Assessing Accuracy and Coverage of Housing Units Sampling Frame using the Post-Frame Update Survey (PFS) Method

Elmogiera Elawad[*], Yioryos Nardis[**] and Mohamed Agied[***]

[*]University of Qatar – SESRI

[**]University of Michigan – ISR

[***]University of Qatar – SESRI

[*]*Corresponding author's email: Elmogiera [AT] qu.edu.qa*

_____

ABSTRACT--- *Survey research using nationally representative samples is growing substantially in the rapidly developing countries of the Gulf Cooperation Council (GCC). However, in such countries the population and their demographic makeup change at some of the fastest rates in the world. In Qatar for instance, government censuses tend to be undertaken every five to six years, while the population and its demographic composition change very quickly (Alrouh, Ismail & Cheema, 2013). In 2004, the population of Qatar was just 798,000 and Qataris represented 24% of the country's residents. The 2010 census measured the population at 1,699,000, only 17% of which were Qataris.*

In such countries therefore, sampling frames based on censuses may quickly become outdated as demographic compositions change considerably over short time periods, more so than in other countries. These contexts present a challenge for survey research, as the target population may differ significantly from the sampling frame shortly after it is constructed. Outdated frames may lead to coverage and content errors that not only affect the estimates produced by a survey but also operational processes associated with data collection. Frame error can negatively impact time spent in the field and survey costs. Periodic evaluation of the quality of a sampling frame used between censuses is required for estimating coverage bias and informing researchers about the quality of the frame data. Adjustments in field operations can then be made to improve the effectiveness and efficiency of data collection.

This paper describes the Post-Frame update Survey method conducted by the Social and Economic Survey Research Institute (SESRI) at Qatar University. It aims to assess the accuracy and coverage of its housing units sampling frame and improve field operations. This method of evaluating sampling frames could be applicable to other research organizations that survey rapidly changing populations.

_____

1. COVERAGE ERROR IN HOUSING UNIT LISTINGS

Coverage error refers to the gap between the target population and sampling frame, and can occur if housing units are overlooked (undercoverage) or if non-housing units are incorrectly included (overcoverage). Housing unit frames are often created by field staff that canvas areas and list their housing units. Such frames will suffer from undercoverage if eligible units are omitted, as people that live in omitted units will have no probability of being selected to participate in surveys. Undercoverage is more common in low-income and rural areas (O'Muircheartaigh, English, & Eckman, 2007). In urban areas, field staff may have difficulty identifying all households in multi-unit buildings, or might be unable to enter gated units (Barrett, Beaghen, Smith, & Burcham, 2002; Eckman & Kreuter, 2013).

Undercoverage can lead to bias if the propensity to be included in the sampling frame is related to a survey variable of interest (Valliant, Dever, & Frauke, 2003). For example, if field staff tend to omit low-income households, which are less likely to have Internet access, then estimates of household Internet access may be biased due to frame errors.

2. SAMPLING FRAME AND DESIGN

Adequate frames are not readily available outside of Government in Qatar. Therefore, the sampling frame was developed by SESRI itself with the assistance of the Qatar Electricity and Water Company (*Kahramaa*). The target population includes persons 18 years or older that live in residential housing units in Qatar. The sampling frame includes information on the location of each household, whether it is private or collective, GIS coordinates, and the nationality of its residents. All housing units in Qatar are listed in the frame with information about the address, and if residents are Qataris, expatriates, or laborers. There are two distinct population groups in Qatar: Qatari nationals and non-Qatari contract employees. The latter group is also composed of two subpopulations: expatriates and laborers. "Expatriates"

refers to the higher income contracted non-nationals who usually live in household units often with family members. "Laborers" refers to lower income contract employees who usually live in labor dwellings unaccompanied by family.

The State of Qatar is divided into seven municipalities. Each municipality contains zones divided into blocks. To assure representation of the population in zones, proportionate stratified sampling is used where each zone is considered to be a stratum. Proportionate allocation implies that the sample in each stratum is selected with the same probabilities of selection. Inside zones, housing units are ordered by geographic location and a systematic sample is separately constructed for Qataris and expatriates. This aims to stratify proportionately, as a block containing a certain percentage of Qatari or expatriate housing units in the frame would be represented by the same percentage of the total number of sampled units. Migrant laborers tend to reside in collective dwellings. Based on the information about the number of people in the labor dwellings, the frame for laborers is divided into strata and proportionate stratified sampling is used to draw separate samples from these strata.

3. THE SESRI POST-FRAME UPDATE SURVEY METHOD

Several post-enumeration methods are available to evaluate sampling frames, including administrative record checks, matching against samples from independent lists, dual-system estimation (Wolter, 1986; Hogan, 1992), and frequent callback (U.S. Census Bureau, 2004). In many developing countries however, alternative sources of housing unit data are not available.

Therefore, the Post-Frame update Survey (PFS) is often the main tool used to assess the accuracy and coverage of a housing units sampling frame. Its objective is to evaluate the frame by estimating measures of coverage error, including omissions and incorrect enumerations, in order to improve field work strategies. The results of the PFS are used to manage the field in terms of cost and time. We outline this method and present the results of our 2014 frame evaluation to illustrate how it can be carried out.

A Frame Update Survey (FUS) was conducted in March/April 2014 and the PFS undertaken in November 2014 to evaluate the FUS. The FUS is a frame maintenance tool whose objective is to update housing unit information. Post-Frame update Surveys are independent surveys that repeat the FUS. The PFS allows deficiencies in the methodology of the FUS to be discovered and adjustments for future frame updates to be made. Comparing the PFS with the FUS produces estimates of coverage and content errors. Post-enumeration surveys typically assess the full census, and consist of evaluating housing units, establishments, and the population. At SESRI we are only concerned with the first step, evaluating housing units, and not a full post-enumeration survey.

To estimate the accuracy and coverage of the housing units frame, a sample of 71 frame blocks that were counted in the Frame Update Survey are randomly selected from each of the 98 areas which constitute the geographic areas of Qatar. This represents about 2.7% of the total number of blocks. To ensure independence, the PFS is undertaken by staff members and interviewers who did not work on the FUS. Field staff used dependent listing, that is, a checklist of existing household addresses in the sample, through which to verify if the information is correct. A sample of 800 Qatari and 800 expatriate housing units was randomly selected. We attempted to collect the following information from the selected housing units:

- Zone and municipality

- Street name and number

- Building type

- Electricity and Water Company (*Kahramaa*) account number for the housing unit

- Head of household nationality

- Geographic Information System (GIS) coordinates

Most information can be collected outside the housing unit and without interviewing respondents. The only measure that requires interviewing is the *head of household nationality*. Interviews were conducted using Computer-Assisted Personal Interviewing (CAPI) technology. Interviewers and supervisors, most of whom had relevant prior experience, were recruited and carefully trained for the survey.

The following indicators were then calculated to evaluate the frame:

- Accuracy of housing units addresses

- Housing units that could not be located
- Percentage of error in the nationality of household head

- Accuracy of housing GIS coordinates

- Destroyed housing units

- Time to locate selected housing units

- Number of interviews completed per day for each interviewer

4. RESULTS

To evaluate frame coverage, the 71 randomly selected blocks were distributed across the municipalities they belong to (Table 1). Around 84.5% of selected blocks come from the Doha and Rayan municipalities. This is consistent with the distribution of housing units in Qatar, according to both the 2010 census and the 2014 SESRI Sampling Frame. In each randomly selected block from the Post-Frame update Survey, the number of housing units is compared to the number of housing units coming from the Frame Update Survey. Around 67% of blocks in the PFS contain the same number of housing units as the frame, 23% of blocks contain more housing units than the frame and 10% contain fewer housing units than the frame.

A total of 2163 interviews were completed, including 1058 Qatari nationals and 1105 resident expatriates. The response rates are 54 percent for Qataris and 68 percent for expatriates[1]. For comparison, the percentage of completed interviews for four other SESRI surveys conducted in 2013 and 2014 ranged from 43% to 56%. On average, it took interviewers about 20 minutes to locate each selected housing unit and each interviewer completed an average of two interviews per day. Whereas around 7% of housing units in the 2014 frame could not be located, only around .1% of housing units in the PFS could not be located. The final dispositions for the sample cases are presented in Table 2.

To evaluate the accuracy of the frame and estimate content errors, eight characteristics of housing units were collected. The results indicate that overall, characteristics of housing units were very accurate (Table 3). Building type, housing unit frame number, and street name and number ranged from 95.5-98% accuracy. Zone and Municipality identification numbers were completely correct. The Electricity and Water Company (*Kahramaa*) account number for the housing unit was the least accurate at around 90%. Moreover, around 96% of the sample had correctly identified the nationality of the head of household. Taken together, the overall average accuracy of housing unit characteristics is around 96%.

With regard to the accuracy of GIS coordinates, around a quarter of housing units were within a range of 10 meters from the coordinates specified in the frame, and approximately 62.5% were within 30 meters (Table 4). Around 20% of GIS coordinates were considered to be quite inaccurate as the housing units were beyond 30 meters of the coordinates specified in the frame.

5. ADJUSTMENTS TO FIELD OPERATIONS

Based on these results, adjustments to field operations can be made to improve the effectiveness and efficiency of data collection for future surveys. The accuracy metrics calculated are averaged and an overall accuracy percentage is assigned to each block. If a given block is estimated to have a given accuracy in the characteristics of its housing units, then the number of interviewers sent to that block and the number of interviews they are asked to attempt are modified accordingly.

For instance, if we'd like to complete one hundred interviews from a certain block and we estimate that it has 90% accuracy in the characteristics of its housing units, then interviewers will be instructed to aim for ninety complete interviews. This reduces the time interviewers spend trying to contact respondents in areas where it is known that the list of housing units may contain errors. As a result, the efficiency of data collection is enhanced and the costs associated with data collection are contained.

---

[1] Response rates are calculated on the basis of Table 2. The response rate (AAPOR RR1) is the ratio between the number of completes or partials and total sample sizes after excluding ineligibles: $RR1 = \frac{C}{C+E+UE}$ where C is the number of completes or partials, E is the number of eligible responses, and UE is the number of unknown eligibility.

6. CONCLUSION

This paper demonstrates how the Post-Frame update Survey can assess the accuracy and coverage of a housing units sampling frame, and ultimately improve field operations in terms of cost and time. This method of evaluating sampling frames is important in contexts where populations and their demographic makeup change at very fast rates and sampling frames quickly become outdated. Without accounting for the accuracy of sampling frames between censuses, frame error can negatively impact time spent in the field and survey costs.

The PFS is implemented through a national survey in Qatar. To evaluate the accuracy of the frame and estimate content errors, information about eight characteristics of housing units was collected. Based on the accuracy of these characteristics, the time interviewers spend contacting potential respondents is adjusted, and ultimately costs associated with field operations can be controlled. This method of assessing a sampling frame could be of use to other research organizations that survey nationally representative populations in rapidly developing countries similar to Qatar and the GCC more generally.

Table 1: Number of selected blocks in each municipality

Municipality	Number of selected blocks	Percent
Doha	45	63.4
Rayan	15	21.1
Alkhor	3	4.2
Alshemal	3	4.2
Wakra	3	4.2
Adayyien	1	1.4
Umsalal	1	1.4
Total	71	100

Table 2: Final dispositions for PFS sample cases

Final dispositions	Overall		Qatari		Expatriates	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Interview						
Complete	2163	46.82	1193	55.15	970	44.85
Partial	5	0.11	2	40.00	3	60.00
Eligible, Non-Interview						
Refusals	385	8.33	322	83.64	63	16.36
No interviewer available for needed language	112	2.42	24	21.43	88	78.57
Appointment made	42	0.91	29	69.05	13	30.95
Response pending	22	0.48	18	81.82	4	18.18
Unknown eligibility, non-interview						
Housing unit is closed	860	18.61	333	38.72	527	61.28
Housing unit under maintenance	28	0.61	23	82.14	5	17.86
Unable to locate housing unit	4	0.09	2	50.00	2	50.00
Not eligible						
Out of sample	210	4.55	75	35.71	135	64.29
Vacant housing unit	172	3.72	105	61.05	67	38.95
Housing unit destroyed	22	0.48	11	50.00	11	50.00
Case not loaded	346	7.49	179	51.73	167	48.27
Case not started	24	0.52	7	29.17	17	70.83
Other	225	4.87	203	90.22	22	9.78
Total	**4620**	**100.00**	**2526**	**54.68**	**2094**	**45.32**

Table 3: Accuracy of housing units characteristics

Housing unit characteristics	Accuracy (%)
Zone ID	100
Municipality ID	100
Housing unit frame number	95.5
Street name	96.5
Street number	97.2
Building type	98.0
Kahramaa account number	90.3
Head of household nationality	96

Table 4: Accuracy of GIS coordinates

Accuracy range of GIS coordinates (distance in meters)	Percentage of housing units
< 10	25.9
10 – 14.99	20.3
15 – 19.99	16.3
20 – 29.99	17.4
30 – 49.99	9.8
> 50	10.2

7. REFERENCES

- Barrett, D. F., Beaghen, M., Smith, D., & Burcham, J. (2002). Census 2000 housing unit coverage study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 146-51. Alexandria, VA: American Statistical Association.

- Eckman & Kreuter (2013). Undercoverage rates and undercoverage bias in traditional housing unit listing. *Sociological Methods & Research,* 42(3) 264-293.

- Kish, L. (1965). *Survey Sampling*. New York: John Wiley.

- Hekmat, A., Awatef, I., & Sohaila, C. (2013). Demographic and health indicators in Gulf Cooperation Council nations with an emphasis on Qatar. J*ournal of Local and Global Health Perspectives*, 3, 1-6.

- Hogan, H. (1992). The 1990 post-enumeration survey: An overview. *The American Statistician*, 46(4), 261-269.

- O'Muircheartaigh, C. A., English, E.M., & Eckman, S.A. (2007). Predicting the relative quality of alternative sampling frames. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 551-74. Alexandria, VA: American Statistical Association.

- Valliant, R., Dever, J.A., & Frauke, K. (2003). *Practical tools for designing and weighting survey samples*. New York: Springer.

- Wolter, K. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81, 338-346.

- *Accuracy and coverage evaluation of Census 2000: Design and methodology*. (2004).U.S. Census Bureau.