

Post-Enumeration Survey

1. Introduction

During the UNFPA-SPC Regional Workshop Reviewing the 2010 Round of Population and Housing Censuses in the Pacific countries agreed to consider implementing plans for Post-Enumeration Surveys (PES) to evaluate the coverage and quality of their censuses. This paper reviews the advantages of the kinds of analyses possible with a PES, as opposed to simple consistency checks and demographic analysis. SPC recommends a simplified methodological approach that is smaller in scope than the type of PES conducted in countries with larger budgets and staff resources.

2. Advantages of a Post-Enumeration Survey

A Post-Enumeration Survey allows the calculation of quality measures that are not available through standard consistency checks or demographic analysis. While those methods are still advised, it is important to recognize the limitations inherent in each. For example, evaluating the internal consistency of the current census with age or sex ratios, digit preferences, or visual checks for outliers, etc., can only help you identify unusual or unexpected data points. These methods cannot be used to distinguish between data collection errors and unusual (but real) population trends. That is, they can be used to identify likely problems, but cannot measure the magnitude of the error without assuming what “normal” might be.

Many countries rely on traditional demographic analysis to evaluate the quality of the most recent data collection. There are several reliable, useful methods available for intercensal analysis, ranging from the simple application of a population balancing equation ($Pop_2 = Pop_1 + Births - Deaths \pm Migration$) to the use of more complex methods such as indirect estimation and cohort component analysis.

The main shortcoming of these methods is that they all rely on assumptions about the accuracy and completeness of prior data collection efforts and/or on assumptions about the nature of population trends between the two censuses (i.e., that there is no migration, that fertility patterns are stable, that the age structure of mortality is constant, etc.). For example, an apparent shortage of males age 30-34 could be attributable to a variety of causes: an undercount during this census; population decline from undocumented emigration; higher than expected mortality, e.g., as seen in many populations with HIV/AIDS epidemics; or an overcount of that cohort of men during the previous census. Without corroborating information from some other data source it is impossible to determine what the real cause of the apparent shortage might be.

Post-Enumeration Surveys, in contrast, do not require the analyst to make any untestable assumptions. If conducted shortly after census enumeration, as recommended, PES estimates apply to nearly the same point in time, making population change and seasonality issues less likely to produce biased results. Utilizing identical census procedures, materials (i.e., maps and household listing), and questionnaires will also help minimize measurement and/or systematic error caused by differences in methods. Finally, PES estimates of coverage can take advantage of the ability to directly match of households and individuals between the two data points. This point is critical to determining whether overcounting, where some people are counted two or more times, might be a problem.

3. Recommended Design and Format

Planning and budgeting for a PES should be done in conjunction with the same preparations for the full census. Most countries should aim to complete their PES in the week following the completion of census enumeration, which, in turn, ideally would be completed within a month of the official Census Day. UN recommendations suggest hiring separate teams of enumerators for the two operations. However, many countries find it more useful to re-hire a select group of experienced census enumerators, making sure to assign them to a new area for the PES.

The sample for a PES recommended by SPC is for a minimum of 800 households in order to provide a reasonably accurate estimate of coverage at the national level. To assess coverage for smaller geographical regions or for special subgroups of the population, basic statistics shows that the sample for each needs to be at least 600 per stratum to ensure a reasonable confidence interval for your coverage estimates.

The sample should be a randomly selected set of sub-areas (EAs, districts, etc.) with a total population of households sufficient to meet your country's sampling need. For example, in Palau we knew the required sample size was equal to about 15% of the total households in the country, so we gave each Enumeration Area an equal 15% chance of selection. The selection should be by geographic area to allow PES enumerators the chance to identify households missed during the listing or census enumeration; selecting directly from the household listing would preclude this step.

PES enumerators should receive the same training as census enumerators, including detailed instruction in the use of maps and identifying occupied households that may have been missed during the listing exercise, or that might contain recent occupants. Rosters should be completed in

reference to occupants that stayed in each household on Census night. This is a major reason that the PES should be conducted as close to Census day as possible to reduce the risk of recall error.

The scope of the PES questionnaire can be as limited as the original household listing or it can include a replication the entire census questionnaire. SPC recommends collecting at least a complete roster for each household, as that will allow matching at both the household and person-level. That is, each country should be able to produce an estimate of coverage for households as well as coverage of persons.¹ Data processing, supervision, and quality control procedures should all be replicated from existing Census systems. This is both cost-effective and helps to improve comparability and consistency between the two collection points.

Coverage is calculated by identifying three key bits of information: the number of households or individuals enumerated in both the Census and the PES; the number enumerated in the Census but not in the PES; and the number enumerated in the PES and not the Census. At the simplest level, coverage is calculated as the number matched in both operations divided by the total enumerated in the PES, and is usually expressed as a percentage. Similarly, for countries that included items from the census questionnaire, item-level coverage is calculated from households/individuals in both enumerations, and is simply as the number of completed or matching items divided by the total.

4. The PES in Practice

For reference purposes, this section is the proposal for the PES conducted in Palau in late June, 2015. At the time this paper is being written data entry for both operations has yet to be completed, so we are only able to provide hypothetical estimates of coverage and accuracy based on the expected number of households to be found in the sample enumeration areas.

We will randomly select a 15 percent sample of all enumeration areas in Palau. A select group of 18-20 of the best-performing census workers will have 2 weeks to completely re-list and re-enumerate the selected EAs. Each worker will be assigned to one or more EAs in the PES; the assigned area cannot be the same one they worked for the Census.

The primary goal is to be able to calculate a net undercount at national level, and to estimate the urban and rural participation rates if possible. Additional objectives are to evaluate the quality and

¹ Estimates at the household- and person-levels may not be the same; single-person households often have poorer coverage than multi-person households and some person-level characteristics are often correlated with selection bias. For example, young males are more likely to be under-enumerated because they are often working in under-covered housing arrangements like worker barracks, and have less-stable living arrangements than older, married men and women.

completeness of work of Census enumerators and to assess the overall quality of questionnaire, training, and other census operations.

The PES instrument is a 2-page questionnaire, designed to fit on the front and back of a single sheet of A4 or letter-size paper. The front page replicates the Household Roster used for both household and group quarters. The back side of the form contains about 20 additional questions, including items selected from household and agriculture sections, and person-level questions for the head of household or spouse.

PES enumerators will be paid at a rate of \$10 (U.S.) for each completed questionnaire (\$5 for the interview, plus \$5 for fuel and transportation), and a per diem of \$10 per week to cover telecommunications costs. All expenses, training, and supervision will be the responsibility of SPC staff. The Senior Analyst at OPS has final approval of all activities, including the hiring of personnel for the PES. However, SPC intends to complete this project while minimizing the impact on OPS in terms of staff time and resources.

The statistical methods used to calculate the undercount is identical to that used by most other countries that have conducted post-Census surveys and is similar to the capture-recapture methodologies used by SPC Fisheries projects and to estimate wildlife populations. We assume that neither the Palau Census nor the PES will be 100 percent complete and accurate. By comparing the two sets of numbers, however, we can estimate the undercount – if any – with a reasonable degree of precision.

Palau will be the first country in the region to conduct a PES in the 2020 round, so there are few points of comparison to tell us what level of undercount might be expected. New Zealand conducted a Post-Enumeration Survey following their 2013 Population Census and found a 2.4 percent undercount. Papua New Guinea estimated a 4 percent undercount in the 1990 and 2000 Censuses based on a Post-Enumeration Check (not a true PES). The PES for the 2011 Australia Census calculated an undercount of 1.7 percent at the national level. Australia also provides state-level estimates of undercounts; for the Northern Territory, which has about the same size population as Palau, the undercount was 6.9 percent.

These calculations for Palau assume a net participation rate of 95 percent in urban areas and 97 percent in rural areas the sample size proposed for the PES will provide an estimate of the undercount with a margin of error of +/- 2.2 percent. Assuming the total enumerated population for the Census comes to 16,587 the estimated undercount will be between 2.4 and 6.7 percent. That is, we would calculate that the actual population of Palau has a lower bound of 16,992 and an upper

bound of 17,771. Applying the same rules at the household-level would give us an estimate of occupied housing units ranging between 5,093 and 5,315. (There were 4,966 occupied housing units counted in the listing exercise.)

The estimates for rural and urban areas will naturally be subject to broader confidence intervals. With an estimate of about 600 households in the selected urban areas the statistical margin of error is expected to be +/- 2.6 percent. The selected rural EAs are expected to have only about 120 occupied households, which gives us an MOE of +/- 4.7 percent. In this case, if the true participation rate in rural areas is 97 percent, we will not be able to determine with statistical certainty whether there was really an undercount, or possibly a small overcount (the 95 percent confidence interval will range from 92.7 to 101.7 percent, implying a possible overcount of about 58 people and 18 households in rural areas).

The estimated workloads and other data in Table 1, below, come from the Palau Household Listing conducted in December 2014.

Table 1. Characteristics of Selected Enumeration Areas

State Name	Hamlet Name	EA	Pop	HHs	Estimated Workload (in days)
Aimeliik	Ngerkeai	A	29	8	2.00
Koror	Idid	C	201	63	15.75
Koror	Ikelau	A	192	83	20.75
Koror	Ikelau	C	204	63	15.75
Koror	Madalai	E	84	21	5.25
Koror	Ngerbeched	B	186	54	13.50
Koror	Ngerchemai	E	302	74	18.50
Koror	Ngerchemai	G	187	49	12.25
Koror	Ngerkebesang	B	60	15	3.75
Koror	Ngerkesowaol	A	212	81	20.25
Koror	Ngerkesowaol	C	178	57	14.25
Koror	Ngermid	B	212	49	12.25
Melekeok	Melekeok	A	16	6	1.50
Melekeok	Ukaeb	A	17	15	3.75
Ngaremlengui	Ngerutchei	A	82	33	8.25
Ngchesar	Ngerngesang	A	68	26	6.50
Ngiwal	Ngermechau	A	55	28	7.00
Ngiwal	Ngersngai	A	105	34	8.50
Total			2390	759	189.8
Total PES Workload				38	person weeks

Table 2 shows the values used to determine how accurately we will be able to calculate whether the Palau Census experienced an undercount. For this mathematical exercise the Household Listing conducted in December 2014 is used as a proxy for the actual census results, which are not yet available. The coverage rates for the Census and PES are hypothetical and used here only for illustrative purposes.

	Estimated Occupied Households	Estimated Population	Average HH Size	Occupied Households in PES	PES Sample Coverage	Census Coverage Rate (<i>Made Up</i>)	PES Participation Rate (<i>Made Up</i>)	standard error	Margin of Error +/-
Total	4,966	16,586	3.3	674	13.6%	95.4%	91.0%	0.011	2.2%
Urban	3,902	13,422	3.4	556	14.2%	95.0%	90.0%	0.013	2.6%
Rural	1,064	3,340	3.1	121	11.4%	97.0%	93.0%	0.024	4.7%

	Net Under-Count (<i>Made up</i>)	Under-Count Minimum	Under-Count Maximum	Corrected Occupied HH Count	Lower Bound HH Count	Upper Bound HH Count	Corrected Pop Count	Lower Bound Pop Count	Upper Bound Pop Count
Total	4.6%	2.4%	6.8%	5,205	5,086	5,330	17,386	16,987	17,803
Urban	5.0%	2.4%	7.6%	4,107	3,999	4,222	14,128	13,754	14,524
Rural	3.0%	-1.7%	7.7%	1,097	1,047	1,152	3,443	3,285	3,617

5. Conclusion

A Post-Enumeration Survey is a very cost-effective means to assess the quality and coverage of any country's Census. For the test conducted in Palau, total fieldwork costs were less than \$10,000 (US) and were completed in less than two weeks. As with other stages of the census, the use of clear and consistent identification procedures for households and individuals is essential. For the PES, accurate identification is essential to matching households and individuals to their census records, and this matching is the key component in the calculation of coverage rates.

Given the lack of experience in conducting PESs in the region it is likely that there will be some unforeseen problems. However, as was found with the first PES conducted in Palau, these problems can be overcome. Having an alternate method available to evaluate the current census will give officials more confidence in the timely release of data, as PES estimates also provide a clear delineation between actual measurement errors and unusual or unexpected and difficult-to-explain population trends. The PES also provides a measure of confidence in current census procedures because it does not rely on assumptions about the quality or completeness of previous data collection efforts.