

Published in final edited form as:

*J Dev Econ.* 2012 May 1; 98(1): 94–107. doi:10.1016/j.jdeveco.2011.08.007.

## A Comparison of EPI Sampling, Probability Sampling, and Compact Segment Sampling Methods for Micro and Small Enterprises

Li-Wei Chao<sup>a,b</sup>, Helena Szrek<sup>c</sup>, Karl Peltzer<sup>b</sup>, Shandir Ramlagan<sup>b</sup>, Peter Fleming<sup>a</sup>, Rui Leite<sup>e</sup>, Jesswill Magerman<sup>b</sup>, Godfrey B. Ngwenya<sup>b</sup>, Nuno Sousa Pereira<sup>c,d,e</sup>, and Jere Behrman<sup>a</sup>

Li-Wei Chao: chao69@wharton.upenn.edu; Helena Szrek: hszrek@wharton.upenn.edu; Karl Peltzer: KPeltzer@hsrc.ac.za; Shandir Ramlagan: SRamlagan@hsrc.ac.za; Peter Fleming: pcfleming@gmail.com; Rui Leite: Rui.Leite@gmail.com; Jesswill Magerman: JMagerman@hsrc.ac.za; Godfrey B. Ngwenya: BNgwenya@hsrc.ac.za; Nuno Sousa Pereira: nsp@egp-upbs.up.pt; Jere Behrman: jbehrman@econ.upenn.edu

<sup>a</sup>Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Room 239, Philadelphia, Pennsylvania 19104-6298, U.S.A

<sup>b</sup>The HIV/AIDS, STIs and TB (HAST) Research Programme, Human Sciences Research Council, Private Bag X41, PRETORIA 0001, South Africa

<sup>c</sup>Centre for Economics and Finance, University of Porto, Rua Roberto Frias, 4200-464 Porto, Portugal

<sup>d</sup>University of Porto Business School, Rua de Salazares 842, 4149-002 Porto, Portugal

<sup>e</sup>Faculty of Economics, University of Porto, Rua Roberto Frias, 4200-464 Porto, Portugal

### Abstract

Finding an efficient method for sampling micro- and small-enterprises (MSEs) for research and statistical reporting purposes is a challenge in developing countries, where registries of MSEs are often nonexistent or outdated. This lack of a sampling frame creates an obstacle in finding a representative sample of MSEs. This study uses computer simulations to draw samples from a census of businesses and non-businesses in the Tshwane Municipality of South Africa, using three different sampling methods: the traditional probability sampling method, the compact segment sampling method, and the World Health Organization's Expanded Programme on Immunization (EPI) sampling method. Three mechanisms by which the methods could differ are tested, the proximity selection of respondents, the at-home selection of respondents, and the use of inaccurate probability weights. The results highlight the importance of revisits and accurate probability weights, but the lesser effect of proximity selection on the samples' statistical properties.

### Keywords

Entrepreneurship; Sampling Methods; Informal Sector; Health; Self Employment; Micro and Small Enterprises

---

© 2011 Elsevier B.V. All rights reserved.

Corresponding address: Li-Wei Chao, Population Studies Center, University of Pennsylvania, 239 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6298 U.S.A. chao69@wharton.upenn.edu, Phone: +1.215.898.6441, Fax: +1.215.898.2124.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 1. Introduction

Micro and small enterprises (MSEs) are small businesses with up to 5 and 100 employees, respectively, although the precise definition of MSEs differs by governmental regime. MSEs have been estimated to provide nearly half of total employment in African countries (Liedholm and Mead, 1999; South Africa Business Guidebook 2004/2005, 2005/2006). Sampling MSEs in developing countries for research and statistical reporting purposes, however, is a challenge. Often there is no official registry of MSEs, and any such registry is incomplete because some owners refuse to register in order to avoid potential taxation and regulation. Even if censuses of MSEs were undertaken occasionally, the listing would soon be out of date given the rapid changes in the existence of MSEs.<sup>1</sup> This lack of a sampling frame usually creates an obstacle in finding a representative sample of MSEs with which to conduct research.

A statistically rigorous sampling strategy, commonly used in household surveys but also useful for surveys of MSEs, is the two-stage cluster sampling design. The first stage selects clusters or enumeration areas (EAs) within the geographic area of interest, and the second stage selects respondents for interview from within each chosen EA. Issues related to the first stage selection are well covered in survey textbooks (e.g., Kish 1965). This paper focuses only on the second stage selection. The statistical gold standard for the second stage is to select a random probability sample drawn from a sampling frame consisting of all *eligible* respondents in the EA. In this paper, this sampling strategy will be referred to as the Probability Sampling Method (PSM). However, without knowing the prevalence of MSE ownership and the location of the owners and non-owners in each EA, it is unclear how eligible respondents can be located, and hence how a probability sample can be drawn. The standard practice of many household surveys that use PSM is to perform a listing of all households in the EA to build the sampling frame of eligible respondents in that EA, and then to randomly select respondents into the study sample.

PSM has been applied to MSE studies in developing countries. For instance, de Mel, McKenzie, and Woodruff (2008) drew their sample using PSM to examine returns to capital in microenterprises in Sri Lanka. They selected enumeration areas, conducted a door-to-door screening survey of all households in the selected EAs to locate businesses, built a sampling frame of businesses using the screening survey, and randomly selected a subset of the business owners to form the final sample to be interviewed in detail. In their pioneering work under the GEMINI Project, Liedholm and Mead (1999) and their doctoral students studied MSE activity in countries in southern and eastern Africa and elsewhere. They first selected a subset of EAs from the whole country, did a listing of every household and business in each EA to canvass for small enterprise activity, and interviewed everyone with an MSE who was also available to answer a survey at the time of the listing. GEMINI's final sample, therefore, instead of being a random subsample of the EA, actually consisted of the universe of business owners available for interview at the time of the listing.

A modification of the PSM, called the Segment Sampling Method (SSM), selects EAs in the first stage as in the PSM, but then slices up the selected EAs into smaller area *segments* of similar sizes (usually defined by the number of households or respondents), and then randomly selects one segment from each EA to conduct a door-to-door listing of all the

---

<sup>1</sup>For instance, the World Bank Enterprise Survey in Nepal (Jain, 2010) initially sought a sampling frame from the Government of Nepal and other appropriate trade associations; however, the lists once obtained were deemed incomplete and out of date, and the survey team had to conduct a new census to build the sampling frame. A study by McKenzie and Sakho (2010) in Bolivia was successful in using an existing business census database collected 18 months prior to the study.

households in the segment to build the sampling frame, from which the respondents are then selected. SSM has been implemented in different ways and widely used in health surveys in developing countries (e.g., UNICEF's Multiple Indicator Cluster Surveys; Turner, Magnani, and Shuaib, 1996). In the traditional SSM (tSSM), a sub-sample of all the listed respondents from each chosen segment is then randomly selected to form the final study sample. This method was used in the Ghana Microenterprise Survey conducted by Fafchamps, et al. (in press). In compact SSM (cSSM), the segments are smaller than those in tSSM and all eligible respondents in the segment are included in the final study sample. The World Bank Microenterprise Surveys for the Informal Sector in southern and eastern Africa and in Nepal used cSSM (Jain 2010; Gelb et al. 2009). The field team first identified clusters of informal sector activity, divided these clusters into segments, randomly selected segments from all possible segments, and then conducted door-to-door surveys with all the households in the selected segments. Because the segments in cSSM are relatively small in area, a potential pitfall occurs when the segment does not contain the minimum acceptable sample size of a specific target population (e.g., larger businesses), and a larger segment had to be redrawn (as occurred in Jain, 2010).

An alternative, potentially cheaper, and easier to implement sampling strategy, designed by the World Health Organization's Expanded Programme on Immunization (EPI) to track immunization coverage, could also be applied to MSE sampling. This EPI method, based on quota sampling, uses a modified version of the two-stage cluster sampling design, with the first stage identical to the first stage of the PSM. In the second stage, the interview team starts at the center of the EA, chooses a random direction by spinning a bottle (or pen), counts the number of households from the center to the edge of the cluster in the chosen direction, and randomly chooses one household as the starting point. Eligible respondents are then sought from the household at the starting point. If an eligible respondent is available, an interview is conducted; if the household has no eligible respondents, the nearest household from this household is screened for eligible respondents, until an eligible respondent is found and interviewed. Subsequent eligible respondents are sought from households closest to the current household, with the process continuing until a quota number of interviews has been conducted (e.g., Henderson and Sundaresan, 1982). This method is referred to as the EPI Sampling Method (ESM) in this paper.

The standard practice of PSM and SSM when respondents are absent during either the listing exercise or the actual survey (after being selected from the census list) is to perform revisits, by making appointments and returning to conduct the interview. For instance, up to 20% of the sample in the World Bank Enterprise Survey in Nepal (Jain 2010) were revisits. ESM, in contrast to PSM and SSM, does not perform revisits. Instead, ESM skips eligible respondents who happen to be absent at the time of the fieldworker's visit and interviews other eligible respondents in subsequent households.

A central question for entrepreneurship research is which sampling method should be used, especially in settings where existing sampling frames for MSEs are lacking. One important consideration is how well the sampling method's sampling distribution represents the study population. A comparison of the different sampling methods reveals three main issues that may affect the properties of the methods' sampling distribution.<sup>2</sup> The first issue is *proximity selection*: respondents that are closer in geographic proximity to each other may be more likely to be selected into the final study sample. Because respondents nearer each other are

<sup>2</sup>A fourth issue is the non-probability selection of respondents. This occurs only in EPI implementations that start by randomly selecting a house among all houses on an imaginary line that extends from the center of the cluster to the periphery. Because houses nearer the cluster's center are more likely to be on the line than houses in the periphery, houses closer to the center have a higher probability of selection (Brogan et al. 1994). This problem can easily be corrected by selecting a random starting point using GPS software or random selection of a grid among grids placed over the cluster's map (Grais et al. 2007).

likely to be more similar, this homogeneity could potentially increase the variance of the sample, reducing efficiency (Kish, 1965). Proximity selection should be most prominent in cSSM which selects respondents within a very small segment, followed by ESM in which respondents are selected based on proximity to the last visited household. Respondents in PSM samples are most likely to be randomly distributed throughout the EA and thus suffer the least from problems associated with proximity selection. This is indeed what is found by Lemeshow et al. (1985), who created 30 artificial EAs with different vaccine coverage, drew computer simulated samples using ESM and PSM, and compared the sampling distributions. ESM samples were more likely to have higher bias and mean square error (which is a measure of both bias and variance) than PSM samples, especially when the location of the target population (e.g., vaccinated children) was unevenly distributed in the EA. Proximity selection can be mitigated by systematic or interval sampling, which skips a set number of adjacent eligible respondents before selecting the next respondent (e.g., Grais et al. 2007; Luman et al. 2007; Rose et al. 2006; Bennett et al. 1994; Fafchamps, et al. in press), by starting at more than one random starting point in the EA in the ESM (e.g., Bennett et al. 1994), or by slicing up the EA into even smaller compact segments and choosing more than one of the compact segments from the EA in cSSM. Bennett et al. (1994), using computer simulation, drew samples from real census data collected in 30 villages in Uganda, and found that the unmodified ESM was comparable to PSM and modifications of the ESM in estimating nutritional status, but that the unmodified ESM had greater mean square error for some household health care related and socioeconomic variables than estimates derived from interval ESM or ESM implemented in all four quadrants of the cluster. Yoon et al. (1997), using computer simulation with real census data collected in Nepal to estimate prevalence for diarrhea and dysentery, also found that ESM and interval ESM produced estimates with more bias than PSM.

The second issue that may affect the properties of the methods' sampling distribution is *at-home selection*, when the sampling method does not implement revisits.<sup>3</sup> Sampling methods that do not entail revisits may yield biased estimates if the reason for the absence is systematically related to the variables being estimated. For example, owners of MSEs that focus on personal services (e.g., hair salons) are more likely to be present on the business premise than owners of MSEs that focus on construction (e.g., plumbers). Owners of MSEs without any employees may be more likely to be present than owners of MSEs with employees, who could substitute for the owner while the owner is offsite. If a sample of respondents who do not own MSEs is included as a comparison group to MSE owners, then those respondents who are found to be home at the time of the interview visit might have systematic reasons for being home, including poor health. Samples drawn with just respondents reached during the initial visit, therefore, may paint a distorted picture of the sampled EAs. Even though it is common practice under ESM not to implement revisits, the impact of at-home selection due to non-completion of revisits has not been tested in the literature that reviews the EPI method. Revisits were also not performed in the studies under the GEMINI Project (McPherson 1996 footnote 16), although revisits are commonly performed under SSM and PSM. In the GEMINI Project, about one-third of the households were unavailable for interview during the first visit (McPherson 1996). However, this did not result in biased statistical estimates, mostly likely because the reason for the absence was not systematically related to the variables being estimated (Parker and Dondo, 1991).

The third issue that may affect the properties of the methods' sampling distribution is *inaccurate probability weighting*, and this occurs when the probability of any given observation being selected into the final sample is unknown or inaccurate, resulting in the

---

<sup>3</sup>Kish (1965, page 532) refers to non-response from absent respondents as problems of *not-at-homes*.

application of incorrect weights during statistical analysis. The probability of a specific observation from a specific EA being selected into the final sample depends on both the probability of the EA being selected in the first stage and the probability of the specific observation from within that EA being selected in the second stage. A common practice is to select EAs into the sample based on probability proportional to the size (*pps*) of the population (or the number of households) in the sampling frame EAs. If the first stage EAs are selected with *pps*, then as long as an equal number of observations are selected from each EA in the second stage, the second stage probability of selection is equal for all the selected observations from all the selected EAs, resulting in a self-weighting sample, and only first stage weights are needed in the statistical analysis. The problem, however, is that the measures of size often are based on outdated population censuses, and the assumptions for *pps* are not met. For MSE-related research, in particular, the proper measures of size are often unknown, without a listing of the EA to determine the number of businesses. Any sampling method that does not entail a census listing but wrongly assumes *pps* will lack the correct probability weights in the statistical analysis, potentially resulting in biased estimates of sample means and underestimates of the sample variances (Brogan et al. 1994). Milligan et al. (2004) selected one cSSM sample and a separate ESM sample in the Western Region of Gambia to study vaccination rates, and tested the effect of no weighting (by assuming *pps* using the 1993 population census) vs. weighting (by using weights from a listing done in 2000–2001 as part of the study). Estimated mean vaccine coverage rates were similar regardless of weights in cSSM and ESM samples, but variance was higher in weighted estimates. Greater population growth since 1993 in some EAs had overstretched existing health care delivery infrastructure, resulting in pockets of high and low vaccine coverage (and thus greater variance) in high growth EAs. Because high growth EAs should be weighted more than EAs with fewer eligible respondents, variance is underestimated when weights are not applied.

Because PSM uses random probability sampling of the entire listed EA (and hence does not suffer from proximity selection), performs revisits to reach those not at-home, and applies weights derived from the listing, its sampling distribution should be most representative of the parent population from which the samples are drawn. While it is the gold standard in representativeness, PSM is also more costly (Yoon, 1997), given that the listing and the revisits are labor intensive cost drivers. PSM that lists the whole EA is expected to cost more than tSSM, which in turn is expected to cost more than cSSM. ESM de facto lists only respondents along the way, so it should require the least amount of labor and time; it also skips absent respondents and thus saves on revisit costs. An important question, then, is whether the more representative sampling methods are worth the cost.

In the present paper, we draw computer simulated ESM, SSM, and PSM samples from a census-universe dataset collected from respondents with and without businesses in townships around the Tshwane (formerly known as Pretoria) Municipality of South Africa. We then compare the characteristics of the generated samples to the characteristics of the universe population. The statistical effects of proximity selection, at-home selection, and inaccurate probability weights are examined to help illuminate the reasons underlying the differences between the various methods. This is possible because we have GPS readings for all households and businesses in our EAs, the census record shows whether the interview was conducted during a first visit or a revisit, and the complete census allows us to calculate the probability of selection of each selected respondent to derive weights. To our knowledge, there has not been a comparison of various sampling methods applied to studies related to entrepreneurship or MSEs. We are also not aware of studies that compare ESM, SSM, or PSM to analyze the importance of revisits, including analyses that consider the extent to which revisits may improve the simple ESM. This paper, thus, makes contributions to the sampling and entrepreneurship literatures in the developing world.



## 2. Method

### 2.1. The survey and the data

The first wave of the South African Panel Study of Small Business and Health, a longitudinal survey that commenced in the fourth quarter of 2009, provides the data for this paper. The main goal of this data collection is to analyze the relationship between health and entrepreneurship in African townships in South Africa. The survey collects data on health, psychology, and entrepreneurship from owners of MSEs and from randomly selected respondents who do not own businesses.

A two stage stratified probability sampling design was used to select the full sample. Based on sample size requirements for a study on health, HIV/AIDS, and entrepreneurship, the first stage selected a total of 22 EAs out of all the African dominated EAs<sup>4</sup> in the Tshwane Municipality stratified by the five regions of Tshwane, by formal and informal geotypes, and by personal income above or below the median personal income at the ward-level.<sup>5</sup> We then obtained aerial photographs of the EAs using Google Earth and proprietary GIS databases containing GPS coordinates, street names, and boundaries of the EAs. The interview team went to each EA, updated the aerial photographs with recent changes to the EA, and assigned a unique number to each property stand. Most stands in the townships have at least one house. We then conducted a census in each of the EAs, by canvassing every stand. All business owners in the EA were selected into the sample. For households without businesses that had a member planning to open a business in the next 12 months, the business planner was selected. For households without owners or planners but with a household member who once owned a business that closed in the last two years, the business ‘closer’ was interviewed. For households without anyone involved in a current, planned, or closed business, a household member was randomly selected for interview using the Kish method (Kish 1949)<sup>6</sup>, which ensures the selection of a random person from the household, rather than the interview of a convenient person who happens to be at home at the time of the initial interview visit.

This “modified” census methodology reaches all stands as well as all businesses in the EA, covers all current business owners, but also includes planners, closers, and a comparison population not involved in business activities. For this paper, current business owners are classified as “Owners.” Planners, closers, and the non-business respondents are classified as “Non-Owners.” In summary, the unit of observation in the census is the stand. For each stand, the census ‘universe’ contains (for non-empty and non-refusal stands) either one (or more) business owner(s) or one and exactly one non-owner respondent.

The interviews were conducted in the preferred language of the respondent (English, Zulu, Pedi, Tswana, Tsonga, or Southern Sotho), using Google Phones with Android 1.5 operating system and interview software Open Data Kit (ODK) Collect version 1.1 (<http://code.google.com/p/open-data-kit/wiki/CurrentDeployments>). For interview visits where no one was home or available to be interviewed, the fieldworkers made revisit attempts at different times of the week when the interview team was in the EA. Only after three revisits and approval by the supervisor would the interview be designated as unsuccessful.

<sup>4</sup>Only non-tribal EAs with over 95% African populations (based on the 2001 South African Census) were included in the EA sampling frame.

<sup>5</sup>Ward is a geographic division bigger than the EA. A ward in our sample contains 25 EAs on average. The ward is the smallest geographical unit for which we had access to income data.

<sup>6</sup>The Kish is implemented by using a grid on which the interviewer needs to match a coversheet ID (column) with the number of respondents in the household (row) to determine which person is selected for interview.

## 2.2 Computer simulation procedure

Because we conducted a census of the whole EA, the data set contains the universe of eligible sampling points within each EA.<sup>7</sup> This universe allows us to draw different samples, *using the same data set*, using different sampling methods. For each sampling method, we simulated 1,000 draws from each EA, each draw with 10 owners and 10 non-owners. We then selected one simulated draw (with replacement) from each of the 22 EAs and combined them to form one simulated sample (with 220 owners and 220 non-owners), and we did this for all the simulated draws from each EA. With 1,000 simulated draws in each EA, we have 1,000 simulated samples for each sampling method.<sup>8,9</sup>

**2.2.1 Probability sampling method with and without revisits**—Using the full universe (with both first visit and revisit respondents) as the sampling frame, we drew 1,000 samples using PSM that allowed for revisits. We also drew 1,000 samples using only the first visit respondents in the universe as the sampling frame, to simulate samples drawn without doing revisits. Each simulated draw using the PSM consisted of 10 owners and 10 non-owners *randomly* selected from all the owners and all the non-owners, respectively, in each EA. The EA-level draws are then combined across EAs, forming 1,000 simulated PSM samples, each containing 220 owners and 220 non-owners. These 1,000 simulated PSM samples are only a subset of all possible random samples that could have been chosen by PSM.

**2.2.2. EPI sampling method with and without revisits**—Using GPS information in our data, we simulated the ESM draws by randomly selecting one stand within each EA as the random starting point for that EA. The respondent on the starting point stand was included as the first respondent. The next respondent came from the stand closest in GPS distance to this first respondent, with all subsequent respondents drawn based on geographic proximity to the previous respondent. This process continued until at least 10 owners and 10 non-owners were found for each EA. We replicated this procedure 1,000 times to form the 1,000 simulated draws from the EA. Combining one simulated EA draw from each of the 22 EAs, we formed 1,000 simulated samples. To simulate ESM samples drawn from the traditional EPI sampling methodology (which does not make revisits but selects a substitute household nearest the one that is absent), we first discarded from the census universe all eligible respondents interviewed during revisits prior to drawing the 1,000 simulated ESM samples. To examine how revisits would affect ESM sample properties, we next drew another 1,000 simulated ESM samples using the entire universe (with both first visit and revisit respondents) as the sampling frame.

**2.2.3 Compact segment sampling method with and without revisits**—We simulated cSSM samples as follows. First, we used GPS to choose a random starting point stand from each EA. We next selected the 10 business owners that were the shortest GPS distance from the starting point stand. We drew 10 non-owners similarly. We next combined one simulated draw from each EA across all 22 EAs to form one simulated cSSM sample. We did this for each of the 1,000 starting points to form a total of 1,000 simulated ESM samples. We repeated this methodology twice to create cSSM samples with and without

<sup>7</sup>We were not able to interview all respondents in the EAs, but we consider our sample of stands that we were able to interview to be the universe of all eligible sampling points.

<sup>8</sup>In a few EAs, we did not reach 10 or more owners or non-owners among first visits, so the total number of respondents of each ownership type could be less than 220 for the sampling methods that did not allow for revisits.

<sup>9</sup>We chose a sample size of 10 owners and 10 non-owners in each EA because our research question on the statistical effects of at-home selection requires a comparison of first visit with revisit samples, and the number of first visit interviews was less than half of the total number of respondents in many EAs. Setting a larger sample size for each EA would increase the probability of ending up with identical respondents in different simulated draws. Issues of how to set an optimal sample size quota have been addressed in the literature (e.g., Lemeshow and Robinson, 1985; Kish, 1965) and are not addressed in this paper.

revisits. Note that this is not a perfect replication of compact segment sampling, which would require a demarcation of segments on the EA map, a random selection of one segment, and then the inclusion of all owners and non-owners in the randomly chosen segment. Because the prevalence of businesses varied between our EAs, and the density of businesses also varied within the EA, creating compact segments with similar numbers of the target population group (e.g., 10 non-owners and 10 owners for each EA) was not feasible in our universe. Instead, we opted for the 10 non-owners and the 10 owners closest to the starting point stand. This essentially created the smallest possible compact segment that still filled the quota for the respondent types.

It is worth noting that the ESM samples consist of respondents that are closest to the previous household visited. The cSSM samples consist of respondents that are closest to the starting point stand. The geographic locations of ESM respondents within an EA fall along a pattern that resembles a crawling snake with the tail located at the starting point stand and the head at the last stand that filled the quota. The location patterns for the cSSM respondents form a compact circular area around the starting point stand. Based on this general pattern, cSSM samples should suffer more from proximity selection problems than ESM samples.

### 2.3 Application of sampling weights

We used sampling weights to adjust for the first and the second stage of our sampling procedure. For the first stage, we calculated weights to reflect the inverse of the probability of that EA being selected out of all the eligible EAs in Tshwane, taking into account the stratification (region, genotype, and income levels) and the number of eligible EAs in each stratum. These first stage probability weights are referred to as *EA-weights* in this paper.

Because we conducted a census of each EA, we know the actual number of owners and non-owners in each EA. This information is not routinely available in ESM or cSSM samples. (Although the number of businesses and non-businesses for a section of the EA will be known, that information may not correctly reflect that for the full EA.) We also know whether the respondent was interviewed during a first visit or a revisit. Knowing both the number of respondents of each business ownership type in each EA and the first visit or revisit status of each interview, we can calculate the probability of any owner or non-owner being selected into the sample from this EA when the sampling frame consists of first visits only or first visits plus revisits. The weight of any respondent within the EA is the inverse of the probability of that respondent being selected from that EA's sampling frame for that respondent type. This weight differs by whether the respondent is an owner, a non-owner, a first-visit respondent, or a revisit respondent. These weights are the second stage weights in our sampling design. The product of the first stage weight and second stage weight is the inverse of the probability of a specific respondent getting selected into the sample (out of all the potential respondents in all eligible EAs in Tshwane). In this paper, we use the term *probability weights* to refer to the joint product of first and second stage weights.

To examine the effect of accurate weights on sample properties, we weight the sample observations by either EA-weights or by the probability weights. The EA-weights do not reflect the actual population or the proportion of businesses to non-businesses within each EA.

### 2.4 Nomenclature of the Samples

To clarify, our final simulated samples consist of (3 sampling methods)  $\times$  (2 kinds of weights)  $\times$  (2 kinds of revisit status) for a total of 12 sets of samples. We compare these 12 sets of samples in the rest of the paper. Each set contains 1,000 simulated samples drawn by



using one of three sampling methods. We also use the following suffix abbreviations to denote the different samples. Methods with a lowercase r are those that use the universe with first visits and revisits as the sampling frame; methods with nr (to stand for ‘no revisits’) use the universe without revisits as the sampling frame. Methods with the lowercase e indicates the application of EA-weights (i.e., first stage weights only), and those with lowercase p indicates the application of probability weights (i.e., both first-stage and second-stage weights multiplied together):

1. ESMnre: ESM with no revisits, EA-weights
2. SSMnre: SSM with no revisits, EA-weights
3. PSMnre: PSM with no revisits, EA-weights
4. ESMnrp: ESM with no revisits; probability weights for no revisits
5. SSMnrp: SSM with no revisits, probability weights for no revisits
6. PSMnrp: PSM with no revisits; probability weights for no revisits
7. ESMre: ESM with first and revisits, EA-weights
8. SSMre: SSM with first and revisits, EA-weights
9. PSMre: PSM with first and revisits, EA-weights
10. ESMrp: ESM with first and revisits, probability weights for first- and re-visits
11. SSMrp: SSM with first and revisits, probability weights for first- and re-visits
12. PSMrp: PSM with first and revisits, probability weights for first- and re-visits

We continue to use ESM, SSM, and PSM (without the suffixes) to refer to each sampling method in general, but we add suffixes when the specific variant of the sampling method needs to be clearly stated. Using different subsets of these types of samples, we can test for the effects of proximity selection, at-home selection, and inaccurate weights in a *ceteris paribus* design, varying one effect at a time while keeping the other two effects constant. For instance, PSM without revisits compared to PSM with revisits would reveal the pure effect of at-home selection when proximity selection is minimal. Comparing ESM (or SSM) without revisits to ESM (or SSM) with revisits would allow us to assess the influences of at-home selection when proximity selection is present. A comparison between PSM with revisits vs. ESM (or SSM) with revisits allows us to address the pure effect of proximity selection with no at-home selection effects, and a comparison between PSM without revisits and ESM (or SSM) without revisits would reveal the effects of proximity selection in the presence of at-home selection. Furthermore, when these comparisons are made across methods (PSM vs. ESM vs. SSM), we can keep the weights constant by using probability weights for all three samples to better isolate proximity and at-home selection. This methodology allows us to understand the relative importance of proximity selection, at-home selection, and the use of inaccurate weights and may inform quality inferences about other sampling methodologies.

## 2.5 Survey measures

Our study compares demographic characteristics, health, risk aversion, and business measures. Health is measured using the 12-Item Short Form Health Survey (SF12), an instrument that assesses health symptoms, functioning, and quality of life along two dimensions: mental and physical health (Ware et al., 1995). We also included a general risk question, taken from the German Social Economic Panel (SOEP) (Dohmen et al., 2005), but modified the scale to range from 1 to 5. The business questions we analyze here include an indicator of the industry of the business (services, manufacturing, or retail), the age of the

business, and whether one or more people are engaged in the business (identifying single vs. multi-person businesses). We also ask about business profit, using a measure developed by de Mel, McKenzie, and Woodruff (2009).

## 2.6 Sampling statistics

For each survey variable, we first calculated the population mean and the population standard deviation. We next calculated the sampling statistics for each of the sampling methods' distributions as follows (See Table A.1 for the definitions of the sampling statistics used in this paper). For each of the 1,000 samples, we calculated the mean of each variable, which we call the *sample mean*. We next used the 1,000 sample means to build a sampling distribution of means (for each variable). The term 'sampling distribution' in this paper refers to the sampling distribution of means; i.e., the distribution of the 1,000 sample means derived from the 1,000 simulated samples. For each sampling method, we calculated the *mean*, *bias*, *variance*, and *mean square error (mse)* of the sampling distribution for each variable. Because  $mse = bias^2 + variance$ , the mean square error penalizes both the bias and the variance of the sampling method. As the different demographic and business variables are measured with different units (e.g., years for age, percent for married or not, normed scale for SF12, etc.), we also calculated the *absolute relative bias (arb)*, which is the absolute value of the bias normalized by the universe mean (Lemeshow et al., 1985). The *arb* essentially converts all the variables with different units to percentages. To facilitate comparison of variances between methods, we calculated the design effect (*deff*), which is the ratio between the variance of the sampling distribution derived from a particular sampling method and the variance of the sampling distribution derived from the PSMrp (i.e., PSM with revisits and probability weights) – the gold standard in our study. We also calculated the *mse ratio*, which is the mean square error of each method's sampling distribution normalized by the mean square error under PSMrp. These sampling statistics are then used to compare the different sampling methods. In the following sections, we motivate how these statistics can be used to explore the effects of proximity selection, at-home selection, and the use of inaccurate weights.

**2.6.1 Proximity selection and its effect on sampling statistics**—If respondents in close geographic proximity within the EA are more homogeneous in their characteristics than respondents far apart in the EA, samples drawn with proximity selection may not be representative of that one EA. Bias in the estimates from a single sample from a single EA may, thus, be large. However, because most studies draw many more EAs than just one (e.g., the traditional EPI methodology samples from 30 EAs, and our study has 22 EAs), the bias averaged across all the sample EAs may become important only if the direction of the bias is consistent across the EAs. Although some samples drawn will have more bias than others, there is no a priori reason for the *mean* of the sampling distribution to be biased. Despite this, ESM has been found to be more biased than PSM (Lemeshow, et al., 1985; Bennett et al., 1994; Yoon et al., 1997), and interval sampling has been found to improve the bias for some measures (Bennett et al., 1994) but not for others (Yoon et al., 1997).

Although the sampling distribution may have a *mean* that does not differ much from the true population mean, proximity selection increases the *variance* of the sampling distribution in the presence of clustering of respondent characteristics. The *deff* statistics found in the literature, however, differ by study and by variables. Lemeshow et al. (1985) found ESM to have a *deff* around 1 for one set of assumptions and 1.88 for another, suggesting that the variance of the sampling distribution from ESM differs by the configuration of the target population in the EAs (especially population density and clustering of homogeneous respondents).<sup>10</sup> However, Yoon et al. (1997) found ESM's *deff* for diarrhea and dysentery to be consistently below 1 (Yoon et al., 1997), suggesting that ESM samples had lower

variance than PSM samples. Bennett et al. (1994) also found the *deff* to be below 1 for 6 out of 19 variables examined when ESM is compared to PSM, with the variance in ESM further reduced by interval sampling or starting from four random points. In summary, while proximity selection is an issue that has dominated the literature in evaluating the EPI sampling method and in attempts to find ways to improve the ESM (such as interval sampling, starting from more than one random point in the EA, etc.), the evidence that proximity selection affects sampling statistics is mixed.

**2.6.2 At-home selection and its effect on sampling statistics**—Sampling methods such as the ESM that do not implement revisits may suffer from *at-home selection* if the respondents who happen to be available for interview have characteristics systematically different from those who are away. If these differences are consistent across all the EAs, the sampling distribution's mean will be more biased than if the samples were drawn from a sampling frame with both first visit and revisit respondents. Many causes of being at-home are consistent across EAs. For instance, non-owners who are at home are more likely to be of poorer health than non-owners who are away; owners of multi-person businesses with employees who can cover for them may be able to work offsite on other duties, and may thus be away from their place of business when the interviewer arrives. These characteristics are likely to persist across different EAs and accumulate, and thus, will produce *bias* in the mean estimates of variables. Also, because the difference between first visit respondents across EAs is likely to be less than the difference between first visit and revisit respondents within and across EAs, samples with just first visit respondents will more likely have a lower variance than samples with both first visits and revisits.

**2.6.3 Use of inaccurate weights and its effect on sampling statistics**—The application of correct weights that reflect the inverse of the probability of selection of each respondent is the norm in statistical reporting. When the first stage EAs are selected with *pps* of the target population in the EA and an equal number of observations are selected from each EA, the sample will be self-weighting – as is assumed by many studies using ESM. However, because of differential growth of EAs since the most recent census and the often unknown eligible population for specific outcome measures (such as the number of entrepreneurs in an EA), assuming the sample to be self-weighting could present a problem. When small EAs with fewer *eligible* respondents are weighted the same as larger EAs with larger numbers of *eligible* respondents, this will pull the mean estimates of the whole sample with all EAs combined closer to the estimates of the smaller EAs than if weights are properly applied. This may produce biased estimates of the mean if the reason that some EAs are larger than others (e.g., availability of clean water) is systematically correlated with the variable of interest (e.g., success of businesses in the service industry). Additionally, the variance will be underestimated without weights, because the much larger number of possible samples drawn from the larger EAs (some of which may have characteristics that differ significantly from the mean of the sampling distribution just by chance) is weighted the same as the much smaller number of possible combinations of respondent draws from the smaller EA. Milligan et al. (2004) found that proper weights did not change the estimates of the mean (for the cSSM method), but did increase the *deff* for 11 out of the 15 vaccines examined.

---

<sup>10</sup>Even though Lemeshow et al. (1985) does not present them, the *deff* values can be calculated from the 'Over all clusters' numbers in the tables, using  $mse = bias^2 + variance$ .

### 3. Results

#### 3.1 Sample characteristics by business ownership and by revisit status

In the 22 EAs in our census universe, a total of 3,117 respondents were interviewed covering 72.4% of all the stands (Table A.2). The stands where no one was interviewed included those with no structure, with no one at home after repeated revisits, or with someone who refused the interview. Over half of the interviews were done during a revisit. Among the interviewed respondents, 35% currently own a business, which is similar to the 31% we found in the eThekweni (Durban) Municipality in 2002 (Chao et al., 2007).

Table 1 presents the demographic, health, risk, and business variables for the universe of the surveyed respondents in our 22 EAs, by business ownership and by first visit and revisit status.<sup>11</sup> Owners are more likely to be older, male, married, more educated, with higher risk preference, and of slightly lower physical and higher mental health than non-owners. Respondents reached in the first visit are more likely to be older, female, unmarried, less educated, of lower risk preference, and of lower health than those interviewed during a revisit. Owners of businesses with at least one employee were more likely to be absent during our first visit (columns 5 and 6, last row), suggesting that perhaps the employee was substituting for them while they pursued activities elsewhere. The table raises the concern that sampling methods that do not perform revisits, such as the ESM, may produce a sample of respondents with characteristics that differ significantly from respondents in a sample drawn from a universe with revisits.

#### 3.2 Comparison of sampling methods

Table 2 presents the average sampling statistics for each of the 12 sets of computer-simulated samples, averaged over all the variables examined, to serve as a summary measure of the overall effect of different sampling methods. (Sampling statistics separately for non-owners, owners, and the business appear in Table A.3.) The row headings denote whether the sampling statistics are for averages of variables from non-owners, owners, or the business, or a total average of all three types of variables combined. The column headings show whether the sampling distribution was built with ESM, SSM, or PSM; whether it was built using a sampling frame with or without revisits; and whether EA-weights or probability weights were used. (Section 2.4 contains the nomenclature of these various methods, and 2.6 describes the sampling statistics.) Using the 12 sets of samples, we can examine the effect of proximity selection, at-home selection, and inaccurate weights one at a time while holding the other two effects constant.

**3.2.1 Comparison of methods to examine proximity selection**—The effect of proximity selection on sampling statistics *ceteris paribus* can be examined by comparing the methods within triplets: columns 1, 2, 3 (with no revisits and with EA-weights), or columns 4, 5, 6 (with no revisits and with probability weights for no revisits), or columns 7, 8, 9 (with revisits and EA-weights), or columns 10, 11, 12 (with revisits and probability weights for first visits and revisits). Within any triplet, proximity selection should be greatest with SSM, followed by ESM, and least with PSM. Nevertheless, we find that the average absolute relative bias per variable does not seem to differ appreciably between samples with and without proximity selection, controlling for revisit status and weighting status. Using the rows for the total averages (which is the average for all the non-owner, owner, and business variables combined), we see that *average arb* is about 7 or 8% for all three sampling methods without revisits and with EA-weights, and higher or lower for the other triplets. The within triplet *average arb* is less variable than the *average arb* for the same sampling

<sup>11</sup>Three respondents with missing values to key demographic variables were deleted from the data set.

method but across different revisit or weight status.<sup>12</sup> This suggests that proximity selection has very little effect on *average arb* relative to the effect from revisits or weights. Additionally, the *average deff* does not seem to vary much within triplets. For instance, ESM<sub>nr</sub> (which is the traditional EPI methodology) has *average deff* that is similar to that for SSM<sub>nr</sub> or PSM<sub>nr</sub>. The *average mse ratio*, which encompasses both the effect of *bias* and *deff* (which indirectly contains *variance*), also shows that the within triplet variation to be much less than between triplet variation. These findings suggest that despite the theoretical importance of proximity selection, its effect is relatively small in our population.

**3.2.2 Comparison of methods to examine at-home selection**—We next compare between samples simulated without vs. with revisits, while keeping sampling method and weighting status constant: 1 vs. 7 (without revisits vs. with revisits, using ESM and EA-weights), 2 vs. 8 (SSM; EA-weights), 3 vs. 9 (PSM; EA-weights), 4 vs. 10 (ESM; probability weights), 5 vs. 11 (SSM; probability weights), and 6 vs. 12 (PSM; probability weights).

The *arb* for most variables is lower with revisits than without revisits, especially in the presence of proper weights. This suggests that not doing revisits may produce samples with more *bias* – highlighting a potential pitfall of the traditional EPI sampling methodology. The *deff* and *average deff* are mostly higher when revisits are done compared to when they are not done. This implies an increase in *variance* of the sampling distribution with revisits, which is an expected finding, since doing revisits adds a sizeable fraction of the population into the sampling frame, and this added population has statistically significantly different characteristics than the population without revisits, increasing the heterogeneity within each EA but with this heterogeneity consistently increased in all the EAs. With revisits, the *average mse ratio* decreases for all the methods (and for almost all the variables if examined separately). This implies that although variance is increased with revisits, this is more than offset by the reduction in *bias*, with a net impact of reducing *mse*. Revisits, therefore, not only reduces the *bias* of the statistical estimates but also reduces the overall dispersion around the population mean.<sup>13</sup>

**3.2.3 Comparison of methods to examine inaccurate weights**—Given the lack of evidence of a statistical effect from proximity selection (Section 3.2.1) and the fact that making revisits improves sampling statistics (Section 3.2.2), the question remains as to how important it is to apply accurate weights. This is not a trivial question since finding the proper weights requires knowing the *eligible* population size and relative proportions of the different populations of interest. A whole-EA listing is required to obtain this information, which *de facto* means a PSM<sub>rp</sub> method might as well be adopted.

In Table 2, the effect of inaccurate vs. accurate weighting, while keeping sampling method and revisit status constant, can be seen when one compares columns 7 vs. 10 for ESM, 8 vs. 11 for SSM, and 9 vs. 12 for PSM (for samples drawn with revisits); and columns 1 vs. 4 for ESM, 2 vs. 5 for SSM, and 3 vs. 6 for PSM (for samples drawn without revisits). The weights used in columns 10, 11, and 12 are the inverse of the probability of selection of each respondent, separately by business ownership, for all the respondents in the EA, including first visit and revisit respondents. The weights (called *non-revisit probability weights*, or *nrProbability*) in columns 4, 5, and 6 are the inverse probability of selection using a sampling frame of only first visit respondents.

<sup>12</sup>The variance for *average arb* within triplets ranges from 0.00001 to 0.00015, but within the same method but across revisit and weighting status ranges from 0.0008 to 0.0014 (not reported in the table).

<sup>13</sup>The *variance* of the sampling distribution is a measure of the dispersion around the sampling distribution's (biased) mean, but the *mse* is a measure of the dispersion around the population mean.



The results show that the application of non-revisit probability weights worsens almost all the sampling statistics. That is, for samples drawn without revisits, the application of probability weights derived from a sampling frame without revisits worsens the sampling statistics compared to the use of EA-weights. However, for samples drawn with first visit and revisits, the application of the correct weights improves the sampling statistics: with a lower *average arb* and a lower *average mse ratio*. Nevertheless, the improvement in *mse* is only marginal. Samples drawn with revisits already have sampling statistics that are very close to samples drawn by PSMrp, the gold standard, even with just the application of EA-weights. Therefore, although the application of proper weights further improves the overall sampling statistics (as measured by the *bias* and the *mse*), the improvement is only marginal beyond that already achieved by revisits, so using EA-weights may not necessarily be a bad option.

### 3.2.4 Comparison of methods by how likely the sample (not sampling) statistics are wrong

—Although we have 1,000 simulated samples from each sampling method, most studies in real life have only one sample drawn using one survey method. Hence, a valid question is how likely it would be for a sampling methodology to produce statistical estimates that differ significantly from the true population parameters. In other words, if only one actual sample is chosen, how likely will it reflect the universe? To answer this question, we first estimated the mean for each variable from each owner and non-owner sub-sample and then used the variance (derived from the sampling distribution of the mean for that variable) to build a 95% confidence interval around this mean. We then examined whether the population mean  $\mu$  was within this confidence interval. The results for the individual variables are shown in Table A.3 as ‘error rate percentage,’ which tabulates the percentage of the simulated samples out of 1,000 with an estimated mean for that variable that is statistically significantly different from the mean of the universe of respondents at the 5% significance level. In Table 2 (in the row called ‘average error rate per variable’), the ‘error rate percentage’ is summed for all the variables and then averaged.<sup>14</sup> The ‘error rate percentage’ measure penalizes *bias* but favors *variance*, so it is not nearly as good a comparison metric as *mse* when the goal is to assess the sampling distribution. However, when logistics dictates one to draw one and only one sample to examine the statistical estimate of some variables from this sample, the *mse* is not very useful at answering questions such as “what is the average age of entrepreneurs, and how likely is the mean age estimated from this sample correct?” The error rate percentage gives an indication of how likely the 95% confidence interval of the mean age estimated using this one sample will indeed contain the true population mean age.

The *average error rate per variable* in Table 2 suggests that among the three methods, PSM still fares the best (with and without revisits, and with and without accurate weights), but ESM and SSM produce estimates similar to PSM as long as revisits are done. ESMre and SSMre give 12–14% error while PSMre gives 10% error. When proper weights are applied on top of revisits, all three methods’ estimates improve further.

The three methods that are most commonly deployed in the field are ESMnre (the traditional EPI method), SSMre (used in some World Bank Enterprise Surveys), and PSMrp (the gold standard in survey sampling). The *average mse* and the *average error rate per variable* suggest that ESMnre is the worst. SSMre greatly improves upon ESMnre in *bias*, *mse*, and *error rate*. PSMrp has the least *bias*, the smallest *mse*, and the fewest number of incorrect estimates. Still, researchers, statisticians, and field study project managers should ask whether SSMre is acceptable and whether ESMnre can be salvaged with revisits to make

<sup>14</sup>Although one could attach importance weights to the different characteristics to obtain a *weighted* average error rate per variable, we did not give differential weights to the various characteristics, which depend on the situation and goals of the analysis.

ESMre a viable alternative sampling method. We turn to a cost effectiveness analysis to assess these methods to help shed light on these questions.

### 3.3 Comparison of methods by cost, cost effectiveness, and cost efficiency

While traditional EPI sampling (ESMnre) produces estimates that are the least representative of the universe and traditional probability sampling (which includes revisits and proper weights) produces the most representative statistics, an issue at hand is whether the methods differ much by cost and cost-effectiveness. To answer this question, we calculated the approximate costs for each of the twelve sampling scenarios. Based on (i) the larger fieldwork cost items (labor and transportation), (ii) the number of interviews conducted each day (calculated separately for first visits and revisits), and (iii) the number of days in each EA for all 22 EAs, we calculated an average marginal cost per first-visit interview (R74, with \$1 = R8 at the time of the fieldwork) and per revisit interview (R103).<sup>15</sup> Because our interviews were conducted immediately after we had determined and found the eligible respondent, our data contained the combined listing-plus-interview costs rather than separated by listing vs. interview costs. In our fieldwork, the listing-plus-interview took on average 40 minutes, and this included listing, informed consent, and the questionnaire survey. If a mere listing of the household were done without the survey, the time needed would have been much shorter. To calculate the cost of the listing alone, we assumed four scenarios whereby the listing cost was the same, one-half, one-quarter, or one-tenth of the listing-plus-interview cost.

Panel A of Table 3 shows the type of encounter with owner and non-owner households -- whether it is a listing-with-first-visit interview, a listing-with-revisit interview, or a listing alone. The columns tabulate, by sampling method, the number of each type of encounter that needs to be completed in an average EA. The proportion of these encounters reflects our actual fieldwork. <sup>16</sup> Several points are worth mentioning. Every EA has interviews with 10 owners and 10 non-owners; however, to find 10 owners to interview, more than 10 non-owners must be listed to determine their business ownership status. For instance, to fill the quota of 10 owners in the sample with first visit respondents using EA-only weights (columns 1 and 2), a total of 15 non-owners on average must be listed – with 10 interviewed and 5 listed. If the sampling method uses PSM (columns 3, 6, 9, 12), then the whole EA must be listed (using either the first visit sampling frame or the first visit and revisit sampling frame) before random selection of respondents can be performed. Revisit costs are, thus, used for all PSM interviews.<sup>17</sup> For sampling methods applying probability weights, the whole EA must also be listed.<sup>18</sup> For instance, sampling methods that apply ‘non-revisit Probability’ weights (columns 4, 5, 6) must perform a listing of the universe of first visit respondents (25 owners and 37 non-owners which includes the 10 owners and 10 non-owners interviewed). In sampling methods with first visit and revisit respondents (columns 7, 8, 10, 11), half the owner interviews are done on first visits and the other half are done on revisits, but for non-owners, who are less likely to be present during a first visit than owners, 4 interviews are conducted during a first visit, and 6 during a revisit. In ESM and SSM samples with first visits and revisits and with EA-only weights (columns 7 and 8), the model assumes a listing of two extra revisit owners and two extra revisit non-owners as standby interviews in the event that some appointments made with the original batch of revisit

<sup>15</sup>Fixed costs that would have been incurred with any of the methods (such as training and PDA equipment and programming costs) were not allocated to the per-interview costs.

<sup>16</sup>Table 1 contains the total number of first visit and revisit interviews, by owners and non-owners, for our universe.

<sup>17</sup>We do not include PSM’s costs of more field time to do the randomized selection and to return to find the randomly chosen respondents.

<sup>18</sup>We assume that EA-weights can be obtained from statistical agencies, and only probability weights must be obtained through a complete listing of each EA.

respondents are not fulfilled. This results in a total of 7 revisit owners and 8 revisit non-owners that must be listed (with 5 of these owners and 6 of these non-owners also receiving interviews). To reach 7 owners during a revisit, the field team will also reach – in an average EA – 7 first visit owners, 10 first visit non-owners, and 15 revisit non-owners.

Panel B of Table 3 tabulates the total cost of each method for our universe of 22 EAs, with a sensitivity analysis of the cost based on how much listing alone would cost relative to the cost of an actual listing plus interview. As the cost of the listing goes down, the actual number of encounters per EA does not change, but the total cost drops precipitously, especially for the listing-intensive methods, such as PSM. Sampling with revisits is invariably more costly than without revisits because it includes the costs of returning to the EA to interview revisit respondents and the costs of listing standby respondents.

Using these cost figures, we could conduct cost-effectiveness analysis, to examine whether PSM would be worth its cost, and whether SSM and ESM or their modifications might also be cost-effective. The problem, however, is finding the proper effectiveness measure. While the best effectiveness measure would be the impact of some policy variable based on these results, we do not have such information in our dataset. Therefore, we use two types of cost-effectiveness measures. The first is the cost per sample means out of the 1,000 simulated samples that are estimated without ‘error’ (i.e., one minus the *average error rate per variable* in Table 2). The second is the cost efficiency ratio between a particular sampling method and PSMrp (Table A.1; Yoon et al., 1997; Kish, 1965, pages 25 and 266).

Panel C of Table 3 lists the cost per *percent error free estimate* and Panel D lists the cost efficiency using *average mse ratios* as the loss function. The two most cost effective methods under each cost scenario are highlighted in boxes. The ranking depends on the cost of the listing relative to the interview cost. It is evident that ESMre and SSMre outperform the other methods with both cost effectiveness measures -- except when the listing cost is high relative to the interview cost, in which case ESMnre and SSMnre become more cost effective. For surveys that are long (and thus costly), ESMre and SSMre may be the preferred method in terms of cost-effectiveness. However, for short surveys, such as the EPI immunization studies that only care about vaccine coverage in the community, ESMnre and SSMnre may be acceptable given their better cost effectiveness, despite the higher bias and lower efficiency. When the listing cost is truly very low compared to the interview cost, PSMrp, which has the lowest error percentage and the lowest *mse*, becomes more competitive. ESMrp and SSMrp in practice will never be implemented because the cost incurred to find the probability weights already pays for the data necessary to do PSMrp, which has the best effectiveness of the three methods that use probability weights.

### 3.3 Comparison of methods by potential policy variables

In any study, because only one sample is drawn in practice, an important question is which sampling method would most likely produce a sample with findings that authentically reflect findings using the population universe. To analyze this question, we built two regression models, estimated using both the population universe and our simulated samples. The first model was an ordinary least squares regression estimating the effect of owner demographic characteristics on business profit (Table 4a). The second model was a logistic regression, examining the effect of owner characteristics on ownership of a business with vs. without employees (Table 4b).<sup>19</sup> The left-hand column in both tables shows the independent variables and, in parentheses, the variables’ beta coefficients from the universe regression. The p-values of significant coefficients are also shown. As before, columns 1 to 12 show the

<sup>19</sup>These two regressions with their simplistic specifications are included only for illustration.

different simulated samples, and the sampling statistics for each variable under each sampling method. Here, we focus on the samples that would be implemented in practice: ESMnre (the usual EPI method), SSMre (a common World Bank Enterprise Survey method), and PSMrp (the statistical gold standard). We also include ESMre to examine whether revisits added to the EPI method could improve estimates. The statistics shown in these tables are similar to those shown in Table A.3, with one main addition, which is the '*correct significance and sign*.' If the variable is significant in the universe regression, *correct significance and sign* shows the proportion of the time out of 1,000 simulations that the variable is also significant and of the same sign in the sample. If the variable is not significant in the universe regression, *correct significance and sign* shows the proportion of the time that the coefficient estimated using the simulated sample is not significant.

In the business profit regression (Table 4a), three demographic characteristics are statistically significant in the universe regression: gender (female), physical health (PCS12), and mental health (MCS12). Women owned businesses make less profit on average (379 rand less per month) than those owned by men, and owners in better health have businesses with higher profits. For the variables that are significant in the universe, ESMnre and (to a lesser extent) SSMre are more likely than PSMrp to produce samples that give a beta coefficient that is significant and of the same sign as the universe beta. ESMnre correctly flags female as negative and significant 54.8% of the time compared to 44.7% of the time for SSMre and 26.9% of the time for PSMrp. However, the beta coefficient in ESMnre and SSMre is estimated with greater bias than that estimated by PSMrp samples. For instance, the coefficient for female estimated by ESMnre is  $-618.2$  instead of the  $-378.5$  estimated by the universe. On the other hand, PSMrp, because of its higher variance, gives a much more conservative value for the correct significance and sign but has very little bias. This suggests that if one is interested in significance level alone, then ESMnre may be sufficient, but that PSMrp must be used to estimate the elasticity of the effect. For example, if policy were based on the single sample drawn, one would more likely (correctly) find a negative relationship between female gender and profit when ESMnre instead of PSMrp was the sampling method. However, any program to reduce such gender-based profit disparity, once implemented, might show disappointing effects, because the magnitude of the gender gap was overestimated by ESMnre in the first place.

There is one important qualification to the above statement: because ESMnre has lower variance, it is more likely to falsely flag insignificant variables as significant. For instance, the married variable is significant in 80% of ESMnre samples, although it is not significant in the universe sample. The conclusion using ESMnre would be a faulty one: businesses owned by married individuals have lower profit (while in fact there is no such effect). It would be unfortunate if lending institutions decided not to lend to married individuals based on the results obtained from ESMnre samples.

The findings from Table 4b are similar to those from Table 4a: ESMnre and SSMre have higher *correct significance and sign* than PSMrp in most or all of the variables that are significant in the universe regression, but some coefficients may be estimated with high bias by ESMnre and SSMre. For instance, the effect of high risk preference on owning a multi-person business is underestimated by ESMnre, while the effect of female gender on decreased likelihood of multi-person business ownership is overestimated by SSMre. PSMrp coefficients generally have the least bias.

Table 4c summarizes these sampling statistics, averaging separately significant and insignificant variables in the universe regression. ESMre and SSMre strike a middle ground between ESMnre and PSMrp. ESMre and SSMre have fewer under- and over-identifications of relationships than ESMnre. They are also superior to PSMrp in not under-identifying

significant relationships. However, their elasticity measure is far more biased than PSMrp and can be more biased than ESMnre.

Our results from this policy analysis suggest that the relative efficacy of a method vis-à-vis another depends on how much a researcher is worried about (i) correctly measuring the elasticity of an effect (which depends on the bias of the estimate, with PSMrp having the lowest bias), (ii) under-identifying a relationship that is significant in the universe (which happens more often when a method has a large variance and when the 95% confidence interval is more likely to include zero, but also depends on whether the bias takes the estimate further away from or closer to zero), and (iii) over-identifying a relationship that is insignificant in the universe and mistaking it for one that is significant (which is less likely to occur in the more conservative methods that have higher variance like PSMrp).

#### 4. Discussion

This paper presents the results of a comparison between sampling methods that can be used to find eligible respondents to conduct interviews to study MSEs when existing sampling frames are lacking. We used computer simulations to draw samples from a census universe, by implementing ESM, SSM, and PSM protocols, and examined the resulting sampling distributions. Our analysis reveals that these different sampling methods produce samples that differ in their sampling statistics of relevant observable variables. Our analysis allowed for the comparison of the four main viable methods, ESM with and without revisits using EA-weights, SSM with revisits using EA-weights, and PSM with revisits using probability weights. We tested specifically for the effects of proximity selection, at-home selection, and the use of inaccurate weights on the sampling statistics. This allowed us to examine why these sampling strategies have differential performance.

Our most important finding is that revisits are very important in reducing the *bias* and the *mse* of the sampling statistics. Eligible respondents who are present during the first visit may be systematically different from those who are away. Drawing a sample only among respondents who happen to be available when the interviewer shows up at the door, thus, may potentially lead to biased conclusions. Another finding is that proximity selection – a major concern in the EPI sampling methodology literature – does not seem to affect the sampling statistics in our data. While we cannot rule out that this is an artifact of our data and context, this lack of an effect from proximity selection is consistent with many studies in the literature. Our finding regarding the use of proper weights suggests that while this further improves upon the sample properties for all three sampling methods, the degree of improvement is much smaller than that which results from revisits. The added benefit of trying to find and apply accurate probability weights may not justify the costs.

Consistent with the above, the results from two types of cost effectiveness analysis suggest that SSM and ESM – both with revisits and using EA-weights - rank as the two top methods when the listing cost is not too high. Our regression analysis of the policy variables adds support to the validity of these sampling methods. Nevertheless, our cost-effectiveness analysis must be interpreted with caution. Without a clear cut effectiveness measure to use, we have based our cost effectiveness analysis on both the percentage of error free estimates and the *mse*, although the rankings of methods are similar with either effectiveness measure. Because cost effectiveness measures do not institute a lower bound effectiveness measure but focuses on overall cost effectiveness, a decision should not be driven solely by the sampling method's cost effectiveness. For instance, if the elasticity of a particular estimate is paramount, then one should not consider any method other than PSMrp. On the other hand, if a survey is very short (and hence the interview cost is similar to that of the listing of a



single respondent), then ESM<sub>re</sub> is a viable alternative that should be considered because it is greatly cost effective.

Our paper suffers from other limitations. First, our universe sample is incomplete. Because our listing excludes those individuals that are difficult to find even after repeat revisits, we believe that our findings underestimate the magnitude of at-home selection. Second, all simulations were run with 10 owners and 10 non-owners per EA. While this simplified our analysis, we may have been able to achieve greater statistical representativeness by optimizing the number of business-owners and non-owners selected from each site. Finally, we cannot ultimately say to what extent our findings are driven by our particular sample. We advise researchers to consider not our specific findings, but rather the general findings regarding proximity selection, at-home selection, and the effect of using inaccurate weights.

To those researchers interested in understanding the differences between SSM and ESM with revisits and EA-weights, we offer a couple of concluding thoughts on these two methods. SSM starts out with some desired sample size, slices the EA into segments to meet that sample size, and then surveys a randomly selected segment. ESM, however, starts with a quota and a random point, continues the survey until the quota is met, thus covering an area of the EA that is irregularly shaped. The advantage of SSM is that clear segments are demarcated, which can help with longitudinal studies in which return visits are planned.

This also allows for discovering new business entries within the boundaries of the segment. However, the advantages of starting with a quota that must be met before the interviewer stops are also important. On one hand, a quota helps to control costs because if a segment is drawn too large, the additional interviews are not required. On the other hand, if a segment is drawn too small – or simply – if the target respondent group is hard to find, ESM allows one to define the area that covers the desired population. Hence, ESM is useful especially when information about the target respondent group is not readily available. Suppose the enumeration team could start with a random point and essentially survey respondents around this point by meandering in an ever-enlarging spiral as required until the quota is met. This method has at its heart the ESM strategy of meeting a quota for some specific respondent type, but also uses the SSM strategy of having a compact segment that can be easily returned to on follow-up visits. While the SSM samples in this paper were simulated using this setup, we leave it to future work to implement this method in the field.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

This project was supported by the U.S. National Institutes of Health R01-HD-051468, (Chao P.I.). The authors would like to thank Mark Pauly, Khangelani Zuma, Adlai Davids, Gina Weir-Smith, Tholang Mokhele, Monica Costa Dias, Pedro Campos, Phil Anglewicz, Joanne Levy, and the participants and organizers of the World Bank Measurement Conference. We also would like to thank the editors of this special issue and the reviewers of this paper for their helpful comments and suggestions that greatly improved the paper.

### References

- Bennett S, Radalowicz A, Vella V, Tomkins A. A computer simulation of household sampling schemes for health surveys in developing countries. *International Journal of Epidemiology*. 1994; 23(6):1282–1291. [PubMed: 7721532]
- Brogan D, Flagg EW, Deming M, Waldman R. Increasing the accuracy of the Expanded Programme on Immunization's cluster survey design. *Annals of Epidemiology*. 1994; 4(4):302–311. [PubMed: 7921320]

- Chao L-W, Pauly MV, Szrek H, Sousa-Pereira N, Bundred F, Cross C, Gow J. Poor Health kills small business: illness and microenterprises in South Africa. *Health Affairs*. 2007; 26(2):474–482. [PubMed: 17339676]
- de Mel S, McKenzie D, Woodruff C. Returns to capital in microenterprises: evidence from a field experiment. *The Quarterly Journal of Economics*. 2008; 123(4):1329–1372.
- de Mel S, McKenzie D, Woodruff C. Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*. 2009; 88(1):19–31.
- Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG. Individual risk attitudes: New evidence from a large, representative experimentally-validated survey. *DIW Berlin Discussion Papers* 511. 2005 Sep.
- Fafchamps M, McKenzie D, Quinn S, Woodruff C. Using PDA consistency checks to increase the precision of profits and sales measurement in panels. *Journal of Development Economics*. (forthcoming).
- Gelb A, Mengistae T, Ramachandran V, Shah MK. To formalize or not to formalize? Comparisons of microenterprise data from southern and eastern Africa. Center for Global Development Working Paper Number 175. 2009 Jul.
- Grais RF, Rose AMC, Guthmann J-P. Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. *Emerging Themes in Epidemiology*. 2007
- Henderson RH, Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bulletin of the World Health Organization*. 1982; 60(2):253–260. [PubMed: 6980735]
- Jain AK. Nepal 2009 Enterprise Survey: Survey description and technical report. 2010
- Kish L. A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*. 1949; 44:380–387.
- Kish, L. *Survey Sampling*. New York: John Wiley & Sons, Inc; 1965.
- Lemeshow S, Tserkovnyi AG, Tulloch JL, Dowd JE, Lwanga SK, Keja J. A computer simulation of the EPI survey strategy. *International Journal of Epidemiology*. 1985; 14(3):473–481. [PubMed: 4055214]
- Liedholm, C.; Mead, DC. *Small enterprises and economic development: The dynamics of micro and small enterprises*. London: Routledge; 1999.
- Luman ET, Worku A, Berhane Y, Martin R, Cairns L. Comparison of two survey methodologies to assess vaccination coverage. *International Journal of Epidemiology*. 2007
- McKenzie D, Sakho YS. Does it pay firms to register for taxes? The impact of formality on firm profitability. *Journal of Development Economics*. 2010; 91:15–24.
- McPherson MA. Growth of micro and small enterprises in southern Africa. *Journal of Development Economics*. 1996; 48:253–277.
- Milligan P, Njie A, Bennett S. Comparison of two cluster sampling methods for health surveys in developing countries. *International Journal of Epidemiology*. 2004; 33:469–476. [PubMed: 15020569]
- Parker JC, Dondo A. Kenya: Kibera's small enterprise sector: Baseline survey report. GEMINI working paper no. 17. 1991 Apr.
- Rose AMC, Grais RF, Coulombier D, Ritter H. A comparison of cluster and systematic sampling methods for measuring crude mortality. *Bulletin of the World Health Organization*. 2006; 84(4): 290–296. [PubMed: 16628302]
- South Africa Business Guidebook 2004/2005, 9th Edition: 241–243; 2005/2006, 10th Edition: 216–218. WriteStuff Publishing;
- Turner AG, Magnani RJ, Shuaib M. A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) cluster survey design. *International Journal of Epidemiology*. 1996; 25(1):198–203. [PubMed: 8666490]
- Ware, JE.; Kosinski, M.; Keller, SD. *Health Institute*. 2d ed.. Boston: New England Medical Center; 1995. How to Score the SF-12 Physical and Mental Health Summary Scale.

Yoon SS, Katz J, Brendel K, West KP Jr. Efficiency of EPI cluster sampling for assessing diarrhea and dysentery prevalence. *Bulletin of the World Health Organization*. 1997; 75(5):417–426. [PubMed: 9447775]

**Table 1**

Mean characteristics of non-owners and business owners, for all respondents and by first visit and revisit

Respondent Characteristics	1		2		3		4		5		6	
	Non-owners		Owners		Non-owners		Non-owners		Owners		Owners	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	39.95	15.49	43.33	14.09**	41.16	17.90	39.06	13.58**	44.77	15.04	41.82	12.92**
Female	0.60	0.49	0.56	0.50*	0.66	0.49	0.56	0.49**	0.61	0.49	0.50	0.49**
Married	0.31	0.46	0.41	0.49**	0.27	0.46	0.34	0.47**	0.37	0.49	0.46	0.49**
Primary school incomplete	0.16	0.36	0.23	0.42**	0.18	0.40	0.14	0.34**	0.27	0.45	0.19	0.39**
High risk preference	0.37	0.48	0.45	0.50**	0.33	0.48	0.40	0.48**	0.40	0.50	0.51	0.49**
PCS12	50.40	8.98	49.52	9.55*	48.99	10.47	51.42	7.67**	48.97	10.44	50.10	8.54
MCS12	46.00	9.97	46.82	9.97*	45.39	10.26	46.44	9.75*	46.01	10.07	47.67	9.80**
Industry (Manufacturing)			0.10	0.30					0.08	0.28	0.12	0.33*
Industry (Retail)			0.54	0.50					0.58	0.50	0.50	0.49*
Industry (Services)			0.24	0.43					0.25	0.44	0.23	0.42
Age of business			5.56	7.39					5.98	8.00	5.11	6.69
Business Income (categories)			4.39	3.06					4.28	3.09	4.50	3.03
Business profits (x 1000)			0.65	2.30					0.59	2.61	0.72	1.95
Multiperson business			0.26	0.44					0.21	0.41	0.32	0.46**

\* p < .05

\*\* p < .01

Asterisks in column 2 denote the significance of difference between owners and non-owners

Asterisks in columns 4 and 6 denote the significance of difference between first visit and revisit respondents, separately for non-owners and owners

**Table 2**  
Bias, design effect, mse ratio, and error rate -- averaged over all variables, by sampling method

	1	2	3	4	5	6	7	8	9	10	11	12
	Sampling Methods											
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMrp	SSMrp	PSMrp
Sociodemographic Variables (non-owners)												
average absolute relative bias per variable	0.073	0.069	0.062	0.096	0.085	0.081	0.041	0.025	0.020	0.016	0.006	0.002
average design effect per variable	0.490	0.513	0.551	0.938	0.951	1.004	0.718	0.743	0.723	0.961	1.020	1.000
average mse ratio per variable	2.481	2.021	1.744	3.619	2.754	2.690	1.092	0.935	0.958	1.052	1.027	1.000
average error rate per variable	0.481	0.386	0.286	0.394	0.296	0.249	0.111	0.085	0.088	0.060	0.054	0.054
Sociodemographic Variables (owners)												
average absolute relative bias per variable	0.046	0.053	0.051	0.072	0.077	0.074	0.033	0.035	0.028	0.017	0.010	0.002
average design effect per variable	0.488	0.493	0.563	0.920	1.037	1.038	0.636	0.686	0.709	0.966	1.007	1.000
average mse ratio per variable	1.993	2.014	2.224	2.436	2.761	2.800	0.965	1.071	0.979	1.152	1.075	1.000
average error rate per variable	0.399	0.424	0.400	0.260	0.273	0.265	0.112	0.115	0.097	0.070	0.057	0.050
Business Variables												
average absolute relative bias per variable	0.112	0.123	0.109	0.105	0.097	0.099	0.113	0.092	0.072	0.047	0.019	0.005
average design effect per variable	0.436	0.470	0.582	0.801	0.829	0.926	0.767	0.818	0.817	0.997	1.060	1.000
average mse ratio per variable	1.362	1.601	1.626	1.634	1.715	1.820	1.454	1.367	1.102	1.192	1.089	1.000
average error rate per variable	0.325	0.366	0.302	0.175	0.182	0.175	0.203	0.162	0.112	0.071	0.054	0.046
Total averages												
average absolute relative bias per variable	0.077	0.081	0.074	0.091	0.086	0.085	0.062	0.051	0.040	0.027	0.012	0.003
average design effect per variable	0.471	0.492	0.565	0.886	0.939	0.989	0.707	0.749	0.750	0.975	1.029	1.000
average mse ratio per variable	1.945	1.879	1.865	2.563	2.410	2.437	1.170	1.124	1.013	1.132	1.063	1.000
average error rate per variable	0.402	0.392	0.329	0.276	0.250	0.230	0.142	0.121	0.099	0.067	0.055	0.050



Table 3

Cost and cost-effectiveness of sampling methods

	Sampling Methods											
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMnr	SSMnr	PSMnr	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMrp	SSMrp	PSMrp
	1	2	3	4	5	6	7	8	9	10	11	12
A. Type of encounter with respondent for an average EA												
Listing & interview of owners reached on first visit	10	10	0	10	10	0	5	5	0	5	5	0
Listing & interview of owners reached on revisit	0	0	10	0	0	10	5	5	10	5	5	10
Listing & interview of non-owners reached on first visit	10	10	0	10	10	0	4	4	0	4	4	0
Listing & interview of non-owners reached on revisit	0	0	10	0	0	10	6	6	10	6	6	10
Listing of owners reached on first visit	0	0	15	15	15	15	2	2	20	20	20	20
Listing of owners reached on revisit	0	0	0	0	0	0	2	2	20	20	20	20
Listing of non-owners reached on first visit	5	5	27	27	27	27	6	6	33	33	33	33
Listing of non-owners reached on revisit	0	0	0	0	0	0	9	9	49	49	49	49
Total number of stands reached in an average EA	25	25	62	62	62	62	39	39	142	142	142	142
B. Cost incurred (in South African rand) for all 22 EAs combined												
Cost (listing cost = 1.00 interview cost)	R 40,416	40,416	113,012	100,305	100,305	113,012	70,308	70,308	242,546	236,808	236,808	242,546
Cost (listing cost = 0.50 interview cost)	R 36,446	36,446	79,097	66,391	66,391	79,097	54,882	54,882	143,864	138,126	138,126	143,864
Cost (listing cost = 0.25 interview cost)	R 34,461	34,461	62,140	49,433	49,433	62,140	47,169	47,169	94,523	88,785	88,785	94,523
Cost (listing cost = 0.10 interview cost)	R 33,270	33,270	51,966	39,259	39,259	51,966	42,541	42,541	64,919	59,181	59,181	64,919
C. Cost effectiveness as cost per percent of error free estimates												
Average error rate per variable	<u>0.402</u>	<u>0.392</u>	0.329	0.276	0.250	0.230	0.142	0.121	0.099	0.067	0.055	0.050
Cost effectiveness (listing cost = 1.00 interview cost)	<u>R 67,542</u>	<u>66,505</u>	168,531	138,589	133,775	146,715	81,953	79,943	269,154	253,904	250,578	255,299
Cost effectiveness (listing cost = 0.50 interview cost)	<u>R 60,907</u>	<u>59,972</u>	117,955	91,730	88,543	102,686	63,972	62,403	159,646	148,098	146,158	151,428
Cost effectiveness (listing cost = 0.25 interview cost)	<u>R 57,590</u>	<u>56,706</u>	92,667	68,300	65,928	80,671	54,982	53,633	104,893	95,195	93,948	99,493
Cost effectiveness (listing cost = 0.10 interview cost)	<u>R 55,600</u>	<u>54,746</u>	77,494	54,242	52,358	67,463	49,588	48,371	72,041	63,453	62,622	68,332
D. Cost efficiency using mse as loss function												

	1	2	3	4	5	6	7	8	9	10	11	12
Sampling Methods												
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp
Average mse ratio per variable	1.945	1.879	1.865	2.563	2.410	2.437	1.170	1.124	1.013	1.132	1.063	1.000
Cost efficiency (listing cost = 1.00 interview cost)	0.324	0.313	0.869	1.060	0.997	1.135	0.339	0.326	1.013	1.105	1.038	1.000
Cost efficiency (listing cost = 0.50 interview cost)	0.493	0.476	1.025	1.183	1.112	1.340	0.447	0.429	1.013	1.087	1.021	1.000
Cost efficiency (listing cost = 0.25 interview cost)	0.709	0.685	1.226	1.340	1.260	1.602	0.584	0.561	1.013	1.063	0.999	1.000
Cost efficiency (listing cost = 0.10 interview cost)	0.997	0.963	1.493	1.550	1.457	1.951	0.767	0.737	1.013	1.032	0.969	1.000

Table 4

a: OLS regression with business profit as dependent variable Sampling statistics: bias, design effect, mean square error, and count of significant coefficient estimates, by sampling method

Explanatory Variables (population coefficient)	Sampling Methods											
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMmrp	SSMmrp	PSMmrp
Age ( $\beta = -0.988$ )												
bias	-2.868	-2.675	-0.895	-1.016	0.666	2.396	-4.372	-1.492	-4.232	0.061	2.713	0.164
absolute relative bias	2.903	2.708	0.906	1.029	0.674	2.425	4.426	1.510	4.285	0.062	2.746	0.166
design effect	0.439	0.676	0.721	0.579	1.201	1.102	0.816	0.940	0.710	1.174	1.473	1.000
mse relative to PSMrp's mse	0.522	0.748	0.728	0.589	1.205	1.160	1.008	0.962	0.890	1.174	1.547	1.000
correct significance & sign	0.945	0.964	0.955	0.968	0.950	0.930	0.915	0.949	0.900	0.934	0.941	0.938
Female ( $\beta = -378.539$ ; $p < .01$ )												
bias	-239.671	-249.856	-208.784	-69.778	-127.501	-95.431	-194.602	-204.043	-81.413	-56.091	-91.676	5.798
absolute relative bias	0.633	0.660	0.552	0.184	0.337	0.252	0.514	0.539	0.215	0.148	0.242	0.015
design effect	1.403	1.446	0.999	1.284	1.438	1.097	1.401	1.364	1.310	1.072	1.041	1.000
mse relative to PSMrp's mse	2.180	2.291	1.589	1.349	1.658	1.219	1.913	1.928	1.399	1.115	1.155	1.000
correct significance & sign	0.548	0.470	0.580	0.361	0.374	0.390	0.415	0.447	0.329	0.312	0.389	0.269
Married ( $\beta = 148.514$ )												
bias	-571.251	-403.824	-377.988	-332.252	-130.163	-133.036	-58.138	-26.912	-42.216	21.429	53.191	14.243
absolute relative bias	3.846	2.719	2.545	2.237	0.876	0.896	0.391	0.181	0.284	0.144	0.358	0.096
design effect	0.290	0.542	0.629	0.476	1.234	0.927	0.903	0.984	1.098	0.794	1.057	1.000
mse relative to PSMrp's mse	4.481	2.635	2.463	1.893	1.448	1.152	0.944	0.991	1.118	0.798	1.091	1.000
correct significance & sign	0.200	0.771	0.817	0.840	0.948	0.943	0.943	0.922	0.933	0.897	0.888	0.901
Education<Primary ( $\beta = -56.641$ )												
bias	-152.090	-113.146	-139.291	-148.352	-164.833	-177.662	-103.187	-171.100	-10.417	-8.478	-84.757	19.314
absolute relative bias	2.685	1.998	2.459	2.619	2.910	3.137	1.822	3.021	0.184	0.150	1.496	0.341
design effect	0.181	0.325	0.237	0.320	0.753	0.477	0.969	0.810	0.865	1.126	1.067	1.000
mse relative to PSMrp's mse	0.362	0.424	0.389	0.492	0.964	0.723	1.049	1.038	0.864	1.124	1.121	1.000
correct significance & sign	0.722	0.852	0.796	0.828	0.859	0.838	0.961	0.930	0.951	0.962	0.943	0.954

a: OLS regression with business profit as dependent variable Sampling statistics: bias, design effect, mean square error, and count of significant coefficient estimates, by sampling method

Explanatory Variables (population coefficient)	Sampling Methods											
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp
High Risk Preference ( $\beta = 126.576$ )												
bias	127.142	15.977	167.431	24.714	-81.978	52.119	12.344	-34.082	33.249	1.805	-39.564	3.237
absolute relative bias	1.004	0.126	1.323	0.195	0.648	0.412	0.098	0.269	0.263	0.014	0.313	0.026
design effect	0.974	1.058	0.880	0.993	1.121	1.092	1.209	1.178	1.078	1.063	1.015	1.000
mse relative to PSMrp's mse	1.158	1.061	1.199	1.000	1.197	1.123	1.210	1.191	1.091	1.063	1.032	1.000
correct significance & sign	0.821	0.936	0.800	0.942	0.972	0.916	0.924	0.958	0.924	0.925	0.947	0.923
PCS12 ( $\beta = 11.584$ ; $p < .01$ )												
bias	-2.079	-0.202	-1.651	-2.883	1.696	0.307	6.316	7.391	2.463	1.691	4.251	-0.072
absolute relative bias	0.179	0.017	0.143	0.249	0.146	0.027	0.545	0.638	0.213	0.146	0.367	0.006
design effect	0.238	0.408	0.312	0.448	0.907	0.581	0.574	0.763	0.834	0.706	1.201	1.000
mse relative to PSMrp's mse	0.288	0.409	0.343	0.543	0.940	0.582	1.031	1.389	0.903	0.738	1.408	1.000
correct significance & sign	0.509	0.405	0.444	0.278	0.304	0.355	0.709	0.627	0.365	0.368	0.296	0.224
MCS12 ( $\beta = 11.137$ ; $p < .05$ )												
bias	2.708	4.503	6.600	-2.900	1.563	2.535	1.335	-3.360	-2.020	3.470	-0.222	0.505
absolute relative bias	0.243	0.404	0.593	0.260	0.140	0.228	0.120	0.302	0.181	0.312	0.020	0.045
design effect	0.301	0.468	0.476	0.527	0.916	1.136	0.783	0.865	0.828	0.958	1.032	1.000
mse relative to PSMrp's mse	0.346	0.592	0.744	0.578	0.930	1.174	0.793	0.933	0.852	1.031	1.031	1.000
correct significance & sign	0.491	0.392	0.435	0.122	0.182	0.189	0.197	0.108	0.126	0.196	0.154	0.143
Sample Size (universe N = 1088)	208	208	208	208	208	208	220	220	220	220	220	220



**b. Logistic regression with multi- vs. single-person business as dependent variable Sampling statistics: bias, design effect, mean square error, and count of significant coefficient estimates, by sampling method**

Explanatory Variables (population coefficient)	Sampling Methods																								
	1		2		3		4		5		6		7		8		9		10		11		12		
	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	
bias	-0.245	-0.443	-0.307	-0.287	-0.458	-0.332	0.002	0.021	0.100	0.019	0.019	0.023	0.010	-0.011	0.023	0.043	0.043	0.026	0.053	0.053	0.053	0.053	0.053	0.053	0.053
absolute relative bias	0.564	1.018	0.705	0.660	1.053	0.763	0.004	0.048	0.230	0.043	0.043	0.023	0.010	-0.011	0.023	0.043	0.043	0.026	0.053	0.053	0.053	0.053	0.053	0.053	0.053
design effect	0.461	0.485	0.520	0.829	0.969	1.138	0.559	0.574	0.636	0.898	0.898	0.898	0.898	0.794	0.898	0.898	0.898	0.794	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mse relative to PSMr's mse	0.916	1.977	1.233	1.453	2.563	1.972	0.557	0.575	0.710	0.897	0.897	0.897	0.897	0.792	0.897	0.897	0.897	0.792	1.000	1.000	1.000	1.000	1.000	1.000	1.000
correct significance & sign	0.124	0.017	0.077	0.068	0.022	0.049	0.364	0.372	0.457	0.265	0.265	0.265	0.265	0.255	0.265	0.265	0.265	0.255	0.232	0.232	0.232	0.232	0.232	0.232	0.232
PCS12 (B = 0.019; p < .01)																									
bias	0.006	0.002	-0.008	0.017	0.013	0.001	0.024	0.015	0.007	0.023	0.023	0.023	0.010	0.010	0.023	0.023	0.023	0.010	0.002	0.002	0.002	0.002	0.002	0.002	0.002
absolute relative bias	0.318	0.078	0.391	0.896	0.676	0.028	1.252	0.779	0.349	1.175	1.175	1.175	0.502	0.502	1.175	1.175	1.175	0.502	0.115	0.115	0.115	0.115	0.115	0.115	0.115
design effect	0.410	0.305	0.390	1.042	0.867	0.883	0.467	0.576	0.694	0.836	0.836	0.836	0.764	0.764	0.836	0.836	0.836	0.764	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mse relative to PSMr's mse	0.485	0.307	0.506	1.660	1.216	0.874	1.690	1.046	0.782	1.909	1.909	1.909	0.953	0.953	1.909	1.909	1.909	0.953	1.000	1.000	1.000	1.000	1.000	1.000	1.000
correct significance & sign	0.423	0.408	0.127	0.342	0.349	0.149	0.835	0.539	0.284	0.538	0.538	0.538	0.328	0.328	0.538	0.538	0.538	0.328	0.161	0.161	0.161	0.161	0.161	0.161	0.161
MCS12 (B = 0.008)																									
bias	-0.011	-0.006	-0.011	-0.006	0.003	-0.004	-0.002	-0.008	0.001	0.000	0.000	0.000	-0.005	-0.005	0.000	0.000	0.000	-0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001
absolute relative bias	1.276	0.654	1.296	0.685	0.365	0.524	0.252	0.893	0.107	0.037	0.037	0.037	0.639	0.639	0.037	0.037	0.037	0.639	0.090	0.090	0.090	0.090	0.090	0.090	0.090
design effect	0.598	0.447	0.642	0.995	0.871	1.130	0.463	0.642	0.668	0.693	0.693	0.693	0.948	0.948	0.693	0.693	0.693	0.948	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mse relative to PSMr's mse	0.957	0.541	1.012	1.097	0.899	1.189	0.476	0.817	0.670	0.692	0.692	0.692	1.037	1.037	0.692	0.692	0.692	1.037	1.000	1.000	1.000	1.000	1.000	1.000	1.000
correct significance & sign	0.947	0.937	0.952	0.940	0.891	0.939	0.920	0.947	0.906	0.921	0.921	0.921	0.948	0.948	0.921	0.921	0.921	0.948	0.925	0.925	0.925	0.925	0.925	0.925	0.925
Sample Size (universe N = 1088)	208	208	208	208	208	208	220	220	220	208	208	220	220	220	208	208	220	220	220	220	220	220	220	220	220

**c. Sampling statistics averaged over all explanatory variables: bias, design effect, mean square error, and count of significant coefficient estimates, by sampling method**

Explanatory Variables (population coefficient)	Sampling Methods																								
	1		2		3		4		5		6		7		8		9		10		11		12		
	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	ESMnre	SSMnre	PSMnre	
bias	-0.011	-0.006	-0.011	-0.006	0.003	-0.004	-0.002	-0.008	0.001	0.000	0.000	0.000	-0.005	-0.005	0.000	0.000	0.000	-0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001
absolute relative bias	1.276	0.654	1.296	0.685	0.365	0.524	0.252	0.893	0.107	0.037	0.037	0.037	0.639	0.639	0.037	0.037	0.037	0.639	0.090	0.090	0.090	0.090	0.090	0.090	0.090
design effect	0.598	0.447	0.642	0.995	0.871	1.130	0.463	0.642	0.668	0.693	0.693	0.693	0.948	0.948	0.693	0.693	0.693	0.948	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mse relative to PSMr's mse	0.957	0.541	1.012	1.097	0.899	1.189	0.476	0.817	0.670	0.692	0.692	0.692	1.037	1.037	0.692	0.692	0.692	1.037	1.000	1.000	1.000	1.000	1.000	1.000	1.000
correct significance & sign	0.947	0.937	0.952	0.940	0.891	0.939	0.920	0.947	0.906	0.921	0.921	0.921	0.948	0.948	0.921	0.921	0.921	0.948	0.925	0.925	0.925	0.925	0.925	0.925	0.925
Sample Size (universe N = 1088)	208	208	208	208	208	208	220	220	220	208	208	220	220	220	208	208	220	220	220	220	220	220	220	220	220

Summary sampling statistics for OLS regression coefficient estimates with business profit as dependent variable, by sampling method



c: Sampling statistics averaged over all explanatory variables: bias, design effect, mean square error, and count of significant coefficient estimates, by sampling method

	Sampling Methods											
	Revisit: No Weight: EA-only			Revisit: No Weight: nrProbability			Revisit: Yes Weight: EA-only			Revisit: Yes Weight: rProbability		
	ESMre	SSMre	PSMre	ESMnrp	SSMnrp	PSMnrp	ESMre	SSMre	PSMre	ESMrp	SSMrp	PSMrp
	1	2	3	4	5	6	7	8	9	10	11	12
For variables that are insignificant in the universe regression												
average absolute relative bias per significant variable	2.610	1.888	1.808	1.520	1.277	1.717	1.684	1.245	1.254	0.093	1.228	0.157
average design effect	0.471	0.650	0.617	0.592	1.077	0.899	0.974	0.978	0.938	1.039	1.153	1.000
average mse ratio	1.631	1.217	1.195	0.993	1.204	1.039	1.053	1.045	0.990	1.040	1.198	1.000
average correct estimates per variable	0.672	0.881	0.842	0.895	0.932	0.907	0.936	0.940	0.927	0.930	0.930	0.929
For variables that are significant in the universe regression												
average absolute relative bias per significant variable	0.352	0.361	0.429	0.231	0.208	0.169	0.393	0.493	0.203	0.202	0.210	0.022
average design effect	0.647	0.774	0.595	0.753	1.087	0.938	0.919	0.997	0.990	0.912	1.091	1.000
average mse ratio	0.938	1.097	0.892	0.823	1.176	0.992	1.246	1.417	1.051	0.961	1.198	1.000
average correct estimates per variable	0.516	0.422	0.486	0.254	0.287	0.311	0.440	0.394	0.273	0.292	0.280	0.212
Summary sampling statistics for logistic regression coefficient estimates with probability of multi-person vs. single-person business as dependent variable, by sampling method												
For variables that are insignificant in the universe regression												
average absolute relative bias per significant variable	0.694	0.577	0.548	0.467	0.493	0.557	0.804	0.656	0.512	0.438	0.448	0.206
average design effect	0.529	0.511	0.574	1.218	1.223	1.115	0.578	0.664	0.680	0.952	1.002	1.000
average mse ratio	0.645	0.662	0.683	1.277	1.252	1.168	0.701	0.767	0.719	0.992	1.038	1.000
average correct estimates per variable	0.851	0.795	0.859	0.916	0.891	0.905	0.878	0.873	0.860	0.907	0.908	0.892
For variables that are significant in the universe regression												
average absolute relative bias per variable	0.304	0.383	0.393	0.550	0.590	0.293	0.454	0.357	0.204	0.409	0.245	0.066
average design effect	0.439	0.384	0.501	1.030	0.928	1.084	0.517	0.578	0.659	0.872	0.875	1.000
average mse ratio	0.617	0.885	0.787	1.455	1.577	1.368	0.940	0.819	0.714	1.229	0.998	1.000
average number of samples with correct sign and significance	0.464	0.458	0.338	0.269	0.333	0.242	0.695	0.611	0.499	0.471	0.421	0.315