The effect of unshrouding relevant information on demand for financial products: Evidence from India

Pre-analysis Plan

Vimal Balasubramaniam Aditi Dimri Renuka Sane[1]

June 3, 2019

[1]Vimal Balasubramaniam and Aditi Dimri are at the University of Warwick. Renuka Sane is at the National Institute of Public Finance and Policy, Delhi. Corresponding address: renukas [at] gmail [dot] com

1 Experimental design and Sample

The experiment has the following key stages: baseline survey, randomization, information intervention, endline survey, follow-up phone survey.

The sampling strategy is random with each surveyor (for baseline) following the right-hand rule in a location. The surveyor will knock on every fifth household on his path and enlist the household meeting the eligibility criterion and with appropriate consent. The experiment will have a sample size of 3201 households. These will be equally divided between the three intervention arms.

The intervention for all participants includes two components: (1) an information video, (2) an information sheet. The content of these varies by intervention group.

The three intervention arms are the following:
(1) C - The control group information video and sheet will contain information on the purposes of purchasing insurance, and will be informed to pay more attention to the standard disclosure made when products are sold.

(2) T1 - In addition to the control group information, the first treatment group information video and sheet will contain information on specific questions that should be asked before making an insurance purchase decision.

(3) T2 - In addition to the first treatment group information, the second treatment group information video and sheet will contain information on an alternative product (combination of term insurance and the public provident fund) that can achieve better insurance cover and return on investment.

Finally, all the groups will be offered the same hypothetical endowment insurance product as an product information sheet. This hypothetical product is called "Jeevan Mitr".

2 Econometric Approach

2.1 Specifications

The empirical strategy will be intention-to-treat (ITT), that is, all households are analysed with the assumption that they remained in the intervention group to which they were initially assigned. The impact of the two treatments can be evaluated by comparing outcomes across groups in a simple regression framework. For each household-level outcome, the main specification is given by:

$$y_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \gamma X_i + \sum_s \delta_s I(S = s) + \epsilon_i \tag{1}$$

where y_i denotes the outcome for household i, t_{1i} is a dummy variable equal to 1 for households in the first treatment group; t_{2i} is a dummy variable equal to 1 for households in the second treatment group; with the reference group as the control group. X_i represents the vector of household level controls and ϵ_i is a robust error term. The randomisation will be stratified on some variables (see Section 3), to adjust our standard errors for stratification we add a dummy variable for each strata with δ_s denoting the randomization stratum fixed-effect.

To address any potential variation in surveyor ability or enthusiasm, we will have an additional specification controlling for surveyor fixed effects:

$$y_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \gamma X_i + \sum_s \delta_s I(S = s) + \sum_t \mu_t I(T = t) + \epsilon_i \quad (2)$$

where t denotes the surveyors conducting the surveys.

We will also directly test the differential impact of the two treatments with the following specification:

$$y_i = \alpha + \beta_3 t_{2i} + \gamma X_i + \sum_s \delta_s I(S = s) + \epsilon_i \quad (3)$$

$$y_i = \alpha + \beta_3 t_{2i} + \gamma X_i + \sum_s \delta_s I(S = s) + \sum_t \mu_t I(T = t) + \epsilon_i \quad (4)$$

where the sample is restricted to the two treated groups making the reference group the first treatment, T1, group.

2.2 Estimators

For continuous variables the OLS estimator will be used. For binary outcomes both OLS and logit estimators will be used. Categorical variables will be analysed in two ways - (1) using the ordered logit estimator; (2) creating a binary outcome (yes and not yes).

2.3 Control variables

In all our specifications, X_i represents household level controls that are potentially strong explanators of the outcome but are not influenced by the intervention. These will all be measured at baseline -

- Age
- Gender
- Education

- Occupation
- Martial Status
- Number of dependents/children
- Number of earning members
- Mother tongue
- Number of years living in Delhi
- Income
- Assets
- Financial investments
- Insurance ownership
- Personal financial stability
- Monthly loan
- Financial Literacy Score
- Hosusehold financial decision making
- Baseline value of outcome, if applicable
- Compliance to intervention variables (listed in Section 4.2)

3 Stratification and Heterogeneity

Of interest in our study is whether the intention to treat effects vary by the *ex-ante* levels of general financial literacy, and specific literacy with insurance products. This is important, in particular, because our information intervention may have differential impact on individuals depending on their ability to contextualise the information (general financial literacy), and the "new"-ness of the information that is provided to them. Additionally, traditional variables such as age (or age-groups) and income (or income-groups) also are of interest since broad-stroke regulatory intervention often restrict the sale of retail financial products along these dimensions. Additionally, insurance premiums payable by individuals will vary by age as it takes into account the conditional survival probability of the individual.

We will use the following specification to study the heterogeneous effects:

$$y_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \beta_{1w} t_{1i} w_i + \beta_{2w} t_{2i} w_i + \beta_w w_i + \gamma X_i + \sum_s \delta_s I(S = s) + \epsilon_i \quad (5)$$

where w_i denotes the variable along which we want to test heterogeneous impact.

We will stratify the randomisation on the baseline values of the following variables and test for differential impact:

1. Age
2. Geographic zone
3. Life insurance ownership
4. Stated preference for savings or life insurance
5. Index of household types
   This index is created by principal component analysis (PCA) of the variables on financial literacy score, risk preference score, time preference score, annual income, self-reported measure of financial stability, income type (self-employed or salaried), gender, education, number of dependents, and number of earning members.

Apart from these five key variables we expect heterogeneous effect along, we will use a LASSO procedure to help predict other dimensions of heterogeneity that have high explanatory power [1]. We will use the following variables in the procedure: the above five variables, all the variables listed in the controls section, and a variable on measuring the importance of insurance features (cover vs return). Finally we will test for heterogeneous effects by compliance to intervention variables (listed in Section 4.2).

4 Outcomes Variables

All outcome variables are measured at the household level and the respondent is always the individual participating in the experiment. We measure outcomes in three ways: (1) Endline: the intervention is immediately followed by and endline survey, in which we measure stated take-up of the product and elicit detailed responses to disentangle potential mechanisms of impact. (2) Follow-up phone call: We make a phone call to all households, few days after the

[1] Esther Duflo, NBER Summer Institute 2018 presentation "Machinistas meet randomistas: useful ML tools for empirical researchers"

intervention, and ask if they are interested in a term insurance product. (3) Follow-up agent call: If participants consent to being contacted by an insurance agent, they receive a call from an authorised agent offering more information, or option to initiate the actual purchase of a term insurance product, thus enabling us to measure the action taken by the respondent to the intervention as well.

We refer to (1) as the "end line" instrument. The last two are classified as 'follow-up" instrument in Table **??**, which lists the various measures of outcome that will be covered in the study. Column (1) lists the question or measure used, Column (2) the nature of the variable obtained, and Column (3) lists the instrument by which this information is obtained.

4.1 Main Outcomes: Insurance Take-up

These outcomes measure the final impact of the intervention on various forms of take-up of the two insurance product.

We hypothesize that the two interventions will reduce the take-up (stated and revealed) of the endowment insurance product and increase the take-up of the term insurance product. This implies that in our specifications, for endowment take-up outcomes we test for: $\beta_1 < 0$, $\beta_2 < 0$, and $\beta_3 < 0$. While for endowment take-up outcomes we test for: $\beta_i > 0$.

Table 1: Insurance take-up

Question	Variable type	Instrument
Panel A: Primary outcomes		
If you were in a situation where you could choose to buy only one of the following products (endowment and term), what would you pick?	categorical	endline
Having been introduced to Jeevan Mitr, would you be interested in purchasing this product?	categorical - yes, no, cannot say	endline
Would you be interested in purchasing this term product?	categorical - yes, no, cannot say	follow-up
Panel B: Secondary outcomes		
May I give your contact details to an insurance agent to follow up and provide you with more relevant details?	binary -yes, no	follow-up
Spoke to insurance agent	binary -yes, no	follow-up
Is HH interested in buying the term product?	binary -yes, no	follow-up
Did the HH purchase a term product?	binary -yes, no	follow-up
Did the HH buy another product? (endowment)	binary -yes, no	follow-up
Having been introduced to Jeevan Mitr, what would your advice be to a relative who is in a similar financial situation as you?	categorical - yes, no, cannot say	endline

4.2 Intermediate Outcomes and Mechanisms

Question	Variable type	Instrument
General Insurance Knowledge		
If your income is Rs.3,00,000 per annum, then what would be the minimum amount of insurance you would need for your family?	binary	endline
If inflation is 4%, and an insurance product gives you 6%, what would be the rate of return after deducting inflation?	binary	endline
Score of above two questions	continuous	endline
If/when you were to buy a life insurance policy, what product features would you look out for?	continuous	endline
Product feature dummies formed using above question - What proportion declare feature x at endline relative to baseline?	binary	endline, baseline
Unshrouding of product features		
Score of below five questions	continuous	endline
What is the guaranteed rate of return of the Jeevan Mitr product?	binary	endline
What do you think is the overall rate of return (guaranteed and non-guaranteed) on the Jeevan Mitr product?	binary	endline
Given the inflation rate of 5%, what will be the guaranteed rate of return, after deducting inflation from it?	binary	endline
What is the cover provided by the product?	binary	endline
Do you always get back all the money you have put in irrespective of when you surrender?	binary	endline

Information seeking		
Would you seek information from a professional financial planner before deciding whether to purchase this product?	categorical - yes, no, cannot say	endline
Imagine you are considering purchasing an insurance product that required you to pay a premium of Rs.10,000 every year for 10 years. How much would pay a professional to help take a decision about the purchase of the product?	continuous	endline
Action: Opened information video link	binary - yes, no	follow-up
Decision making		
Having seen the video about the Jeevan Mitr product, what do you think about the product?	categorical - good, bad, neither, cannot say	endline
Product feature dummies formed using the question - How important are the following features in determining that Jeevan Mitr is good/bad product?	binary	endline
Why do you not want to share this information?	categorical	follow-up
Compliance to intervention		
Score of below questions	continuous	endline
Did the respondent watch the video till the end?	binary	endline
How many times did the respondent watch the video?	continuous	endline
Did the respondent read the information sheet?	binary	endline
Did the respondent read the product sheet? (product sheet)	binary	endline
Where did you infer this number from?	categorical	endline

How attentive was the respondent for the entire duration of the HH visit?	continuous	endline
Score of questions on: sounding impatient, irate, backgroud noise	continuous	follow-up
Asking questions on the product		
Do you have any questions about, Jeevan Mitr, the product presented in the video ?	binary	endline
Number of questions asked on the product	continuous	endline

5 Other analyses

5.1 Randomisation balance

To check the validity of the randomisation and balance across the three groups, we perform balance tests on all the variables listed in Section 2.3 as control variables.

5.2 Multiple Hypothesis Testing Correction

To minimise multiple hypothesis testing, multiple measures of the same hypothesis (like knowledge, unshrouding) will be combined to form one score or index. When not possible, the statistical tests will report the corrected $p-$values to account for multiple hypothesis testing within the same type of outcome variables.

5.3 Selective Attrition

Typically, the concern with attrition is to do with the disappearance or non-response of individuals in the study, in particular, if the attrition is differential across the intervention groups then the sample suffers from selective attrition bias. In our study, we do not foresee a high rate of attrition as the intervention and endline happens in quick succession to the baseline, with little risk of households moving residences. If any, attrition could happen for a households

that took part in the baseline but has a non-reponse to the endline or follow-up survey.

Test: We will test for differential attrition by regressing whether the participant responded to the endline on the treatment assignment at randomisation:

$$response_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \epsilon_i \tag{6}$$

where $response_i$ is a dummy that is 0 if the household did not respond to the endline survey and 1 otherwise. The same regression will be run with $response_i$ measuring response to the follow-up survey to test for differential attrition at the follow-up survey stage.

In addition to this, we will also test if the individuals leaving the study are different in terms of any of the control variables,

$$response_i = \alpha + \gamma X_i + \epsilon_i \tag{7}$$

where X_i is the vector of controls variables discussed in Section 2.

In the case of differential attrition, we will address the arising bias in two ways. First, we will use the re-weighting procedure as described in DiNardo et. al. (1996)[2]. Second, we will attempt to use imputations and bounds methodologies as proposed in King et al. (2001)[3] and Lee (2009) respectively.[4]

5.4 Selective Compliance

All the three groups are exposed to similar interventions via an information video and an information sheet. Only the content of these varies by group. The intention to treat (ITT) strategy assumes that each participant watches the video and reads the sheet originally assigned to it. While the surveyors are instructed to ensure playing the video atleast once and giving the participant at least 2-3 minutes to read the sheet, it is possible that participants do not comply with the intervention. This can be a problem if, like attrition, it is differential by treatment group.

Test: To test for differential compliance by treatment, we regress a compliance index on the treatment assignment at randomisation:

$$compliance_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \epsilon_i \tag{8}$$

[2]DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. No. w5093. National bureau of economic research, 1995.

[3]King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incom- plete political science data: An alternative algorithm for multiple imputation.

[4]Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies, 76(3):1071- 1102.

where $compliance_i$ is an index created using the variables listed under the compliance block in Table 2 of Section 4.2.