

An Alternative Approach for Dealing with Inaccessible Sampled Persons in Registry-Based Samples

Valerie Hsu, Leyla Mohadjer, Tom Krenzke
Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract

Many participating countries in the first round of the Programme for the International Assessment of Adult Competencies (PIAAC) utilized their national population registries as sampling frames. A list of residents is an ideal frame for PIAAC if it provides adequate coverage and includes reliable information to sample and locate individuals for the interview. One challenge faced by some countries with registry samples is the inability to locate some of the selected persons at the registered address, due to erroneous registry data, resulting in high nonresponse rates. For countries with a high proportion of inaccessible persons, we developed an alternative approach that employs address-based sampling in conjunction with person-based sampling from a registry. This paper describes the multi-frame design and discusses the advantages of such a design as compared to single-frame designs that employ person-level sampling from registries.

Key Words: Population registry, noncoverage, frame deficiency, sample design

1. Introduction

In countries where citizens are required to register their place of residence with the authorities, the population registry would be an ideal sampling frame for a household survey provided that it is current and continuously being updated, complete with respect to the target population, and contains reliable information for sampling and interviewing selected respondents. The adequacy of a population registry for sampling purposes depends largely on its quality. An outdated registry contains inaccuracies and if used as a frame source leads to difficulties locating and interviewing sampled persons not residing at the registered address, which may result in high noncoverage rate and the potential for bias. However, for countries with population registries that are of insufficient quality, it may not be feasible to completely forgo the use of registries and adopt a new sampling approach, such as a multi-stage area probability design involving traditional listing for creating dwelling unit sampling frames and dwelling unit sampling, due to resource and budget constraints. As a way of dealing with inaccessible sampled persons resulting from the use of outdated population registries, this paper describes a method of sampling individuals using the registry in conjunction with field enumeration as an alternative to sampling persons directly off the registry. This investigation is most relevant for countries with in-person household surveys that sample from their population registries with insufficient data quality.

The remainder of this paper is organized as follows: Section 2 gives an overview of the Programme for the International Assessment of Adult Competencies (PIAAC), including the requirements for sampling frames and definitions of frame and sample exclusions.

Section 3 describes the treatment of the inaccessible sampled persons in PIAAC and presents an alternative design for dealing with inaccessible sampled persons. The observed noncoverage rates in PIAAC Round 1 are discussed in Section 4. A summary is provided in Section 5.

2. The PIAAC: Round 1 (2011–2012)

Our investigation used data from Round 1 (2011-2012) of the PIAAC, which is an international in-person household survey sponsored by the Organization for Economic Cooperation and Development. The core PIAAC target population consists of all non-institutionalized adults between the ages of 16 and 65 who reside in the country at the time of data collection. The design and implementation of PIAAC is the responsibility of the Consortium, which is a group of companies led by the Education Testing Service, each assigned to a component of the study (e.g., Westat was assigned sampling, operations and quality control). Given the survey complexities and the possibility that different countries would use different survey practices, the Consortium developed a set of technical standards and guidelines to ensure that the survey yields high-quality and internationally comparable data. Participating countries can choose a sample design that is most optimal and cost effective for their survey. However, they must follow the PIAAC survey procedures to produce a probability-based sample that is representative of the target population. Among the 24 participating countries in Round 1, 9 used dwelling unit listing procedures to create local area frames, and 15 drew the samples from their person registries.

2.1 PIAAC Sampling Frames

PIAAC standards require that the sampling frame include sufficiently reliable information for sampling individual units and ultimately locating eligible adults for the interview. It must cover at least 95% of the PIAAC target population. In other words, exclusions (eligible persons with no chance of selection) may represent no more than 5% of the target population. If using the registry to sample persons does not yield a 95% or more coverage of the target population, then participating countries in PIAAC must look for ways to improve the coverage. In Round 1 of PIAAC, the quality of the population registries varied across registry countries. Some countries did not have a centralized system to record address information about their residents, but instead municipality-based registries were available where they were current in some places and outdated in other places.

2.2 PIAAC Exclusions

There is a potential for bias whenever part of the target population is excluded from the frame. Therefore, exclusions had to be limited as much as possible. There are two categories of exclusions in PIAAC – ex ante exclusions by design (frame exclusions) and ex post exclusions following data collection (inaccessible sampled persons). Both contribute to the overall noncoverage rate, which has an upper limit of 5% for PIAAC to guard against high noncoverage bias in the survey estimates.

2.2.1 Exclusions by Design

Exclusions by design or frame exclusions result from persons who are eligible under the definition of PIAAC target population but are not included in the sampling frame. They include exclusions resulting from a decision about certain population groups in the sampling frame (e.g., the populations of remote and isolated regions) given operational or

resource considerations. Additionally, the use of a particular frame source may lead to the exclusion of certain groups in the population by virtue of the rules that determine which individuals are included in the list constituting the frame. For example, many population registers include only those members of the population with valid residence permits and, therefore, exclude illegal immigrants from the population from which the sample is drawn.

2.2.2 Exclusions Related to Data Collection

Exclusions related to the data collection process occur as a natural part of the survey process. It involves persons who are included in the frame, and selected into the sample, but in practice are impossible to be interviewed because no information can be found about their location or it is confirmed that they have moved to areas inaccessible by interviewers. This type of exclusion is only applicable to registry-based samples. Unlike household-based samples where eligible individuals are not identified and selected until the households are screened for their presence, registry-based samples select individuals for interview in advance of any attempted contact. Therefore, it will not be possible to make contact with some individuals because the registry may be out-of-date at the time of sample selection and sampled individuals may have moved to somewhere other than their address on the registry between sample selection and the initial contact.

Sampled persons not residing at their registered address could fall into several categories. Table 1 provides an overview of the different types of non-contact and their status. Some individuals will be found to be out of scope when the contact is attempted. For example, some may have deceased, emigrated, or moved to an institutional setting, such as prison, hospital and nursing home. Others will be untraceable in that no information about their whereabouts is available. Finally, some will be inaccessible in that they have moved to another location not accessible by interviewers for operational and cost reasons or because the information about their residential address was incorrect or invalid¹ and therefore cannot be interviewed. These individuals are the focus of this paper and are referred to as inaccessible sampled persons (henceforth termed simply IAs). If the population registry is not well maintained in the first place, the exclusion due to IAs might be much higher than what the PIAAC standards would allow.

Table 1: PIAAC Categories of Non-Contacts and Their Status

<i>Description</i>	<i>Status</i>
Deceased	Out of scope
Moved outside country	Out of scope
Moved to institution	Out of scope
Unknown whereabouts	Unknown exclusion status
Moved inside country	Inaccessible
Invalid address	Inaccessible

3. Treatment of Inaccessible Sampled Persons in PIAAC

IAs could not simply be classified as ineligible since they are part of the target population. If IAs were treated as ineligible, people with higher mobility would be under-covered in some countries and the rate of under-coverage would be different due to the

¹ An invalid address could be one that is nonexistent, demolished, under construction, non-residential, or not occupied.

variation in the quality of the registries across countries. This would create a great deal of bias when cross-country comparisons are made.

As mentioned earlier, the noncoverage rate, combined over all stages of sampling in PIAAC, could not exceed 5%. Although IAs are regarded as eligible since they are part of the target population, PIAAC standards allow countries to include this group in their “excluded” population up to the threshold of 5%, overall. The information about the IAs could be used to reduce the bias associated with undercoverage and, thus, reduce inconsistencies between country data. If the proportion of the target population that is excluded (including the inaccessible cases) exceeds the 5% threshold, an alternative approach of using an address-based rather than a person-based design is recommended.

In keeping with the goal of producing high-quality unbiased estimates that are comparable across countries, the Consortium developed two options that are in accordance with the PIAAC standards to deal with IAs in the sample. These two options are described below in further detail.

3.1 Option 1: Person-Based Sampling and Treating Inaccessible Sampled Persons as Exclusions

IAs are part of the PIAAC target population, but are under-represented since they are not accessible. Option 1 is to count IAs as eligible but being outside the sampling frame, and thus classified as excluded. Under this approach, countries that have lists of persons with sufficient quality would employ a self-weighting sample design, such as simple random sampling or systematic sampling from a sorted list, to select individuals and IAs would be treated as exclusions conditional on the total exclusion rate staying at or below 5%.

For countries that use Option1 and have an overall exclusion rate greater than 5%, up to 5% are reported in the exclusion rate and the portion greater than 5% are treated as nonresponse in the response rate calculations. For example, if the frame exclusion amounts to 2.8% of the target population, and IAs represent 3% of the eligible sample, then 2.2% of the eligible sample who are IAs can be treated as additional exclusions with the remainder (0.8%) being treated as eligible nonrespondents in the computation of response rates.

During weighting, an adjustment for persons with unknown exclusion status is performed if the exclusion status of some sampled persons could not be determined due to the inability of the survey to locate these selected persons (not residing at their registered address). In the unknown exclusion adjustment, an adjustment factor is computed as the proportion eligible among those with known exclusion status to down-weight the cases with unknown exclusion status (accounting for an estimated proportion that is ineligible). The down-weighted cases with unknown exclusion status are then treated as eligible nonrespondents in the nonresponse adjustment. It is important to use high quality external data and consider alternative adjustment methods during weighting to reduce the bias associated with the exclusion of IAs.

Since not all countries can use this approach because treating IAs as excluded could increase the exclusion rate considerably above the 5% threshold, the Consortium developed a second option of implementing an address-based sample design in which sampled persons are selected from the addresses in the registry.

3.2 Option 2: Address-Based Sampling from Population Registry

Under the alternative design, the following steps need to be taken to arrive at a random representative sample of persons:

- a) Construct a listing of addresses with adults 16 to 65 years old from the registry information
- b) Select a sample of addresses with probability proportional to the number of eligible adults in the household
- c) Pre-select a person from the household's list of eligible adults in the registry within each sampled address
- d) At the time of contact, confirm the listing of adults 16 to 65 years old at the address as given in the registry:
 - i. Take the original sampled person if the listing has not changed
 - ii. If the listing has changed, then create a new listing and randomly select a person from the new list

It is of note that the listing of eligible adults obtained from the population registry needs to be checked by an interviewer when each sample household is contacted. The household composition between the registry and the household screening must coincide otherwise the household will be re-enumerated and a person will be re-selected. The interviewer will take the originally sampled person if the listing of eligible adults has not changed. If it has changed, even if the originally sampled person is at the residence, the interviewer will enumerate the household and randomly select a person from the new listing.

During the weighting process, the household-level adjustments need to include the creation of the household base weights that reflect the household selection probability and adjustments for unknown eligibility (if the eligibility status of some households could not be determined) and nonresponse to the screener. Unlike Option 1, an adjustment for unknown exclusion status is not warranted at the person-level under the address-based design.

3.3 Comparison of the Two Options

Both of the aforementioned options satisfy the PIAAC survey procedures and lead to results that are comparable across countries using registry or household sampling frames. Each method offers its own set of advantages and disadvantages.

Overall Person Probability of Selection

Under Option 1, each sampled person has an equal probability of selection resulting from the self-weighting design of persons. In Option 2, since the household probability of selection is proportional to the number of eligible household members, the overall person probability of selection will be equal across sample persons if the household screening reveals no discrepancy in the number of eligible adults reported in the registry. As a result, under Option 2, there is variability in the weights only if the number of eligible adults based on the household screening differs from that in the registry.

Preparation Work

Under Option 1, individuals are selected in-house for interview in advance of any attempted contact. In contrast, there is more in-house and field work involved with Option 2. Not only do countries need to glean more information from the registries to

construct the address frame, they also need to develop a household screener and administer the screener.

Ability to Send Advance Letters

Both options allow for the mailing of personalized advance letters to motivate response cooperation. However, if there is a serious concern about the quality of the registry, countries may want to send a generic, as opposed to a personalized, advance letter.

In-Field Data Collection

To arrive at an optimum and consistent approach across all registry countries, under Option 1, countries should attempt to find the location of the sampled persons and try to interview them if they moved to one of the primary sampling units in the sample or are in a location where it is possible for PIAAC interviewers to visit and conduct the interview. Unlike Option 1, Option 2 eliminates the need for following up with sampled persons who do not reside at their registered address by switching the design from a person-based sample to an address-based sample.

Noncoverage and Coverage Improvement

Option 2 covers the IAs and will be able to cover those who would have moved into vacant dwelling units since the frame creation by including a coverage enhancement procedure such as a missed dwelling unit procedure². If there is a high proportion of IAs, the address-based design improves the representation of the sample by lowering the sample exclusion rate.

4. Noncoverage Rates

The noncoverage rates of the IAs, sampled persons with unknown whereabouts, and out-of-scope sampled persons are calculated in the same manner as the non-contact sample divided by the eligible sample as follows:

$$\text{Inaccessible Exclusion Rate} = \frac{\text{Inaccessible Sampled Persons}}{\text{Total Sample Size}}$$

$$\text{Unknown Exclusion Rate} = \frac{\text{Sampled Persons with Unknown Exclusion Status}}{\text{Total Sample Size}}$$

$$\text{Out - of - Scope Rate} = \frac{\text{Out - of - Scope Sampled Persons}}{\text{Total Sample Size}}$$

These individual noncoverage rates as well as the overall sample exclusion and out-of-scope rates are shown in Figures 1 through 4 for each country in succeeding pages. Note that specific results from PIAAC Round 1 could not be presented at the time of this writing since the survey results have not been officially published. Country names were masked and were denoted by a randomly assigned letter.

In PIAAC Round 1, 14 out of 15 countries with registries that were of varying degrees of accuracy used Option 1. Some of these countries were bound to use their registries as is

² The missed dwelling unit procedure is applied to identify dwelling units that are missed on the original sampling frame. Once confirmed, missed dwelling units are added to the frame and given an opportunity to be selected into the sample.

and were unable to implement the screening process involved in Option 2 due to resource and budget constraints. Only one country (Country Z) adopted Option 2. In this country's field test, they found that the registry information supplied by their municipalities was sometimes outdated or incorrect. This country agreed with the Consortium's recommendation to switch from a person-based sample to an address-based sample.

4.1 Noncoverage Rate of IAs

Figure 1 displays the inaccessible exclusion rates for the 15 countries with registries in ascending order. In PIAAC Round 1, the inaccessible exclusion rates ranged from zero to 13%. Country Z, which used the alternative address-based design, had a relatively low inaccessible exclusion rate close to 2%. It was not zero because invalid addresses in the registry still contributed to a portion of the sampled persons being inaccessible.

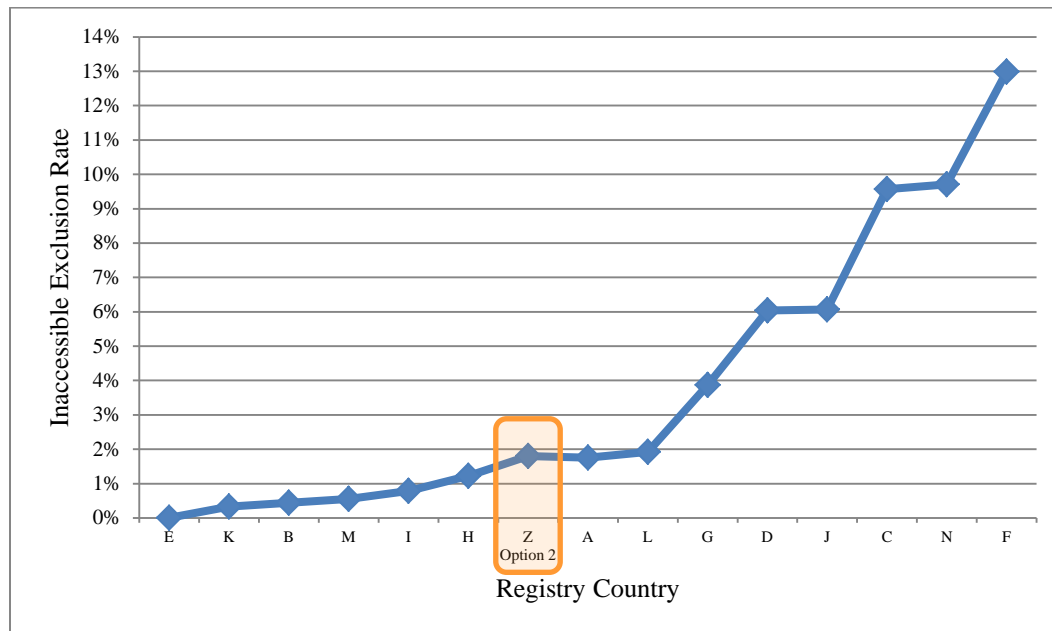


Figure 1: PIAAC Round 1: Rate of Inaccessible Sampled Persons

4.2 Noncoverage Rate of Sampled Persons with Unknown Whereabouts

Figure 2 presents the noncoverage rates of sampled persons with unknown whereabouts, which ranged from zero to 9%. Sampled persons with unknown whereabouts differed from the IAs in that no information about their whereabouts was available and their exclusion status could not be ascertained. For Country Z that used Option 2, there were no sampled persons with unknown whereabouts by design since the households were re-enumerated and a person was reselected when these cases were identified.

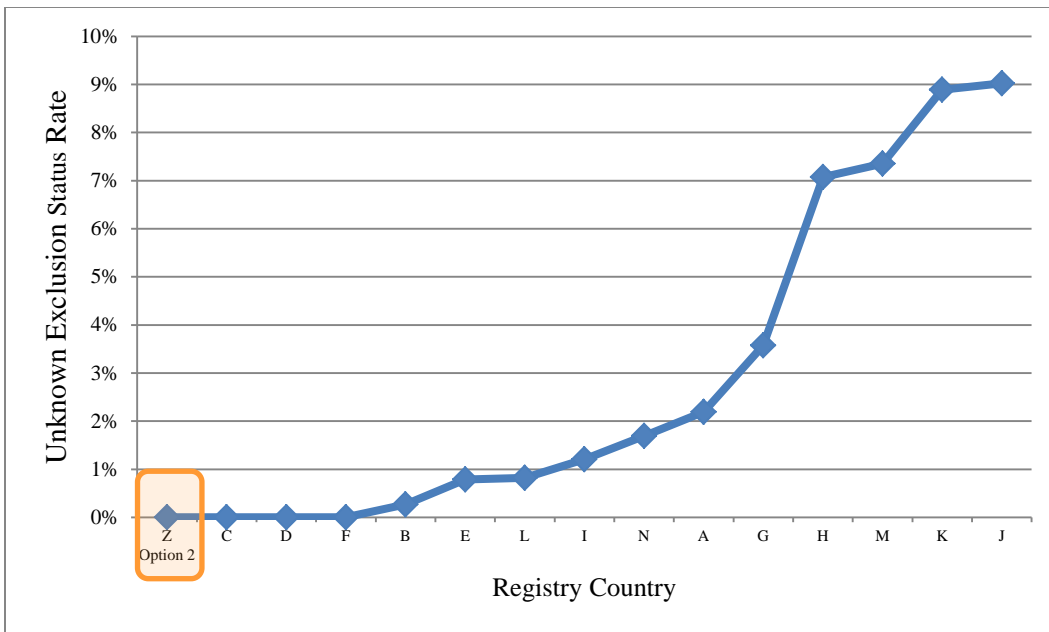


Figure 2: PIAAC Round 1: Rate of Sampled Persons with Unknown Exclusion Status

4.3 Noncoverage Rate of Out-of-Scope Sampled Persons

Figure 3 shows the noncoverage rate of the out-of-scope sampled persons, ranging from zero to just above 8%. Out-of-scope sampled persons are those who had deceased, moved outside the country, or moved to an institutional group quarter during the interval between the sample selection and the interview. Under the alternative design, there were no such cases by design.

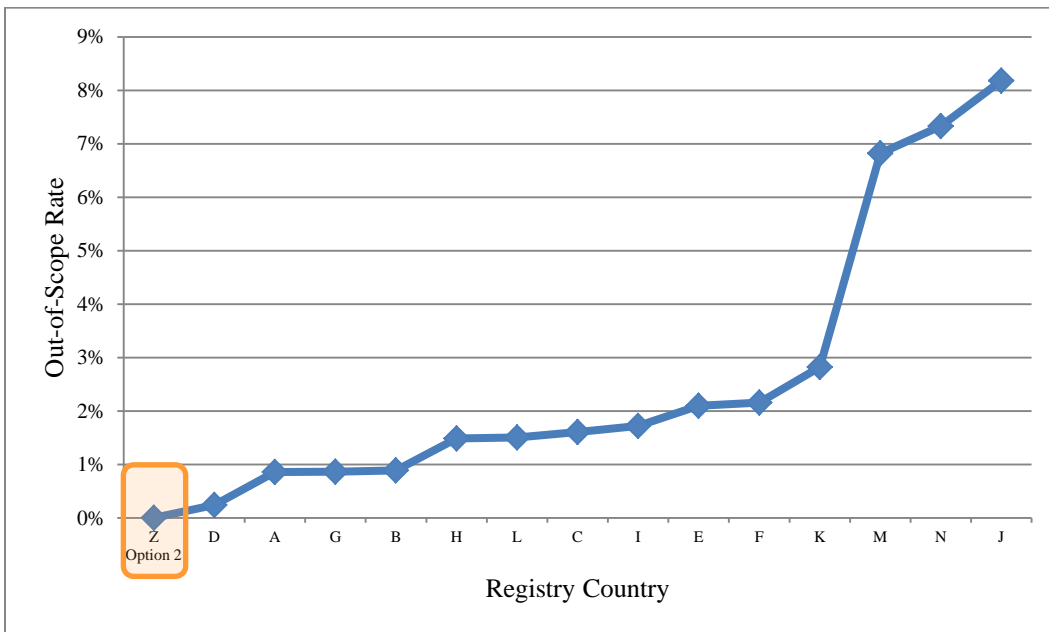


Figure 3: PIAAC Round 1: Rate of Out-of-Scope Sampled Persons

4.4 Total Sample Exclusion and Out-of-Scope Rate

Figure 4 shows the sum of the three sample exclusion and out-of-scope rates presented in Figures 1 through 3. Overall, Country Z that used the alternative design is next to Country D with the best registry data. If Country Z did not employ the alternative design, based on what we know from their field test, we would expect it to end up on the right of the plot with countries that had poorer quality of registries. This demonstrates the effectiveness of the alternative address-based design in lowering the sample exclusion rates.

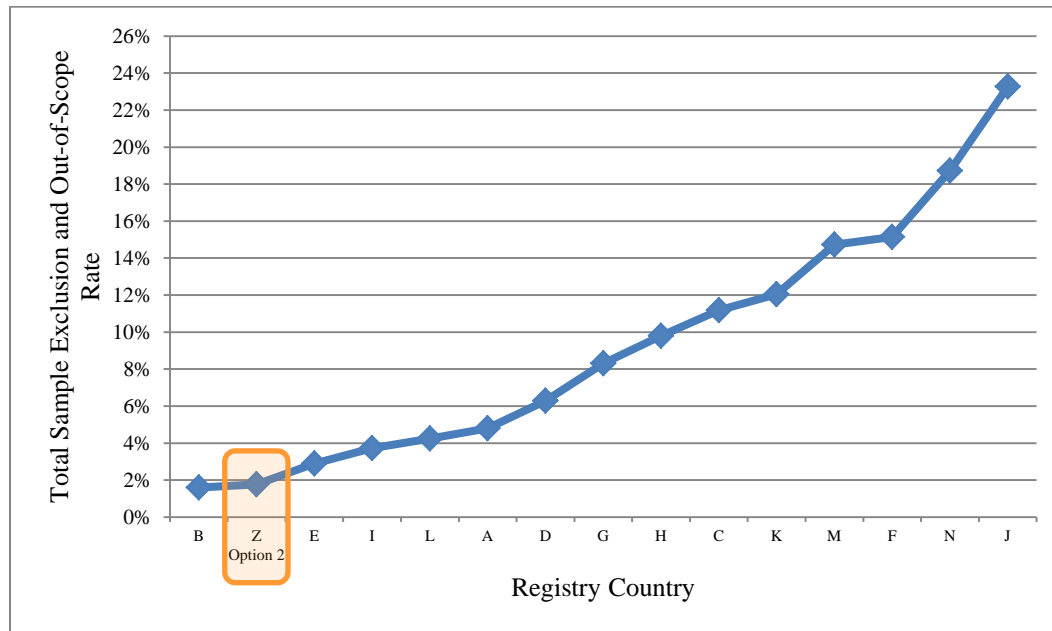


Figure 4: PIAAC Round 1: Total Sample Exclusion and Out-of-Scope Rate

5. Conclusions

In summary, it is important to consider the quality of the population registries and the consequences of using them as sampling frames, especially the implications of noncoverage and the potential for bias in survey results if part of the target population is excluded. As a way of dealing with IAs in registry-based samples, the population registry could be used in conjunction with field enumeration to improve the representation of the sample. In future cycles of PIAAC, countries using registries with insufficient quality will be encouraged to employ this alternative design. Although this investigation draws the conclusion from a case study, we believe this recommendation is reasonable given the dramatic improvement in the sample coverage rate that was observed in the country that employed the alternative address-based design.