual's social network and risk preferences as discussed. An individual may be more likely to be employed if anyone in the immediate family has previously migrated. Thus, all regressions control for a dummy variable that takes the value of 1 if anyone in the household has previously migrated. We further control for the respondents speaking ability in two common languages of India; Hindi and English.[13] Looking at state level variation, Odisha has the highest proportion of registrants who have completed training (36%) while Tamil Nadu has the lowest (4%). This suggests that there exists differential state policies, impetus and expenditures towards skill-training. To account for this, we include district fixed effects (D) in the regression which allows us to control for variation in state policies at an even more granular level. Standard errors are clustered at the district level.

4.1 Canonical Differences-in-Differences

The above specification is not completely exogenous as there could still exist selection bias. In particular, one could always question why some trainees choose to postpone training by 6 months. If the observable characteristics between the treatment and control group are different, there is a strong case for the unobserved characteristics to also be dissimilar. Thus, we use the KP portal dataset and the phone survey to capture the outcome and preferences in two time periods and estimate the following equation:

$$Y_{ids} = \alpha_0 + \alpha_1 Training_{ids} + \alpha_2 Post + \alpha_3 Post * Training_{ids} + X'_{ids}\alpha_4 + D + \epsilon_{ids} \quad (2)$$

[13] According to the 2011 Census data, most Indian States, apart from those located in north and central India, do not consider Hindi as the primary language. However, these states have adopted Hindi as a secondary language. Similarly, while states located in the south and northeast are not Hindi-speaking, these states have adopted English as their secondary language.

where, α_1 accounts for average difference between treatment and control group, α_2 captures the effect of post-treatment dummy and α_3 gives the true effect of treatment. Post takes the value 1 for the phone survey data (after training) and 0 for the KP portal (before training). Remaining variables are the same as in equation (1). The vector X excludes variables which were time invariant such as number of children and marital status. Causal interpretation of α_3 requires the identification assumption that there are no changes in employment trends due to other confounding factors. Given that the time period between the KP registration and phone survey was short, it is plausible to assume that any change between treatment and control groups over this period will be purely due to the training. The above equation is equivalent to a differences-in-differences specification with only two time periods.

4.3 Non-Parametric Estimates

The empirical strategy relies on strong assumptions regarding self-selection into the programme and comparability of trends. While we cannot directly test the identification assumptions, we use two separate non-parametric estimation methods to confirm the results from the canonical DID model.

Propensity Score Matching, given observable pre-treatment characteristics, approximates randomization by balancing on observables and determines an appropriate control group (Becker and Ichino, 2002). Unlike parametric techniques, the advantage of PSM is that no underlying functional form assumptions are required to estimate the relationship between outcomes and independent variables. Matching by propensity scores assumes that Conditional Independence Assumption (CIA) holds, which means potential outcomes are independent of treatment status. Using observable pre-treatment characteristics (age, gender, marriage status, education, risk attitude, social network, minority status, whether the household is BPL, whether any household member worked as NREGA worker and belonged to a SHG), we conduct an overlap test to fully determine the selection process and outcomes. The matching technique in PSM helps to ensure that the distribution of the pre-treatment characteristics of individuals in the treatment and the control groups overlap or there is common support, thereby making the groups more comparable and causal inferences more valid.

As shown in figure 1, the overlap assumption is not violated and thus we proceed to estimate the treatment effect of interest. The figure shows plots of the estimated densities of the probability of getting each treatment level. Neither plot indicates a high probability mass near 0 or 1, and the two estimated densities have most of their respective masses in regions in which they overlap each other. This assumption ensures that there is sufficient overlap in the characteristics of treated and untreated units to find adequate matches.

The drawback of PSM is that identification relies on the assumption of CIA. If individuals choose to obtain training based on the unobserved characteristics (or if these unobserved characteristics correlate with the observed characteristics that influenced/motivated them for training), then the CIA does not hold. To address this potential bias, we rely on using the Minimum Biased Estimator (MBE) approach.

The minimum bias estimator provides treatment effects estimates when the CIA fails and appropriate exclusion restrictions are unavailable. Millimet and Tchernis (2013) proposed the minimum biased (MB) estimator, which uses a propensity score-based estimator but trims the estimation sample to reduce bias caused by the CIA failure. The value of propensity score at which the bias is minimized, the bias minimizing propensity score (BMPS), is fixed for ATT, and it is 0.5[14]. They recommended a method to decide on the region, which depends on the minimum percentage of treatment and control groups present in the trimmed sample (θ). We use the MB estimator to calculate the average effect of training at the recommended value of 0.05 and 0.25 and compare the results.

5. Results

We first show the naïve probit estimates of the effect of treatment on employment. Table 3 shows marginal effects from a probit regression where the outcome variable is the probability of wage employment (columns 1 and 3) and self-employment (column 2 and 4). While columns 1 and 2 compare the trainees with non-trainees, columns 3 and 4 use the subsample of non-trainees who showed interest in completing training within the next 6 months as a control group. Those who are trained under DDU-GKY are 11

---

[14] Shown in Black and Smith (2004), Heckman and Navarro-Lozano (2004), and discussed in Millimet and Tchernis (2013). The BMPS is not fixed for ATE, and it is estimated by minimising the bias due to unobserved characteristics. Once the BMPS for ATE has been estimated, only observations with a propensity score in a neighbourhood around the BMPS are used to obtain the MB estimator for ATE. The extent of bias that must be traded off against variance necessitates a subjective decision about the support region. If the support region is too wide, the sample size will also remain large, but this will limit the extent to which the bias is reduced. If the support region is too narrow, many observations will be discarded from the analysis. The estimated ATT will be less precisely measured, resulting in larger standard errors.

percent more likely to be wage employed compared to non-trainees who are interested to undertake training within the next 6 months. Skill training under DDU-GKY has no effect on self-employment rates in the sample which is expected as the DDU-GKY training is specifically designed to address gaps in wage employment and the government of India has a separate infrastructure in place for those individuals who want to be self-employed (footnote 4).

Since the naïve estimates could be confounded by selection bias, we next use the DID strategy from equation 2 to see the causal relationship between training and wage employment. In table 4, the results suggest that after treatment, the average wage employment of the treatment group has increased to 13% compared to non-trainees and to 17 percent compared to non-trainees who are willing to undergo training in the following six months.[15]. Once again, the results from the preferred control group show no effects on the probability of being self-employed.

We next show results using alternate nonparametric empirical strategies; propensity score matching (PSM) and minimum biased estimator (MBE). Using PSM we find that the average treatment effect on treated (ATT) in our study is 12 percent that is, those who took DDU-GKY training are 12 percent more likely to be wage employed than if they have not taken the training. Table 5 shows that ATE is 10 percent and ATT is 12 percent, both significant at a 1 percent level of significance.

Next, we create a post-match balance across all covariates to see how well the matching replicated the experimental benchmark. As shown in figure 2 the covariates were not balanced in the raw data but after matching the covariates are balanced in treatment and control group. Table 6 shows balancing tests for each covariate in both unmatched (U) and the matched (M) samples. In matched samples, the difference in means between the treated group ($\bar{x1}$) and the untreated group ($\bar{x0}$) is much smaller than in unmatched samples. This is confirmed by a t-test: while the differences ($\bar{x1}$ - $\bar{x0}$) are large and often statistically significant in the unmatched samples, the differences become small and are always insignificant in the matched samples. The summary measures of the

[15] 94 percent of the individuals in our data were not employed in the first period (in KP data) while 6 percent were engaged in self-employment occupations. Thus, there is no significant variation in the outcome variable. As a robustness, we redo the same analysis with Y_{ids} as employment revealed *preferences* (as opposed to outcomes) to capture the long-term changes in the employment decision. The main results are robust to this exercise.

overall (im)balance of all conditioning variables are taken care in matched sample (refer Appendix Table A.3).

The ATE and ATT of skill training are further estimated using the MB and MB-EE estimators. 95 percent confidence intervals based on the percentile method are obtained by bootstrap using 250 repetitions. As such, following Millimet and Tchernis (2013), the preferred estimator is MB (MB-EE) for the ATE (ATT). Table 7 presents the MB estimator with $\theta=0.05$ ($\theta=0.25$) and shows that the ATE of training on the probability of being wage employed is 0.12 (0.12). The MB-EE estimator with $\theta=0.05$ ($\theta=0.25$), on the other hand, shows that the ATT of training on the probability of being employed is 0.15 (0.11). These estimates indicate that DDUGKY training causes an increase in the probability of being wage employed by 12 percentage points for the average individual and by around 11-15 percentage points for the average individual who did the training. Thus, regardless of the econometric methods used, we find a positive and significant effect of skill training on the probability of being wage employed with comparable estimates across the different specifications.

5.1. Impact of training by gender

Next, we investigate the impact of training for males and females separately on their employment status using the preferred DID strategy. The usual notion that training programs are more effective for females is not always the case, as discussed in the literature. However, the results of our study support the popular belief.

Table 8 shows that training has increased the average probability of being wage employed by 19 percent for females and approximately 17 percent for males, however, the difference of 2 percentage point between males and females is not statistically significant. Interestingly, females have a higher likelihood of engaging in self-employment post training. Evidence suggests that women are less likely to be self-employed than men (Clain, 2000). Thus, skill training is an important policy tool to improve employability of women as it increases the probability of being employed in both paid employment and self-employment.

Though we do not have the necessary data to explain this result, we can suggest some reasons by drawing on the existing literature. First, note that the Indian government also facilitates entrepreneurship training for individuals under the Rural Self-Employment Training Institutes (RSETI) programme. While the DDU-GKY training targets wage employment, RSETI is aimed at improving self-employment outcomes. Further, individuals can register for either of these programmes via the KP portal, there is no additional cost of choosing one over the other. One could thus argue that skill training helped in making women aware of the different opportunities in addition to wage employment, information that was not available to them at the time of registration. A recent study based on grassroot Indian health workers suggested that women have lack of clarity or information on what kind of business they can start. [16] Skill training may have helped in reducing the barriers to information. Similarly, training could have also made women aware that self-employment allows more flexibility in location and schedule relative to wage employment.

Second, there is strong evidence of gender differences in risky behaviour and in particular in financial decision making. Women tend to invest less, and thus appear to be more financially risk averse than men (Charness & Gneezy, 2012). Skill training may have empowered women to engage in risky alternatives such as self-employment.

Finally, we conduct sensitivity analysis to ascertain the robustness of the main results. First, we drop one state at a time from the sample. As shown in the appendix, results are robust and are not driven by any state level differences in policies. Next, we include a control for the time elapsed between KP registration and our survey. Our administrative data includes participants who registered in the portal between October 1, 2017 to 5th February 2019 while the phone survey was conducted 6 months. We control for months since registration in the DID estimates, the results do not change. We also checked if there is any heterogeneity in effects by the time elapsed since registration. Since we do not have data on when the candidate completed DDU-GKY training, we used the time since registration as a proxy for this variable. We find no significant effects. These results are available upon request.

6. Conclusion and Policy Implications

We present one of the first evaluations of Deen Dayal Upadhyay Grameen Kaushalya Yojana, a skill training program for wage employment in India. The literature on the randomized evaluations of training programs suggests marginal or insignificant effects on labour market outcomes (McKenzie, 2017; Heckman et al.,1999). However, for

[16] Available at https://iwwage.org/wp-content/uploads/2020/10/Policy-Landscape-Sudy_Summary-report.pdf

developing nations where the skill gap is one of the primary barriers to employment, skilling might improve candidate employability.

We deal with the endogeneity concerning trainee selection using two methods. First, to difference out unobserved characteristics, we use a control group which is the subset of non-trainees interested in completing the training within six months. This control group is arguably similar to the treatment group since unobservable characteristics such as the desire for work, alternative employment options, and talent would not differ much. Then, we use data from two time periods, one before and one after training, to study the causal effect of DDU-GKY on employment. Second, we also estimate training effects using two nonparametric approaches: propensity score matching (PSM) and minimum bias estimator (MBE).

We control for a host of background characteristics including risk index and social networks, two variables for which data is often missing leading to questionable treatment estimates. Our results suggest a positive and significant impact of the program on the probability of wage employment. Regardless of the methodology employed, DDUGKY training is associated with 11 to 17 percent increase in the probability of being wage-employed. We also find that skill training programs are more beneficial for women because in addition to wage employment, women also increase participation in self-employment. While we do not have the necessary data to test the mechanisms, we hypothesize that training reduces information gaps and increases female empowerment.

Women in India, particularly in rural areas, are uneducated, subject to gender discrimination in the labour market and lack the necessary skills to participate in the labour market. Thus, such policy initiatives can help them acquire specific skills and provide the opportunity to become part of the labour force.

One caveat of our study is that we measure short term employment effects. We do not know if these short term effects translate to long term employability or to an increase in wages. We leave that for future research.

References:

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., & Vitali, A. (2020). Tackling youth unemployment: Evidence from a labour market experiment in Uganda. *Econometrica*, *88*(6), 2369-2414.

Allen, W. D. (2000). Social networks and self-employment. *The Journal of socio-economics*, *29*(5), 487-501.

Attanasio, O., Kugler, A., & Meghir, C. (2011). Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *3*(3), 188-220.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *The Review of Economic Studies*, *79*(1), 128-161.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The stata journal*, *2*(4), 358-377.

Bergemann, A., & Van Den Berg, G. J. (2008). Active Labor Market Policy Effects for Women in Europe — A Survey. *Annales d'Économie et de Statistique*, *91/92*, 385–408.

Betcherman, G., Dar, A., & Olivas, K. (2004). Impacts of active labour market programs: New evidence from evaluations with particular attention to developing and transition countries.

Black, D. A., & Smith, J. A. (2004). How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching. *Journal of Econometric*, 121, 99-124.

Brown, S., Dietrich, M., Ortiz-Nuñez, A., & Taylor, K. (2011). Self-employment and attitudes towards risk: Timing and unobserved heterogeneity. *Journal of Economic Psychology*, *32*(3), 425-433.

Card D, Kluve J, Weber A (2010). Active labour market policy evaluations: a meta-analysis. *Economic Journal* 120(548):452–477

Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D., & Soares, Y. (2011). The labour market impacts of youth training in the Dominican Republic. *Journal of Labour Economics*, *29*(2), 267-300.

Charness, G & Gneezy, Uri (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization*, Volume 83, Issue 1, 2012, Pages 50-58,

Clain, S. H. (2000). Gender Differences in Full-Time Self-employment, *Journal of Economics and Business* 52 (6), 499-513.

Daly M, Delaney L (2013). The scarring effect of unemployment throughout adulthood on psychological distress at age 50: Estimates controlling for early adulthood distress and childhood psychological factors. *Social Science and Medicine* 80(0):19–23

Dustmann, Christian & Fasani, Francesco & Meng, Xin & Minale, Luigi. (2020). Risk Attitudes and Household Migration Decisions. *Journal of Human Resources*, 1019-1051, 2020.

Fougère D, Kramarz F, Pouget J (2009) Youth unemployment and crime in France. Journal of European Economic Association 7(5):909–938

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. "Evaluating government training programs for the economically disadvantaged." *Journal of economic literature* 35, no. 4 (1997): 1809-1855

Hagen, T. (2019). Evaluation of a placement coaching program for recipients of disability insurance benefits in Switzerland. *Journal of occupational rehabilitation*, *29*(1), 72-90.

Hardeweg, B., Menkhoff, L., & Waibel, H. (2013). Experimentally Validated Survey Evidence on Individual Risk Attitudes in Rural Thailand. *Economic Development and Cultural Change*, *61*(4), 859–888.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking measured risk aversion to individual characteristics. *Kyklos*, *55*(1), 3-26.

Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). The economics and econometrics of active labour market programs. In *Handbook of labour economics*, Vol. 3, pp. 1865-2097.

Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and statistics*, *86*(1), 30-57.

Hirshleifer, S., McKenzie, D., Almeida, R., & Ridao-Cano, C. (2016). The impact of vocational training for the unemployed: experimental evidence from Turkey. *The Economic Journal*, *126*(597), 2115-2146.

Kahn LB (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics* 17(2):303–316

Killingsworth, Mark and Heckman, James, (1987), Female labor supply: A survey, ch. 02, p. 103-204 in Ashenfelter, O. and Layard, R. eds., *Handbook of Labor Economics*, vol. 1, Elsevier.

Lönnqvist, J. E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *Journal of Economic Behavior & Organization*, *119*, 254-266.

Maitra, P., & Mani, S. (2017). Learning and earning: Evidence from a randomized evaluation in India. *Labour Economics*, *45*, 116-130.

Mamerow, L., Frey, R., & Mata, R. (2016). Risk taking across the life span: A comparison of self-report and behavioral measures of risk taking. *Psychology and Aging*, *31*(7), 711.

McKenzie, D. (2017). How effective are active labour market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, *32*(2), 127-154.

Millimet, D. L., & Tchernis, R. (2013). Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program. *Journal of Applied Econometrics*, *28*(6), 982-1017.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labour market. *The Quarterly Journal of Economics*, *118*(2), 549-599.

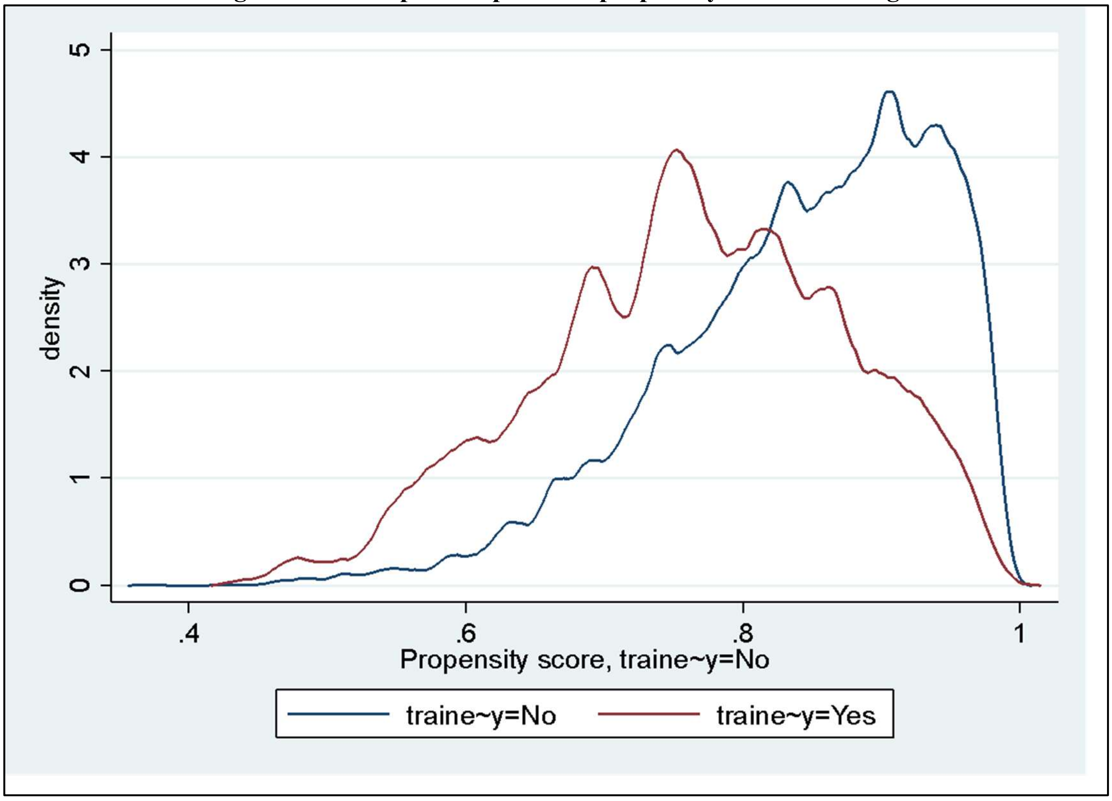
Muralidharan, K., Niehaus, P., Sukhtankar, S., & Weaver, J. (2021). Improving last-mile service delivery using phone-based monitoring. *American Economic Journal: Applied Economics*, *13*(2), 52-82.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169-188.

Schochet, P. Z., McConnell, S. M., & Burghardt, J. A. (2003). *National Job Corps Study: Findings using administrative earnings records data*. Mathematica Policy Research, Incorporated.

Tables and Figures:

Figure 1: Overlap assumption for propensity score matching



The figure shows plots of the estimated densities of the probability of getting each treatment level.

Figure 2: Balance plots for covariates

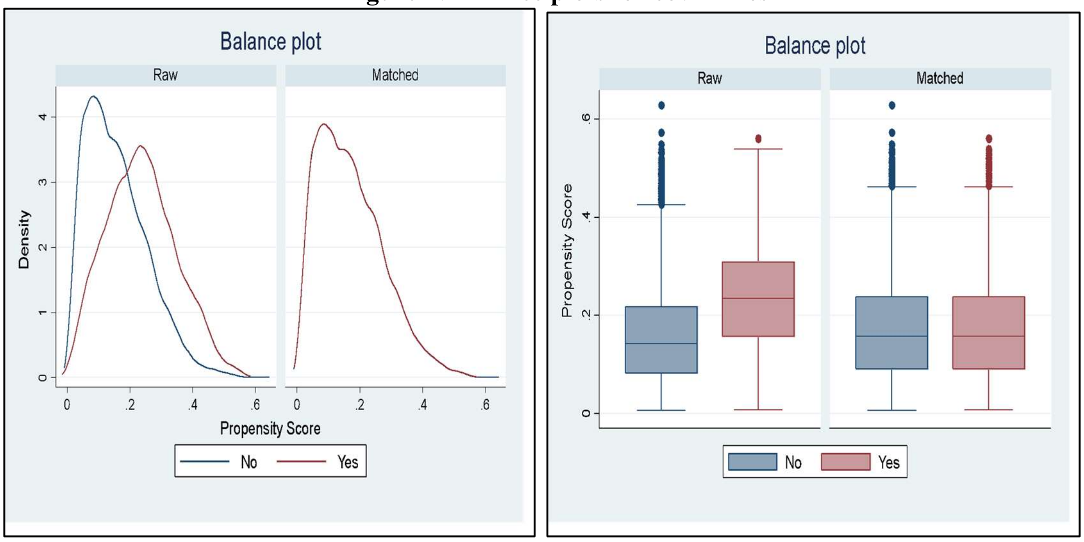


Table 1: Descriptive Statistics by State in KP Data

Variables	All States	Gujarat	MP	Odisha	TN
Willing to move to a different state for a job	0.38	0.24	0.33	0.75	0.24
	(0.48)	(0.43)	(0.47)	(0.43)	(0.43)
Willing to move to a different district for a job	0.58	0.43	0.53	0.89	0.51
	(0.49)	(0.50)	(0.50)	(0.31)	(0.50)
Prefers Wage Employment	0.32	0.14	0.14	0.63	0.42
	(0.47)	(0.35)	(0.35)	(0.48)	(0.49)
Prefers Self-Employment	0.24	0.29	0.31	0.16	0.17
	(0.43)	(0.46)	(0.46)	(0.37)	(0.38)
Percentage Currently Employed	0.06	0.02	0.03	0.01	0.15
	(0.23)	(0.13)	(0.18)	(0.11)	(0.36)
Any member of household belongs to SHG*	0.36	0.23	0.31	0.37	0.50
	(0.48)	(0.42)	(0.46)	(0.48)	(0.50)
Working in NREGA**	0.07	0.03	0.05	0.08	0.11
	(0.25)	(0.18)	(0.21)	(0.27)	(0.32)
Below Poverty Line	0.16	0.24	0.20	0.14	0.05
	(0.36)	(0.43)	(0.40)	(0.35)	(0.21)
Belong to a minority	0.07	0.07	0.03	0.04	0.14
	(0.26)	(0.26)	(0.17)	(0.19)	(0.35)
Observations	**12108**	**2511**	**3844**	**2509**	**3244**

*Self Help Group, **National Rural Employment Guarantee Act. Standard errors are in parentheses.

Table 2: Descriptive Statistics of Demographic Characteristics by State: Survey Data

Variables	All States	Gujarat	MP	Odisha	TN
Age	25.37	26.11	24.88	23.75	26.65
	(5.56)	(6.22)	(5.07)	(4.74)	(5.77)
Female	0.38	0.44	0.20	0.43	0.51
	(0.49)	(0.50)	(0.40)	(0.49)	(0.50)
Married	0.41	0.49	0.42	0.23	0.49
	(0.49)	(0.50)	(0.49)	(0.42)	(0.50)
Number of Children	0.58	0.69	0.60	0.23	0.75
	(0.98)	(1.08)	(1.03)	(0.65)	(1.00)
Household Size	4.02	4.53	4.59	3.63	3.25
	(1.84)	(1.79)	(1.96)	(1.72)	(1.41)
Mother's Education: None	0.47	0.58	0.63	0.35	0.27
	(0.50)	(0.49)	(0.48)	(0.48)	(0.44)
Mother's Education: Up to 10th grade	0.46	0.36	0.34	0.55	0.63
	(0.50)	(0.48)	(0.47)	(0.50)	(0.48)
Either Parent Self Employed	0.27	0.32	0.29	0.12	0.31
	(0.44)	(0.47)	(0.45)	(0.32)	(0.46)
Own Education: Up to 8th grade	0.08	0.09	0.09	0.03	0.09
	(0.27)	(0.28)	(0.28)	(0.18)	(0.29)
Own Education: 9th to 12th Grade	0.50	0.56	0.49	0.69	0.33
	(0.50)	(0.50)	(0.50)	(0.46)	(0.47)
Own Education: More than High School	0.42	0.35	0.43	0.28	0.58
	(0.49)	(0.48)	(0.49)	(0.45)	(0.49)
Land	0.58	0.47	0.59	0.65	0.61
	(0.49)	(0.50)	(0.49)	(0.48)	(0.49)
Smartphone	0.74	0.80	0.80	0.71	0.65
	(0.44)	(0.40)	(0.40)	(0.45)	(0.48)

Variables	All States	Gujarat	MP	Odisha	TN
Speak English	0.33	0.28	0.29	0.42	0.33
	(0.47)	(0.45)	(0.46)	(0.49)	(0.47)
Speak Hindi	0.62	0.68	1.00	0.74	0.02
	(0.49)	(0.47)	(0.04)	(0.44)	(0.14)
Preference to move state	0.20	0.12	0.22	0.32	0.16
	(0.40)	(0.32)	(0.41)	(0.47)	(0.37)
Preference to move district but not state	0.53	0.57	0.62	0.71	0.29
	(0.50)	(0.50)	(0.49)	(0.46)	(0.45)
Wage Employed	0.21	0.13	0.12	0.25	0.34
	(0.41)	(0.34)	(0.32)	(0.44)	(0.47)
Unemployed	0.39	0.45	0.41	0.36	0.35
	(0.49)	(0.50)	(0.49)	(0.48)	(0.48)
Self Employed	0.08	0.07	0.11	0.09	0.05
	(0.28)	(0.26)	(0.32)	(0.29)	(0.22)
Number of blood relatives in district	33.57	32.15	34.43	28.87	39.22
	(39.35)	(44.37)	(42.44)	(22.78)	(41.95)
Number of friends met in a week	7.64	7.14	8.43	5.78	9.06
	(11.49)	(7.38)	(8.84)	(10.68)	(18.35)
Ever Migrated	0.33	0.41	0.36	0.50	0.09
	(0.47)	(0.49)	(0.48)	(0.50)	(0.29)
Risk Score (Self-Scored)	5.98	4.33	5.38	7.88	6.21
	(2.83)	(2.28)	(2.64)	(1.82)	(3.14)
Risk Score (Lottery)	2.85	2.87	3.30	3.18	2.04
	(1.54)	(1.41)	(1.56)	(1.59)	(1.20)
Prefers Wage Employment	0.49	0.45	0.42	0.65	0.50
	(0.50)	(0.50)	(0.49)	(0.48)	(0.50)
Prefers Self-Employment	0.30	0.29	0.31	0.18	0.38
	(0.46)	(0.45)	(0.46)	(0.38)	(0.48)
Received Training	0.14	0.13	0.10	0.36	0.04
	(0.35)	(0.33)	(0.30)	(0.48)	(0.19)
Observations	**12,144**	**2,513**	**3,846**	**2,541**	**3,244**

*Standard errors are in parentheses

Table 3: Effect of Skill Training on Employment Outcomes (Marginal Probit Estimates)

	Non-Trainee		Control Group	
	(1)	(2)	(3)	(4)
Variables	**Wage Employment**	**Self Employment**	**Wage Employment**	**Self Employment**
DDUGKY training	0.087***	-0.007	0.108***	-0.012
	(0.010)	(0.008)	(0.011)	(0.010)
Demographic Controls	Yes	Yes	Yes	Yes
KP Controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Observations	**9,246**	**8,777**	**6,172**	**5,677**

Table shows marginal effects from a probit regression where the outcome variable is the probability of wage employment (columns 1 and 3) and self-employment (column 2 and 4). While columns 1 and 2 use all non-trainees, columns 3 and 4 use the subsample of respondents who showed interest in completing training within the next 6 months. All regressions include district fixed effects. Controls include age, age square, gender, marital status, number of children, family size, mother's education, own education, asset ownership, knowledge of Hindi and English, migration status of family members, both measures of risk and social networks. Also includes KP data variables NREGA, SHG, BPL and Minority. Standard errors are clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effect of Skill Training on Employment Outcomes (canonical DID)

	Control Group 1		Control Group 2	
Variables	**(1) Wage-Employment**	**(2) Self-Employment**	**(3) Wage-Employment**	**(4) Self-Employment**
Post*Treatment	0.128***	0.016*		
	(0.022)	(0.015)		
Treatment	-0.019**	-0.014*		
	(0.008)	(0.008)		
Post*Treatment			0.169***	0.011
			(0.023)	(0.015)
Treatment			-0.026***	-0.015**
			(0.009)	(0.007)
Post	0.191***	0.038***	0.149***	0.042***
	(0.015)	(0.015)	(0.016)	(0.014)
Constant	0.044	-0.235***	0.141***	-0.166***
	(0.043)	(0.044)	(0.053)	(0.049)
Demographic controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Observations	18,719	18,719	12,666	12,666

Table shows difference in difference estimates of wage employment (columns 1 and 3) and self-employment (column 2 and 4) after training. While columns 1 and 2 use as control group all non-trainees, columns 3 and 4 use as control the subsample of respondents who showed interest in completing training within the next 6 months. All regressions include district fixed effects. Controls include age, age square, gender, family size, mother's education, own education, asset ownership, knowledge of Hindi and English, migration status of family members, both measures of risk and social networks. Also includes KP status on NREGA, SHG, BPL and Minority. Standard errors are clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of Skill Training on Employment Outcomes (Propensity score matching)

Treatment effects	**Wage employed**
ATE	0.102***
	(0.0181)
ATT	0.122***
	(0.0166)
Total	**9613**

The table gives the average treatment effect (ATE) of training and average treatment effect on treated (ATT) after the propensity score matching using phone survey data. Out of 12144 individuals, we find the adequate matches for 9613 individuals (1649 individuals completed the training and 7964 individuals were non-trainees). Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Balancing of Covariates

Variable	Unmatched/Matched	Mean Treated	Mean Control	% bias	% reduction	t	p>|t|
Age	U	23.72	25.42	-32.5	88.3	-11.6	0
	M	23.72	23.52	3.8		1.21	0.23
Female	U	0.40	0.35	10.4	83.1	3.88	0
	M	0.40	0.39	1.8		0.5	0.62
Married	U	0.22	0.43	-44.8	93.7	-15.44	0
	M	0.22	0.21	2.8		0.88	0.38
9th to 12th Grade	U	0.59	0.51	15.7	84.5	5.79	0
	M	0.59	0.58	2.4		0.71	0.48
High School and above	U	0.38	0.41	-6.7	42.7	-2.47	0.01
	M	0.38	0.40	-3.8		-1.11	0.27
Self-Scored Risk	U	6.99	5.76	46.7	97.7	16.29	0
	M	6.99	6.97	1.1		0.32	0.75
blood relatives	U	40.93	34.52	15.7	91.1	5.81	0
	M	40.93	41.50	2.4		-0.37	0.71
Risk Lottery Score	U	3.35	3.17	12	48.1	4.45	0
	M	3.35	3.44	-6.2		-1.71	0.08
Any member of household belongs to SHG	U	0.26	0.36	-21.5	92.7	-7.74	0
	M	0.26	0.27	-1.6		-0.47	0.64
Working in NREGA	U	0.08	0.06	7.3	93.6	2.8	0
	M	0.08	0.08	0.5		0.13	0.90
Below Poverty Line	U	0.23	0.15	22.7	95.2	8.92	0
	M	0.23	0.23	1.1		0.29	0.77
Belong to a minority	U	0.05	0.07	-5.8	95.5	-2.06	0.04
	M	0.05	0.05	-0.3		-0.08	0.94

The table shows the balancing tests with respect to each covariate used in PSM, for both unmatched (U) and matched (M) samples.

Table 7: Effect of Skill Training on Employment Outcomes (MB and MB-EE estimates)

	Control and PSM variables same		*Controls and PSM variables are different*	
	Wage employment (ATE)	Wage employment (ATT)	Wage employment (ATE)	Wage employment (ATT)
MB (θ=0.05)	0.122	0.153	0.126	0.153
	[0.032, 0.217]	[0.092, 0.207]	[0.039, 0.206]	[0.100, 0.217]
MB (θ=0.25)	0.113	0.11	0.12	0.112
	[0.066, 0.150]	[0.078, 0.142]	[0.078, 0.156]	[0.083, 0.145]
MB-EE (θ=0.05)	0.137	0.152	0.151	0.157
	[0.053, 0.206]	[0.088, 0.204]	[0.084, 0.227]	[0.100, 0.215]
MB-EE (θ=0.25)	0.098	0.106	0.102	0.112
	[0.072, 0.150]	[0.078, 0.142]	[0.078, 0.153]	[0.082, 0.147]

The outcome variable is wage employment. The treatment variable is skill training. Controls include age, gender, marital status, number of children, family size, mother's education, own education, asset ownership, knowledge of Hindi and English, migration status of family members, both measures of risk and social networks. Also Includes KP status on NREGA, SHG, BPL and Minority. For Propensity score matching we used a subset of all controls and focused on variables that determine selection into DDUGKY training. This includes age, gender, marital status, education, risk and social networks, knowledge of Hindi and English, KP status on NREGA, SHG, BPL and Minority. Other demographics are not used since they could be potentially endogenous. MB (MB-EE), minimum-biased (minimum-biased Edgeworth expansion) estimator using θ = 0.05 or 0.25. 95% empirical confidence intervals in brackets are obtained using 250 bootstrap repetitions.

Table 8: Skill Training effect on Employment Outcomes separately for male and female (canonical DID)

Variables	**(1) Wage-Employment**	**(2) Wage-Employment**	**(3) Self-Employment**	**(4) Self-Employment**
	Female	**Male**	**Female**	**Male**
Post*Treatment	0.189***	0.167***	0.044***	-0.003
	(0.029)	(0.026)	(0.015)	(0.019)
Post	0.085***	0.181***	0.001	0.063***
	(0.012)	(0.020)	(0.013)	(0.016)
Treatment	-0.030**	-0.035***	-0.010	-0.018*
	(0.012)	(0.009)	(0.009)	(0.009)
Constant	-0.123	-0.059	-0.115*	-0.180***
	(0.085)	(0.074)	(0.062)	(0.069)
Demographic controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Observations	4,428	8,238	4,428	8,238

Table shows difference in difference estimates of wage employment (columns 1 and 2) and self-employment (column 3 and 4) after training for female and male separately. All regressions use as control the subsample of respondents who showed interest in completing training within the next 6 months. All regressions include district fixed effects. Controls include age, age square, family size, mother's education, own education, asset ownership, knowledge of Hindi and English, migration status of family members, both measures of risk and social networks. Also includes KP status on NREGA, SHG, BPL and Minority. Standard errors are clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix

Table A.1: Kaushal Panjee and Survey Data by State

	KP	Viable*	Attempted	Surveyed	Identified
Gujarat	71,408	14,683	9,823	2,513	2,028
Madhya Pradesh	75,976	42,619	19,480	3,846	3,385
Odisha	1,83,537	69,289	14,007	2,541	2,313
Tamil Nadu	3,15,758	1,58,648	13,951	3,244	2,426
Total	**6,46,679**	**2,85,239**	**57,261**	**12,144**	**10,151**

*Viable number (validation criteria): Numbers that are 10 digits in length, begin with 6, 7, 8 or 9, and have not more than 10 people associated with the same phone number.

Table A.2: Differences in Means by DDU-GKY Training Status

Variables	Mean(Not Trained)	Mean(Trained)	t-statistics	Std. Error	Obs.
Wage Employed	0.19	0.31	-0.13***	0.01	12092
Self Employed	0.08	0.08	0.00	0.01	12092
Unemployed	0.41	0.30	0.11***	0.01	12144
Age	25.64	23.77	1.87***	0.14	12143
Female	0.38	0.40	-0.03**	0.01	12144
Married	0.44	0.23	0.21***	0.01	12144
Number of Children	0.63	0.29	0.34***	0.03	12143
Household Size	4.05	3.86	0.19***	0.05	12130
High School and above	0.43	0.38	0.04***	0.01	12144
Land	0.57	0.65	-0.08***	0.01	12069
Speaks English	0.31	0.41	-0.09***	0.01	12144
Speaks Hindi	0.59	0.79	-0.21***	0.01	12144
Number of blood relatives	32.23	40.86	-8.63***	1.04	10852
Friends met in a week	7.61	7.82	-0.21	0.30	10828
Migrated	0.30	0.53	-0.23***	0.01	12144
Self-Scored Risk	5.79	6.99	-1.20***	0.07	10515
Risk Lottery Score	2.78	3.29	-0.52***	0.04	12144

Our sample consist of 12,144 individuals, only 14 per cent completed the DDUGKY training

Table A.3: Overall (im)balance of covariates used in PSM

Sample	Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R
Unmatched	0.085	746.91	0	20.1	15.7	76.5*	0.83
Matched	0.002	7.52	0.821	2.2	1.7	9.6	1.03

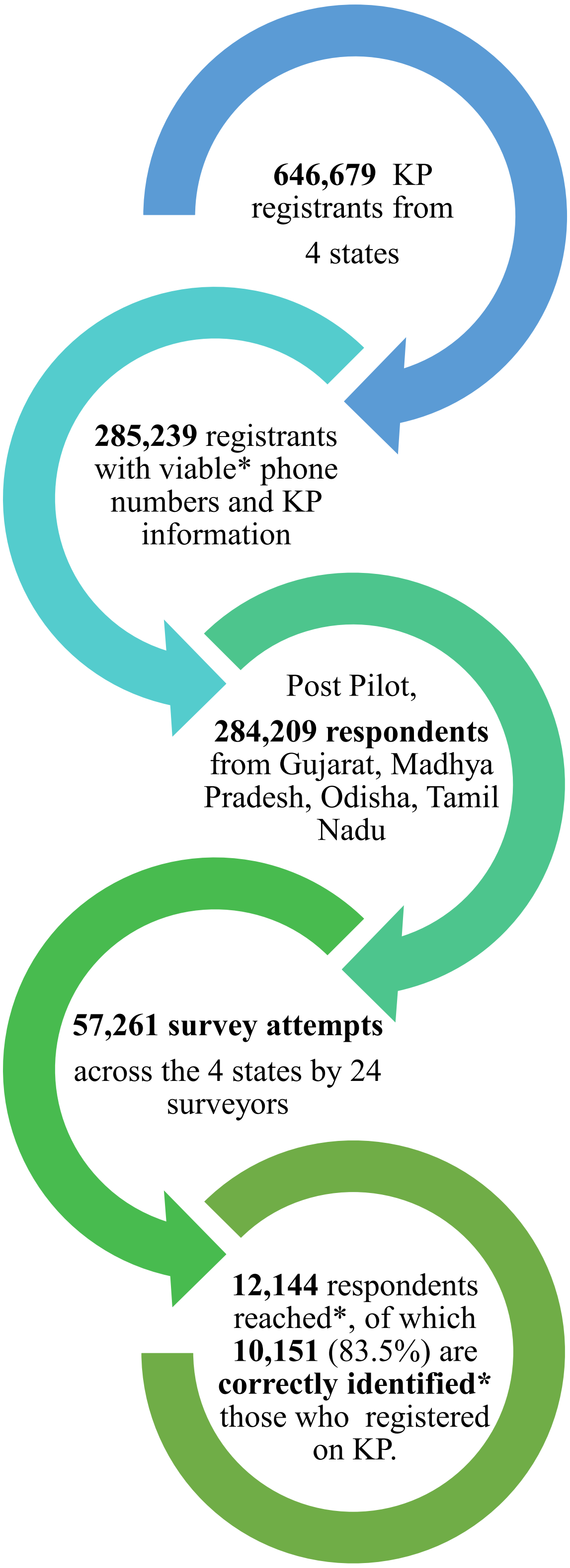
For overall balance of covariates, the value of B should be less than 25 and R should lie between 0.5 to 2. Both the conditions are satisfied for matched sample. The Pseudo-R^2 is from a probit estimate of the propensity score equation in the unmatched and the matched samples. The fact that the Pseudo-R2 is near zero in the matched samples indicates that after matching, the conditioning variables no longer have any predictive power for the participation. This is a further indication that differences between treated and control individuals are balanced (Hagen, 2019). Due to the matching procedure, the mean standardized difference is reduced from 20.1 to 2. Rubin's B is the absolute standardized difference in the means of the linear index of the propensity score in the treated and the (matched) untreated group. Rubin (2001) specifies that a B below 25 indicates a balanced control group. For the matched sample, it is 9.6. Rubin's R is the ratio of treated to (matched) untreated variances of the propensity score index. Rubin's R should be between 0.5 and 2. This is again the case in the matched sample.

Table A.4: Robustness Check – dropping one state at a time

	Control group			
	(1)	(2)	(3)	(4)
VARIABLES	Wage-Employment	Wage-Employment	Wage-Employment	Wage-Employment
Post*Treatment	0.149***	0.148***	0.147***	0.211***
	(0.027)	(0.031)	(0.024)	(0.020)
Treatment	-0.026**	-0.011	-0.010	-0.049***
	(0.011)	(0.011)	(0.008)	(0.007)
Post	0.172***	0.192***	0.144***	0.103***
	(0.019)	(0.024)	(0.018)	(0.007)
Demographic controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Observations	9,759	8,007	9,419	10,855

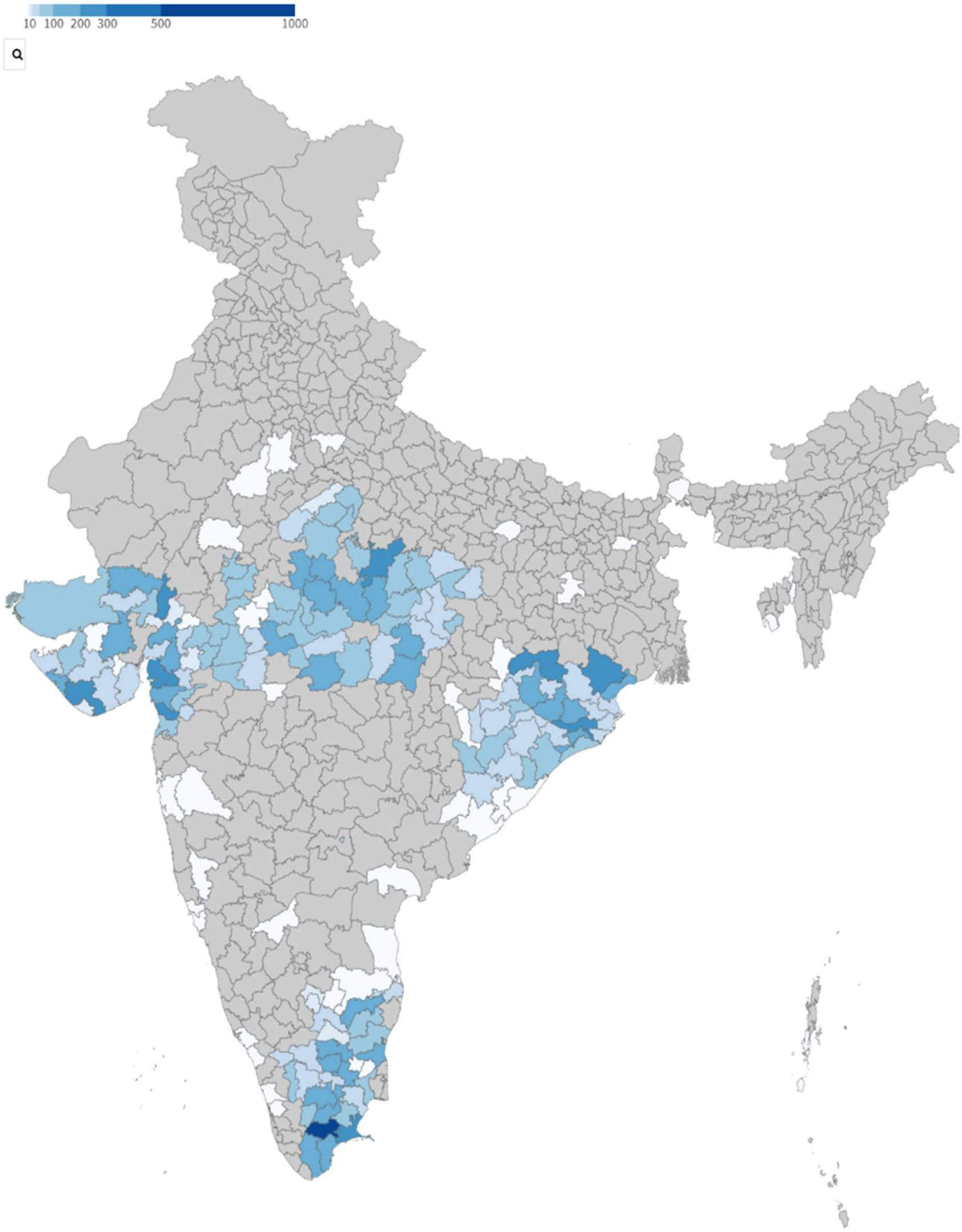
The table shows the result of pre-post analysis after dropping one state at a time. All regression includes the control group, i.e., non-trainees who will complete the training within six months. Column 1 excludes observations of Gujarat, Column 2 excludes observations of Madhya Pradesh, Column 3 excludes observations of Odisha, and Column 4 excludes observations of Tamil Nadu. We found that the treatment effect after dropping one state at a time is also positive and significant. Standard errors are clustered at district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure A.1: Spatial Distribution of KP Registrants Surveyed in Follow-Up.



- *Viable numbers are those that meet the criteria of being 10-digit in length, starting with numbers 6, 7, 8 and 9 and having 10 or fewer people associated with it.
- To test the survey instrument, we conducted a pilot exercise. The 1,030 KP registrants whose names and information were a part of the pilot exercise were excluded from the final survey exercise.
- The 284,209 respondents and their information was divided among 24 Surveyors. This division was based on the State the respondents belonged to and surveyors' fluency in the language of that State. Over a data collection period of 45 days, 57,261 phone surveys were attempted.
- Of the total surveys attempted, ~21 percent were reached.
- *Reached individuals were those who consented to being surveyed.
- A respondent is *correctly identified if their name and identity matches with that of the individual associated with the phone number used to reach them.
- Attrition was on account of numbers not being reachable, respondents not giving consent to be surveyed, or respondents' inability to understand surveyors due to audibility issues and language incompatability.

Figure A.2: Spatial Distribution of KP Registrants Surveyed in Follow-Up (Map)



Data collection activities spanned over a period of 45 days in the months of August and September 2019. For the data collection exercise, we set up field offices in three cities; Gandhinagar in Gujarat, Bhubaneswar in Odisha and Madurai in Tamil Nadu. In these locations, 24 surveyors were hired and trained to administer the follow up survey in the language of the KP registrant. Surveyors in Gandhinagar and Bhubaneswar conducted phone surveys for individuals from the states of Gujarat and Odisha respectively. Those surveyors who were fluent in Hindi also conducted phone surveys for the state of Madhya Pradesh. Surveyors in Madurai were fluent in Tamil and conducted phone surveys for KP registrants from the state of Tamil Nadu