Briefing Paper 1 July 2015

Why Does Sampling Matter? Answers From the Gender Norms and Labour Supply Project

In this briefing paper you will find out how to:

- understand and explain different sampling designs
- design surveys using multi-stage sampling techniques
- contrast randomness and representativeness of a sample
- address problems raised by missing data
- determine the adequacy of a survey design (power analysis)

Introduction: This paper forms part of a project which investigates the gender aspect of the impact of poverty alleviation schemes in rural India and Bangladesh. Combining a variety of sources, we aim to offer a fresh view on the effect of anti-poverty interventions by focusing on how women's involvement in the labour market is mediated via local gender norms. To interpret our findings we draw on an innovative combination of approaches from a number of disciplinary backgrounds including sociology, economics and social policy.

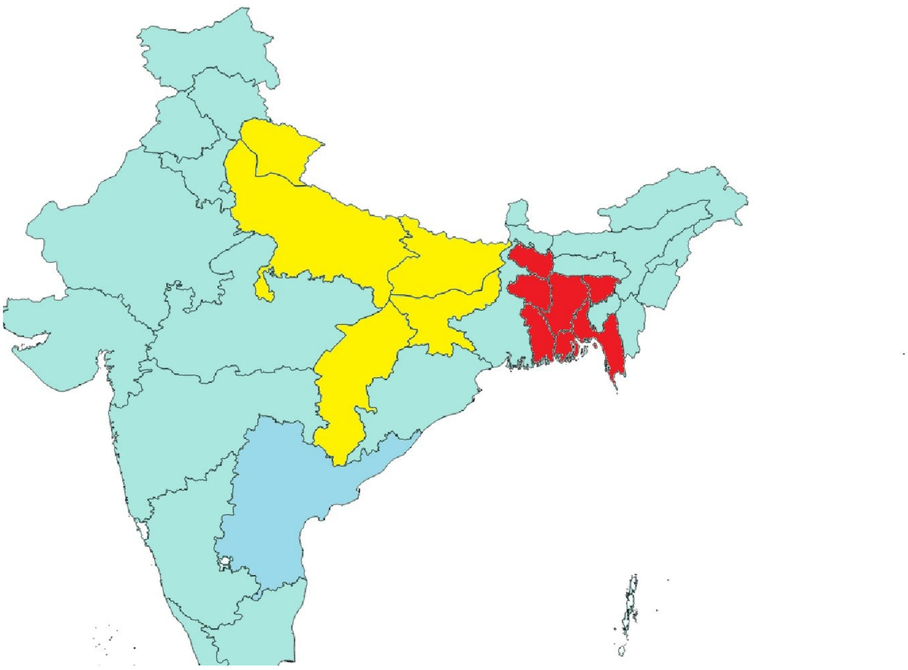


Figure 1. Map Highlighting the Regions Investigated

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project

What Counts as Random Sampling? "Probability sampling" is any sampling method where each element in the sampling frame (see Box 2) has an equal chance of being sampled, or at least the probability of a case being sampled is set in advance. Probability sampling is useful when we wish to make statistical inferences, about a larger population, based on a (cheap) sample.

Box 1: Probability Sampling

Saunders, et al (2012) suggest a four step procedure:

1. Identity a suitable sampling frame based on your research questions and objectives.
2. Determine a suitable sample size.
3. Select the sample using the most appropriate sampling technique (see Figure 2).
4. Check that the sample is indeed representative of the population.

Figure 2. Three-Stage Sampling (Saunders, et al., 2012: Ch. 7)

Stage 1	Stage 2	Stage 3
• Cluster Sampling • Usually Geographic Areas	• Random Stratified Sampling • May Need to Create a Census List of Households	• Intra-household Sampling • Can Use Kish Grid Approach • Or Map 1:1 Sample to the Household, e.g. Choose the Eldest Dependent Child

In a three-stage sampling procedure, statisticians would argue that if any stage is non-random then the whole procedure is non-random. Figure 1 shows that the arbitrary location of the clusters introduces a non-random element. Every cluster in the whole geographic area does **not have an equal chance of getting into the sample.**

On the other hand, the purpose of randomness is to gain representativeness of a population. In multi-stage sampling with limited resources, we can't try to represent the whole geographic area shown in Figure 1. Instead we can try to represent '**Villages like those in the chosen clusters**.' Suddenly random selection at Stage 2 is very important.

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project.

The Stage 3 sampling options are also important. One option is 'Kish Sampling' where individual adult are chosen randomly within each chosen household (Kish, 1995). This is typically used in large surveys. The Kish grid allows us to choose k adults from each house with household size varying from 1 to (say) 13 adults. The odds of each k are set in the grid to optimize the speed at which the number of chosen individuals reaches the desired target number. Kish sampling is a random option.

The second option is non-random selection of one person or one couple per household. In our project we are choosing one adult married woman, of any age, preferring the youngest if there are more than one couples in that household. We then also choose her husband if he is available. When using this 1:1 mapping of households to the 'EGO' (first) respondent, the randomness of Stage 2 sampling is transferred to the Stage 3 sampling. However, the **population now omits those couples, children and absent men who are not included by this intra-household sampling method.**

Omitting these changes the population to which inferences would refer. We infer from the sample to **this population.**

Staff at the BRAC Development Institute, Bangladesh, choose Upazilas as the Stage 2 sampling units. Staff at the Benares Hindu University, Varanasi, India, choose to select census villages as their Stage 2 sampling units across three states. The inferences would typically be 'from this sample data to villages and Upazilas of the kinds chosen as the sample Clusters'. The logic is based on **representativeness of what is represented**. This has internal validity as well as external validity as long as the claims are worded carefully.

Box 2: Sampling Frame.

A sampling frame is **the set of source material from which the sample is selected**. It is common to use a number of frames, especially in multi-stage designs.

For example in the type of design outlined in Figure 2 we would have a different sampling frame for each stage. In Stage 1 our sampling frame would be geographic areas of interest such as villages, in Stage two our sampling frame would be specific households within these villages and in Stage 3 it would be specific individuals within these households.

Example of Inferences Depending on Sample Size

Once the sample is chosen, the inference from sample to population **depends on the absolute sample size,** not on the proportion of each village or state that responds. The inferences are weakened by "Non-Response". The non-response is likely to be non-random. In the face of Stage 1 non-random selection and Stage 2/3 non-response, openly admitting all weaknesses of the sampling is recommended.

Contrasting Randomness with Representativeness of a Sample

We use secondary data from the 2007 DHS the 2011 DHS and the 2005/2006 NFHS. All surveys had a high response rate; the household response rates for DHS 2007 and DHS 2011 (Bangladesh) were 99.4% and 97.9% respectively. These are representative samples.

For the NFHS 2005/2006 (India) the household response rate was 97.5% for the male sample and 97.7% for the female sample. For both sets women were more likely to respond than men.

If a sample has non-response, design weights are used to make the results ***representative.***

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project.

For all DHS datasets, the response rate varied between rural and urban areas with households in rural areas being slightly ***more*** likely to respond. Non-respondents were typically absent from their household during repeated visits.

It is common for researchers to re-attempt contact with non-respondents, for example re-visiting a household in which the respondent was absent at the time of the first visit. However this does not necessarily increase the representativeness of the sample. As Stoop *et al* show using examples, sometimes people who do not respond **have characteristics which differ significantly from the rest of the sampled population** (Stoop, *et al.,* 2010. p.189)**.**

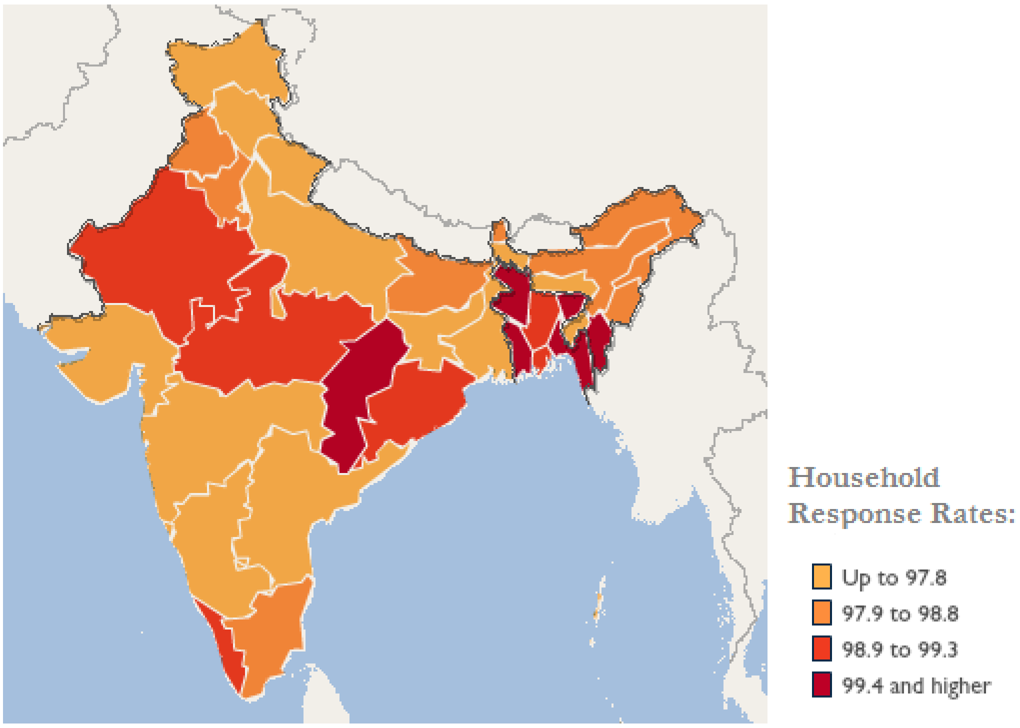


Figure 3. Household Response Rates from the India DHS 2005/6 and Bangladesh DHS 2007

The traditional inference from sample to population is then framed as follows: 'The sample data allow us to conclude with (1-p)% confidence that in multiple samples of this kind, the result would be replicated within 2 standard errors.' Usually p is 0.05 and we have 95% confidence. When the N of cases is large enough, the 'standard error' of an estimate (e.g. the mean) becomes rather small in typical studies, so you have a tight confidence interval.

How Large is Large Enough N? "Power Analysis"

The power analysis shows the minimum N for a given inference. You need to know what the expected mean or % is for a key variable to do 'power analysis for a mean' (Anderson, et al., 2005). Your team typically needs 30 cases for a mean to have a reasonable small confidence interval. In regression, around 15 cases per 'regressor' (the independent variables, including each dummy variable) is a minimum. There are even formulas for Power Analysis for Structural Equation Models (Kaplan, 2009: section 6.2) and for advanced models of married couples' correlated regressions (Kenny, Cashy and Cook, 2006).

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project.

A power analysis applies mainly to random sampling. Multi-stage designs that give representativeness with design weights may not allow a simple power analysis.

Example of power analysis

Our project requires at least 360 cases in each country. This allows us to represent the means at the State level for the types of clusters chosen. We can also conduct a complex regression in each country. In Bangladesh, for instance, N=360 gives us up to 720 men and women in the chosen couples, and each has 2 observations because we revisit in the second growing season. N of 1440 gives a good power for a regression. Our model with two time-periods of data can use a 'change-score' dependent variable for Y, allowing the rest to be a cross-sectional regression. N is then reduced back to 720 per country

Box 3: Weighting, or Design Weights and Non-Response Weights

Weighting is a method used to adjust a sample to allow for possible bias. Consider a situation where we wish to survey a population that we know to be 50% female but, due to non-response, only 40% of our sample is female. Gender is linked to many other characteristics so this creates a bias on other variables. We can apply a weight to correct for this imbalance. We calculate weights by dividing the percentage of a category in the population by the percentage of the category in the sample. In our example the weight for female respondents would therefore be 50/40=1.25 and non-female respondents 50/60=0.83 (see also De Vaus, 2001. pp.151-152). Design weights correct for different strata sampling proportions.

In fact we will aim for N=450 households.

When the sample is non-random but representative, the power analysis is only indicative.

Adjust Your Target Sample Size Upward for Expected Non-Response

In Bangladesh we expect rural residents to give only a 76% response rate. This will be fine for academic publication. But if N is only 76% of 450, our sample size is too low [N=337]. We therefore select from Upazila Household Lists a target of 592 cases. 592 is (1/.76)*450 cases.

Internal Validity of Sampling

The sampling can be tested for internal validity by considering whether the claims made are consistent with the methods used. Avoid **generalizing to the larger population that wasn't reached in the non-random cluster sampling.** See Figure 1. Each state is huge. We cannot genearlise to the State level. Large government surveys use random cluster selection across all states so that they can generalize to the whole country (e.g. National Sample Surveys in India). Internal validity exists if the whole study is consistent and coherent without making unfounded claims.

To validate your own sampling, write carefully and in full detail about what happened during the choice of population, selection of clusters, selection of stage 2 units, and in intra-household sampling; report the Non-response rate at each Stage; and give the N and Standard Error of every estimate (such as 'means' or prevalence %s). Notice that the s.e. of a % does not use the same formula as the s.e. of a mean of a continuous variable. That is why SPSS is crucial. If using Excel, avoid the formulas 'Std_Pop' and 'Std_Sample'. Move your data to SPSS or STATA to get the standard errors. If it is necessary to use Excel, buy the add-on, XLStat, so that you control the formulas.

External Validity of Sampling

To test the external validity of a sample, compare the sample means on crucial indicators with the local Census data. Take the mean age and its standard error, compare with local average for that type of person, e.g married women. Take the mean percentage who have primary-education only, and compare it with local Census data. Is the Census figure within the 95% confidence interval for this sample? If not, why not? If not your sample is not typical of this area. Report the results.

Validity is very important to increase the authoritativeness of your results. Raise your project's credibility by improving the sampling methods during the field visits. Later, improve your personal and team credibility by publishing tests of the external validity of the sample.

Correcting for Non-Response and Missing Data

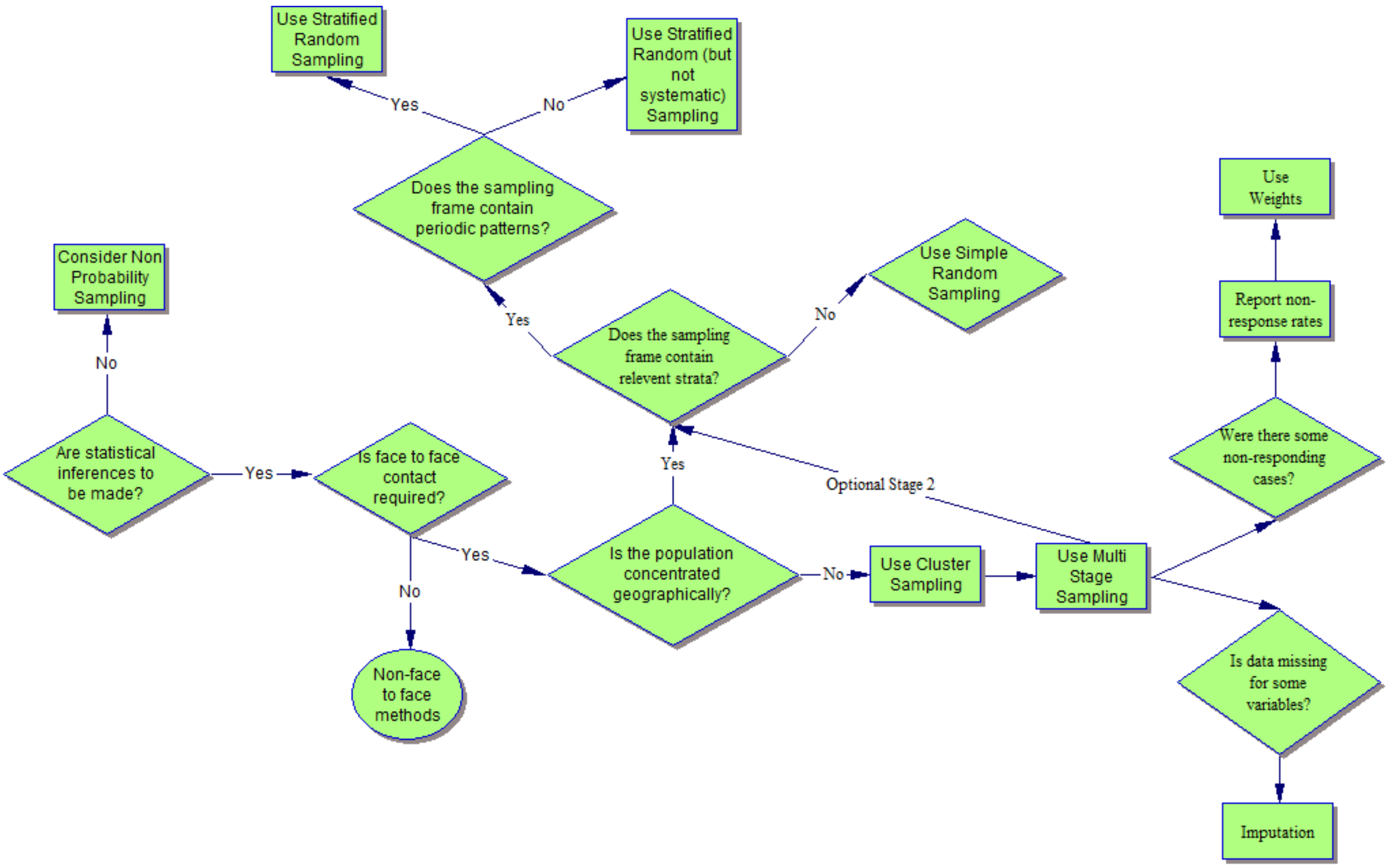
Missing data creates a number of issues, it can reduce sample size and increase sample bias. Whilst ultimately the best way of minimising missing data is a well designed survey, a variety of techniques exist to minimise the impact incomplete data has on our analysis. Two methods were used in this project: Imputation and Weighting.

Design weighting involves allowing for known bias in the sample by giving more 'weight' to underrepresented cases and less to over-represented cases. This was used to correct for cases missing due to non-response (see Box 3). Most surveys including the DHS and the NFHS use design weights.

Some cases were missing more selectively, for example if a respondent refused to answer a particular question. In such cases we can use a method called **imputation**. This is when we replace missing data with other values. There are a number of methods to decide which values are substituted for the missing data. The choice of method varies depending upon the type of study (for example in a longitudinal study design we would look to see how the respondent has answered in past waves of the study), what effect of the missing data we wish to minimise (sample size or sample bias) and how randomly the data is missing.

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project.

Figure 2. Selecting a Sampling Model (adapted from Saunders, et al., 2012: Ch. 7. Fig 7.3)



A word of warning… if a survey team decides not to use the Target List for sampling, but instead starts choosing households upon meeting people in the street, this will be **non-random, non-representative.** The Target List is the list of households chosen at Stage 2 from the agreed list of available households within the Stage 1 chosen unit, e.g. within the village.

Example: Household list for Village A has 360 households. A sample of size 50 is wanted. The expected non-response rate is 20%. Therefore 60 households (1/6) are randomly put on the Target List. Their household number or address is needed. 50 are reached, from these 60.

Why Does Sampling Matter? Answers from the Gender Norms and Labour Supply Project.

Project Partners:

The project ‘Gender Norms, Labour Supply and Poverty Reduction in Comparative Context: Evidence from India and Bangladesh” runs from January 2014 –December 2016. The Cathie Marsh Centre for Census and Survey Research (CCSR) at the University of Manchester is working on this in partnership with the BRAC Development Institute, Bangladesh, the Indian Institute for Dalit Studies, Delhi, and the German Institute of Global and Area Studies (GIGA).

See the web site http://www.cmist.manchester.ac.uk/research/projects/norms-labour-supply-and-poverty-reduction/labour-supply-additional-information/

Briefing Papers:

This briefing paper is the first in a series.

To join a mailing list about this project, and get our updates and Briefing Papers, please send an email to wendy.olsen@manchester.ac.uk.

Wendy Olsen

References:

Anderson, et al. (2005) *Statistics for Business and Economics*, 2nd ed. Hampshire: Cengage Learning.

De Vaus, David. (2001) *Research Design in Social Research,* New York: Sage Publications.

International Institute for Population Sciences (IIPS) and Macro International. 2007. *National Family Health Survey* (NFHS-3), 2005–06: India: Volume II.Mumbai: IIPS. pp.18-19.

Kaplan, David (2009). *Structural equation modeling : foundations and extensions*. 2nd ed. Los Angeles, London: Sage.

Kenny, David A., Deborah A. Kashy, and William Cook. (2006). *Dyadic Data Analysis*. New York: Guilford Press.

Kish, Leslie. (1995) *Survey Sampling.* London: Wiley.

National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Macro International. 2009. *Bangladesh Demographic and Health Survey 2007.*

National Institute of Population Research and Training (NIPORT), Mitra and Associates, and ICF International. 2013. *Bangladesh Demographic and Health Survey 2011.*

Saunders, Mark, Philip Lewis, and Adrian Thornhill (2012), *Research Methods for Business Students*, 6th ed. London: Prentice Hall/ Financial Times.

Stoop, Ineke, Jaak Billiet, Achim Koch and Rory Fitzgerald (2010), *Improving survey response: lessons learned from the European Social Survey.* London: Wiley.