Considerations for Selection and Release of Reserve Samples for In-Person Surveys

Lin Li[1], Tom Krenzke[1] and Leyla Mohadjer[1]

[1]Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract

Especially in designing the samples for first-time surveys, there is uncertainty about the response rates and eligibility rates. To address this uncertainty, a large sample may be initially selected, then randomly separated into a main sample and a reserve sample. In this cost-efficient approach, a reserve sample will be on hand for release if needed. In multi-stage samples there are options of drawing a reserve sample at different stages. This paper discusses the issues to consider when choosing the sampling stage for reserve samples and discusses options for drawing reserve samples at various stages. We also look at other issues related to the selection and release of reserve samples, such as the timing of selection and release, amount of release, etc.

Key Words: replicate samples, sample design, sample selection, sample monitoring

1. Introduction

Our motivation for looking into reserve samples arose from the Program for the International Assessment of Adult Competencies (PIAAC). Sponsored by the Organisation for Economic Cooperation and Development (OECD), PIAAC is an international literacy survey of non-institutionalized adults ages 16-65. In Cycle 1 of PIAAC, twenty-four countries participated in the first round of data collection that occurred in 2011-2012, and nine other countries are taking part in the second round in 2014. PIAAC sample designs vary across countries ranging from a 1-stage sample of persons from a population registry to a 4-stage sample including primary sampling units (PSUs), secondary sampling units (SSUs), dwelling units (DUs), and persons. As a member of the PIAAC consortium, Westat provides statistical guidance to produce survey estimates that are reliable and comparable across countries.

To meet the analytic goals of PIAAC, each country is required to obtain at least 5,000 completed interviews. The cost per completed case is high because the survey is conducted by face-to-face interview, and the data collection includes both an in-depth questionnaire and a direct assessment of literacy, numeracy, and problem-solving proficiencies. Given the cost, a goal is to tightly control the number of completed interviews. Therefore, the consortium developed general plans for reserve samples to help countries achieve this goal. However, no specific guidance on selecting and releasing reserve samples was initially given to each country. To improve the PIAAC guidelines, we considered various scenarios for selecting and releasing reserve samples under different types of sample designs and survey practices.

2. What are Reserve Samples?

Reserve samples can be used to address the uncertainty about the assumptions for response rates and eligibility rates when selecting a sample to arrive at a target number of completed cases. In first-time surveys, such as in Cycle 1 of PIAAC, there can be considerable uncertainty about these rates. Inaccurate assumptions can lead to either a shortfall or an excess in the sample yield, which either sacrifices the precision of survey outcome or increases the cost of data collection. As Kalton and Anderson (1986) discuss, to address the uncertainty, a large sample may be initially selected, then randomly separated into a main sample and a reserve sample, with the size of the large sample being sufficient to generate the desired number of completed cases based on the lowest estimates of response and eligibility rates. Since reserve samples are essentially replicate samples, care should be taken to ensure that reserve samples represent the target population.

After the reserve sample is selected, it is set aside and released to the field only when needed. If too few cases are released in the early phase of data collection, more need to be released later. This is less efficient operationally, since the interviewers may have less than an optimum number of cases to work in the early phase of data collection but they may have too many cases in the later phase to be carried out in a limited amount of time. Insufficient release may also extend the period of data collection and increase costs. On the other hand, as discussed in Lavrakas (2012), if too many cases are released from the reserve sample, some of the cases may not be fully worked and they are therefore more likely to become nonrespondents of unknown eligibility, which reduces response rates and may cause nonresponse bias. To help tailor the amount of release, a large reserve sample can be split into random subgroups, referred to as release groups (sometimes called replicates). Instead of being released all at the same time, the reserve sample can be released in groups until the target sample size is achieved. Each group should be created to be representative of the population. Therefore the reserve sample needs to be large enough so that each release group represents the target population. For example, under a stratified design, each release group needs to have cases in each and every stratum.

Reserve samples should be distinguished from substitution, which is another approach sometimes used to address sample shortfall. Substitution typically is understood to mean replacing a nonresponding unit with another unit that matches on characteristics known for all cases on the sampling frame. Vehovar (1999) discusses field substitution as it relates to unit non-response. Substitution has the advantage of possibly reducing nonresponse bias by attempting to select similar cases, but it is not probability-based. Unlike substitution, reserve sample is probability-based and not susceptible to the choice by home-office staff or interviewers to get convenient cases.

3. Special Circumstances in PIAAC

Each country conducting a PIAAC survey was encouraged to consider selecting a reserve sample. Table 1 shows the sizes of the reserve samples for PIAAC participating countries in the first two rounds. The reserve sample sizes ranged from 0% (no reserve) to 63% of the main sample size. The variation across countries partly reflects the different confidence levels that countries have in their assumptions for response and eligibility rates, but some variation is due to the country preferences or traditions in the amount of reserve to select.

Table 1: Size of reserve sample as a percentage of the main sample for PIAAC countries

Country	*Size*	*Country*	*Size*	*Country*	*Size*
Australia	0%	Germany	63%	Norway	20%
Austria	50%	Greece	0%	Poland	28%
Belgium	56%	Indonesia	50%	Russian Federation	10%
Canada	10%	Ireland	0%	Singapore	50%
Chile	44%	Israel	10%	Slovak Republic	17%
Cyprus	20%	Italy	38%	Slovenia	13%
Czech Republic	60%	Japan	10%	Spain	50%
Denmark	10%	South Korea	10%	Sweden	20%
Estonia	20%	Lithuania	35%	Turkey	50%
Finland	25%	Netherlands	10%	United Kingdom	20%
France	20%	New Zealand	10%	United States	50%

As discussed by Valliant, Dever, and Kreuter (2013), reserve samples can be formed in different ways with multi-stage designs. In a multi-stage household sample, a reserve sample may be composed of PSUs, SSUs, or DUs. Often a large sample of SSUs or DUs is selected initially and randomly divided into the main sample and the reserve sample. For example, the National Survey on Drug Use and Health (NSDUH) selects a reserve sample of SSUs (Morton et al. 2012) and the National Health and Nutrition Examination Survey (NHANES) has been selecting a reserve sample of DUs since the 1980s (National Center for Health Statistics, 1992, 2014). However, the following different circumstances across countries in PIAAC call for different types of reserves:

1) **Size of PSUs**. The PSUs were counties (median size is about 10,000 DUs) or groups of counties in the US, while some of the PIAAC countries have much smaller PSUs. For example, in South Korea PSUs were Census enumeration districts which contained about 60 DUs each. When the PSUs are large, a relatively small number of them is sampled (80 in the US PIAAC), whereas when the PSUs are small, a much larger number of them are selected (883 for the South Korea PIAAC).
2) **Sampling frame**. A population registry is available for use as a sampling frame in many PIAAC countries. In this case, it is straightforward to draw a reserve sample when a single stage design is used. In multi-stage designs, the reserve samples can be selected from the register in the already sampled or newly sampled PSUs or SSUs.
3) **Organization of the field work**. In most of the PIAAC surveys data collection was carried out in all sampled PSUs simultaneously. However in two of the PIAAC participating countries it was carried out sequentially, from PSU-to-PSU, i.e., interviewers completed the data collection in a PSU before moving to the next.

In PIAAC Round 2, several countries planned to select reserve samples at the PSU or SSU stage. This prompted the consortium to look into the pros and cons of selecting reserve samples at different stages. Hereafter we will discuss the selection and release of reserve samples in one-stage or multi-stage designs in turn.

4. One-Stage Designs

Several PIAAC countries selected a one-stage sample of persons from a population registry with either explicit or implicit stratification (i.e., sorting the frame on auxiliary variables, and selecting the sample through systematic sampling). An issue with reserve samples in stratified (implicit or explicit) designs is that gaps in representation are created when splitting the sample into the main and reserve samples. When the sample design includes stratification, and the reserve sample is selected from a large initial sample, the sampler needs to ensure that there is at least one main sample unit remaining in each of the strata. This is not a serious issue in one-stage designs since reserve samples are large samples of persons, but it becomes an issue with multi-stage designs in which the reserve sample is selected at the PSU level.

The reserve sample can be selected at the beginning with the main sample or during data collection after the main sample is selected. From the sampling perspective, selecting the reserve together with the main sample is easier and less error prone since only one process is involved. However, the sampler will have a better idea about how large the reserve sample should be if the reserve sample is selected later. From the operation perspective, it is more efficient to select the reserve sample at the same time as the main sample, since the reserve sample can be loaded into the survey management system at the same time and will be ready for release whenever needed. However, the cost may be higher if addresses need to be keyed or advance letters need to be mailed.

For the single stage sample, with equal probabilities of selection within strata, it is straightforward to either select the reserve sample with the main sample, or select the main sample first, and then the reserve sample if needed in a sequential manner. Suppose a systematic random sample of 500 units was selected initially from a frame of 10,000 units, and then an additional 100 reserve units were selected with equal probability from the remaining 9,500. Each unit in the combined sample will have the same selection probability of 0.06.

5. Multi-Stage Designs

With multi-stage designs, the reserve sample can be selected at the PSU stage, at the SSU stage (where applicable), or at the DU stage for household samples or persons for samples from registers. If the reserve sample is selected from DUs or persons within SSUs, the issues involved are basically the same as those for a single stage sample. The only extra factor to consider is that the SSUs have to be large enough to support the main and reserve samples. If that condition does not hold, then the reserve sample must include some additional SSUs.

If the reserve sample is to be made up of additional PSUs or SSUs, then the fact that these units are selected with unequal probabilities needs to be taken into account. As discussed in Valliant et al. (2013), if the sample design includes a PPS sample at the stage the reserve is selected (e.g., PSU level), it is not straightforward to select a sequential, or subsequent, reserve sample separately in a way that the overall combined sample is PPS[1].

[1] Cochran (1977) illustrates several approaches to achieve PPS in the overall combined sample.

For this reason, reserve PSUs or SSUs should be selected at the same time as the main sample units as a simple way to achieve their desired selection probabilities.
The applicability of a reserve sample of PSUs depends on the numbers of PSUs selected for the main sample, which in turn depends on the sizes of the PSUs. When a small number of large PSUs is selected, as in the case with the 80 PSUs selected for the US PIAAC, selecting a reserve sample of PSUs is problematic for several reasons.

First, detailed PSU stratification is particularly beneficial with such designs, often with two PSUs selected per explicit stratum and implicit stratification by systematic sampling from an ordered list within the explicit strata. Adding a reserve sample of one PSU per explicit stratum with two main sample PSUs results in a 50 percent increase in the sample. When a reserve sample of other than 50 percent is needed, it can be obtained by using either smaller or larger sampling fractions within the reserve sample PSUs. However, that may lead to inefficiencies in data collection.

Second, selecting the reserve sample together with the main sample may turn some PSUs into certainty selections when they were not certainties for the main sample alone. If the certainty threshold is based on the total sample (including both main and reserve), and if all reserve PSUs are released, then an optimal sample of PSUs was selected. However, if not all of the reserve samples are released, the optimal design would not have included the smallest certainty PSUs with certainty. On the other hand, if the certainty threshold is based on the size of main sample only, the optimal design would have included the largest non-certainty PSUs as certainty if some of the reserve sample PSUs were to be released. A possible solution is to include them with certainty with a set of SSUs selected for the main sample and additional SSUs selected for the reserve sample, but the workload for the main sample is then suboptimal. Another related issue is whether to include certainties in the reserve. For example, certainties could be left out of the reserve selection. If they are included, some certainties in the reserve sample may not get released for data collection. Either way, there is some loss in efficiency – the sample is not optimal once the reserve is released.

Third, with designs of this type, the workload in each sampled PSU is sufficiently large that local interviewers can be recruited and trained for each PSU, with little reliance on travelling interviewers. Staffing the PSUs in a reserve sample, if needed, is then a major task. For all these reasons, we recommend against selecting a reserve sample of PSUs when the main sample consists of a small number of large PSUs.

The situation is different when the PSUs are small and many of them are selected for the main sample, as was the case with the South Korea PIAAC. In that survey the PSUs were census enumeration areas and 883 of them were selected for the main survey. In this case, detailed stratification can be carried out while still retaining a reasonable number of PSUs in each stratum. For example, with 100 strata and 6 sampled PSUs per stratum, the full reserve sample might add another, say, 3 PSUs. Adding one or two reserve PSU per stratum would increase the sample size by 16.6 percent or 33.3 percent. Sampling within the reserve PSUs at different rates can be used to fine tune the sampling rate for the reserve sample to the rate desired. Note that in this kind of design, the problem of certainty PSUs is unlikely to arise. Also, the workload in a PSU of around 10 interviews per PSU is not large enough to provide enough work for even a single interviewer. Thus travelling interviewers are likely to be used.

The costs incurred by setting up operations in a new PSU also need to be considered. These costs include the cost of listing addresses in a new PSU, cost for hiring and training local interviewers, and travel costs. Not all these costs are relevant issues for all countries. For example, listing addresses is not an added cost if the country has a population registry or address frame and uses it as sampling frame. Also the cost for hiring and training local interviewers can be mitigated if a data collection agency already has local interviewers. The cost of travel is another important component for large countries if not using local interviewers. In the US, it is expensive to select a large number of PSUs across the country due to the long distance and consequently high travel costs. However, for a country with a much smaller area, the number of PSUs has less impact on the cost, hence it is more feasible to select PSU reserves in these countries.

When the sample design includes an SSU selection stage, it is generally preferable to select the reserve sample from the SSUs in the main sample PSUs rather than selecting additional PSUs. By remaining in the main sample PSUs, the interviewers conducting the main sample can also contribute to the data collection for the reserve sample if it is needed. Selecting a reserve sample of SSUs avoids the complications arising with selecting a reserve sample of PSUs when the PSUs are large units (as is usually the case when an SSU sampling stage is used).

6. Release of reserve samples

6.1 How Much and When to Release

As discussed earlier, samples can be released in random groups to help control the amount of release. Release groups can be of whatever size desired and the size can vary across groups. The initial release should be based on optimistic projections of the response and eligibility rates, and at the same time should provide a large enough sample for an efficient start of field operations. For PIAAC, the consortium guidelines for decisions related to when and how much more to release included the following three collaborating sources of information:

1) Sample yield projections - (more discussion below).
2) Pace of data collection - Prior to the start of data collection, establish goals for the number of cases to reach in a month. At the end of each month, compare these goals with the actual counts.
3) Field staff circumstances - Have the field director or field managers project the number of completes that could be obtained by the end of data collection, given the amount of fieldwork left to do and the current field conditions as they relate to interviewer performance and types of initial nonresponse.

For item 1), the number of completed cases and response rate at each stage of the data collection may be projected using finalised cases and assumptions for interim cases and cases not yet worked. A response rate toolkit has been provided to PIAAC countries to compute response rates and project sample yield weekly. Continuous monitoring of sample yields during data collection is recommended to help determine the amount of release and ensure timely release as well. To help with the projection, models can be developed based on past experience and using interim disposition codes, contact history, and other para data from the initial release as predictors. Also note that response rates for interim cases tend to tail off fast in the later phase of data collection. It is beneficial to

consider the time in field period (e.g., first month of field period, last month of field period, etc.) in the projection model.

The reserve sample should be released without any delay once a reliable projection is available and a decision is made about needing a reserve. The release should be made early enough to give time to work the cases fully. All the sampled cases in the released groups must be fully processed according to the contact rules (Lavrakas, 2012), otherwise there could be a potential for bias in the sample estimates due to nonresponse. The extent of bias depends on how different the harder-to-reach (potential nonrespondents) and easiest-to-reach cases are in terms of survey outcome. As pointed out by Valliant et al. (2013), an alternative would be to work the release group sample units in a random order, in which case data collection could be stopped partway through the release group. However, working cases in a random order is typically impractical in most in-person survey data collection practices.

6.2 Where to Release

There are many factors that influence the decision on where to release, including budget, the survey practice of the data collection agency, sample design, sample size requirements for both domains and overall, response rates, etc. For instance, if a data collection agency conducts the survey sequentially among PSUs (i.e., one PSU after another), it is not feasible to release a DU or person reserve across all the PSUs. Also, in a stratified (implicit or explicit) design, care should be taken to avoid creating gaps in representation when deciding where to release.

Some survey practitioners may consider releasing more in analytic domains with low response rates in order to obtain more respondents in those domains. However, because of the difficulty in gaining response in these domains, the overall number of respondents will be fewer than that obtained by releasing the sample across the board. Samples in the over-released domain will have higher probabilities of selection hence smaller base weight, but the impact on weight variation may cancel out after nonresponse adjustment since they will have a larger adjustment factor. The decision ultimately depends on sample size requirements, i.e., if a domain is a reporting domain for a published estimate.

7. Summary

To meet the challenges of PIAAC and provide guidance to participating countries, we looked into the issues related to the selection and release of reserve samples. This paper summarized our investigation. By considering, in turn, one-stage or multi-stage designs, we discussed the pros and cons for drawing reserve samples for sampling units at different stages. We recommended selecting a reserve sample at the beginning with the main sample rather than later during data collection. We also discussed issues concerning the release of reserve samples, focusing on how much, when, and where to release them. Dividing the reserve sample into release groups is a useful way to control the amount of release. We reviewed issues regarding the creation and working of release groups. We recommended continuously monitoring sample yield and discussed the development of projection models to aid in determining the amount of release. There are many factors affecting the selection and release of reserve samples. As discussed by Valliant, et al. (2013), the ultimate goal is to ensure a sufficient number of completes to meet the analytic objectives, keeping in mind any ramifications on the budget, time, and, if appropriate, the effects of unequal weighting.

Informed by this research, we made recommendations to PIAAC countries about reserve samples while considering each country's special circumstances. We also note that although reserve samples help improve sample yield, they have no impact on non-response bias. To reduce the risk of non-response bias during data collection, response rates need to be improved through interviewer training, nonresponse follow-up, and other approaches, such as offering incentives, providing endorsement from a credible agency, increasing publicity, etc. We encouraged PIAAC countries to focus on both sample yield and nonresponse bias during data collection.

References

Cochran, W.G. 1977. *Sampling techniques*. 261-263. Hoboken, NJ: John Wiley & Sons.

Kalton, G. and Anderson, D.W. 1986. Sampling rare populations. *Journal of the Royal Statistical Society A*. 65-82.

Lavrakas, Paul J. 2008. *Encyclopedia of Survey Research Methods 2*. 780-781.

Morton, K.B., P.C. Martin, B.E. Shook-Sa, J.R. Chromy, and E.L. Hirsch. 2012. 2011 National Survey on Drug Use and Health sample design report. Available at http://www.samhsa.gov/data/2k12/NSDUH2011MRB/NSDUHmrbSampleDesign2011.pdf (access September 2014).

National Center for Health Statistics. 1992. Sample design: Third National Health and Nutrition Examination Survey. *Vital and Health Statistics 2*. Available at http://www.cdc.gov/nchs/data/series/sr_02/sr02_113.pdf (access September 2014).

National Center for Health Statistics. 2014. National Health and Nutrition Examination Survey: sample design, 2011–2014. *Vital and Health Statistics 2*. Available at http://www.cdc.gov/nchs/data/series/sr_02/sr02_162.pdf (access September 2014).

Valliant, R., J.A. Dever and F. Kreuter. 2013. *Practical tools for designing and weighting survey samples*. 176-177. New York: Springer.

Vehovar, Vasja. 1999. Field substitution and unit nonresponse. *Journal of Official Statistics 15*. 335-350.