RAND

Sampling with Field Burden Constraints: An Application to Sheltered Homeless and Low-Income Housed Women

Marc N. Elliott, Daniela Golinelli, Katrin Hambarsoomian, Judith F. Perlman, Kirsten Becker, Kathryn G. Anderson, Crystal Kollross, and Suzanne L. Wenzel

DRU-3057-NIDA

May 2003

CONTENTS

Tables and Figures

SUMMARY

In this paper we present a statistically efficient means of collecting a probability sample in the presence of field burden constraints. We suggest the use of disproportionate stratified random sampling as an alternative to two-stage sampling and illustrate how to account, via weighting, for the subjects' differential probabilities of inclusion. We describe our approach with reference to a study of impoverished women.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute on Drug Abuse (R01 DA11370, S.L. Wenzel, PI). The authors thank Stephanie Thompson for assistance in the preparation of this manuscript.

1. INTRODUCTION

We provide an example of how to obtain a probability sample under certain field burden constraints. In particular, we consider situations in which a survey requires face-to-face interviews at the sites from which subjects are sampled and the total number of eligible sites in the population is limited. Under these circumstances, the need to limit burden on the sites may dictate a maximum sample size per site and needs for field efficiency may dictate a minimum sample size per site. In such situations, a proportionate-to-size (PPS) sample may not be feasible if the size of sites varies substantially.

We describe one technique for such situations applied to a study of impoverished women, both housed and sheltered homeless (Wenzel, 1999). Sampling homeless populations is a challenging task and has resulted in a variety of creative approaches (e.g., Koegel, Burnam, and Morton, 1996; Burnam and Koegel, 1988; Rossi, Fisher, and Willis, 1986; Sumner et al., 2001). The goal of the study to which this approach was applied was to collect two parallel, comparable samples of sheltered homeless and low-income housed women in central Los Angeles County. In this example shelters and housing facilities vary substantially in the number of women that they house. For more information on the findings of this study see Wenzel et al. (2003) and Tucker et al. (2003).

The proposed sampling approach will apply in many other instances that are characterized by field constraints similar to the ones mentioned above, as long as the total number of potential clusters/strata is low enough for all to be included. For example, if we needed to interview a sample of teachers within schools, we might face similar challenges. A similar situation might occur for face-to-face surveys of military personnel on military bases (Buddin et al., 1999).

2. METHODS

In this section we describe how two disproportionate stratified samples were obtained for the study of impoverished women. We also illustrate how to construct and evaluate the sampling weights induced by this sampling design.

In our study we were interested in collecting two probability samples of English-speaking women aged 18-55 in a defined, central region of Los Angeles County (approximately a 15 mile radius from downtown). One sample is of sheltered homeless women; the other is of women living in Section 8 private project-based HUD-subsidized apartments. (HUD refers to the U.S. Department of Housing and Urban Development.) The second sample was intended to be comparable to the first in terms of geographic region, age, and socio-economic status (SES). Each sample is a (disproportionate) stratified random sample, rather than a cluster sample, because women are sampled from all eligible shelters and housing units. The probability of a given woman being selected is higher in a small shelter or small HUD site (i.e., Section 8 building), but these differential probabilities are known and accounted for via weighting.

The study was executed in three consecutive replicate-waves to allow us to adjust sampling proportions to satisfy sample size targets as we progressed. Each replicate was preassigned to three non-adjacent geographic sub-regions. The low-income housed and sheltered homeless samples were interviewed concurrently. The three replicates contained approximately 20%, 40%, and 40% of the anticipated sample size, respectively, in order to maximize our ability to make use of information from early replicates to modify sampling proportions in later replicates.

The construction of such disproportionate samples is made up of a series of steps: definition of the eligible population, definition of the strata, selection of a sampling mechanism within strata, and determination of stratum sizes. We describe these steps in detail below.

2.1. Definition of the eligible population

Sheltered homeless women: Our goal for the sheltered homeless sample was to achieve a sample of women representative of those living in the diverse array of temporary lodging options available within Los Angeles County. We did not employ a lengthy field screener to determine homelessness on an individual basis, but rather limited our sample to women staying in facilities with a simple majority of homeless residents. For the purpose of identifying shelters, we defined as homeless, “persons who would otherwise live in the streets or who sleep in shelters and have no place of their own to go”.

Potentially eligible temporary lodging options were emergency shelters, transitional living facilities, single room occupancy (SRO) hotels, board-and-care and voucher hotels, detox centers, rehabilitation centers, and HIV/AIDS transitional homes. We excluded facilities that limit services to persons less than 18 years old, facilities that only serve men, domestic violence shelters, and facilities whose population was not majority homeless. Additional shelters were excluded based on the following eligibility criteria:

(a) average length of stay could not exceed one year;

(b) the number of English-speaking female residents on a typical night must have been eight or more;

(c) shelters not consistently operable in 2000 must have had plans to be fully operable by December 1, 2000;

(d) fee-based programs could not charge more than the monthly grant a homeless person may receive through public assistance from such sources as General Relief (GR), Temporary Aid to Needy Families (TANF), and Supplemental Security Income (SSI);

(e) shelters could not limit services to a special needs population that our budget did not allow us to accommodate (e.g., monolingual Spanish speakers and deaf persons);

(f) shelters that had refused to participate or interfered with the sample integrity of previous RAND research efforts were excluded (three shelters);

(g) shelters for whom we could not obtain a complete telephone survey were excluded.

The number of eligible shelters decreased further when we combined separately listed eligible programs located at the same physical address. The final revision of the number of shelters was made when we excluded all shelters with fewer than eight women staying there on an average night. Of the 350 separately listed programs reported to serve homeless individuals in Los Angeles County, 57 facilities were eligible. Once in the field, we found an additional six sites to be ineligible (non-existent, domestic violence shelters, or would not allow random sampling of residents, reducing our shelter total to 51.

In addition to the shelter screening criteria, a brief field screener was used to exclude women not between 18 and 55 years of age, women cognitively incapable of completing the interview, and women who had already been interviewed by our research team.

Low-income housed women: The sample of housed women was designed to be as similar to the sheltered homeless sample as possible and at the same time representative of the experiences of impoverished women living in Los Angeles County. To efficiently locate low-income women, we limited our sample to women living in Section 8 private project-based HUD-subsidized apartments. To qualify for

Section 8 housing, a person can make no more than 50 percent of the median income for Los Angeles County (though most of those receiving the subsidy actually make closer to 30 percent of the median). We included all such apartment buildings within the study boundary reported by HUD to consist entirely of Section 8 project-based apartments not specifically designated to house elderly or disabled tenants. A further requirement was that we were able to complete a telephone verification of address and size with property managers. Eligible women were 18-55 year olds whose primary residence was one of these apartments and who were capable of completing a 1½-hour interview in English.

Of the 653 properties listed in the HUD database that were within the study area, 531 were initially excluded in the following order due to ineligibility: 228 were designated to house elderly tenants or had missing information on tenant designation, 16 had missing or incorrect address information, 54 did not actually contain any Section 8 units, 27 housed exclusively developmentally disabled tenants, 82 contained a nontrivial proportion of non-Section 8 units, and 124 had fewer than nine units. The resulting database contained 122 potentially eligible properties. It was determined by careful examination of the database and through phone-calls with property managers, that these 122 properties corresponded to 189 physical locations. We found through an exhaustive telephone survey that of these 189 sites, 69 were eligible. The remaining 120 properties were deemed ineligible because managers reported that more than 50% of units housed elderly and/or non-English-speakers or because we were unable to verify resident composition from managers. While in the field, the Survey Research Group excluded three additional sites (one outside study boundaries, two elderly), decreasing the total to 66.

2.2. Sample design: stratified versus clustered sampling

We obtained a total of 460 completes from homeless sheltered women and 438 completes from low-income housed women. Since the number of eligible shelters and Section 8 buildings was relatively few (51 shelters and 66 Section 8 buildings), we adopted a stratified sampling design instead of a clustered one. In a traditional two-stage cluster sample, only a subset of eligible shelters/ buildings would be selected, and no one would be interviewed from the remaining shelters/buildings. If there is even moderate similarity or homogeneity of residents within the same shelters/buildings (as compared to differences between residents of different shelters/buildings), this approach can lead to large design effects (DEFF), which measure the loss of statistical power due to sample design.

In the present approach, each shelter or Section 8 building represents a stratum. Although weighting and disproportionate stratification do induce some design effects, these are typically considerably smaller than those found in cluster samples. We describe the actual moderate design effects obtained in Section 3.2.

2.3. Sampling mechanism

We used disproportionate stratified random sampling to obtain parallel samples of the two impoverished women populations. If the population size of a shelter or if the number of units within a Section 8 building was smaller than a specified cutoff, all the women or units were sampled. If the population size was larger than an upper cutoff, the number of units sampled was limited to a "ceiling". If the population size was between the lower and the upper cutoff, then we sampled a fixed proportion of the population. This mechanism was used for both samples. However while for the shelters the end result was indeed a sample of women, for the housed women this scheme provided a sample of individual residential housing units. The extra step required to obtain a sample of women for the low-income housing case will be described below.

This sampling scheme requires an accurate estimate of the population size, which will be described in the next section. It also requires the specification of a set of parameters: a "floor," a "ceiling," and a sampling proportion. These parameters were specified differently for the two samples and were adjusted during the three replicates.

The reason for not using a strict proportionate probability sample is field practicality. A "ceiling" was set because we wanted to limit the burden on the very large shelters or housing facilities. A "floor" was set because it was not logistically efficient to travel to a shelter or housing facility to interview only one or two women. For statistical efficiency, the sampling procedure was as close as possible to a proportionate sample, within those constraints. Of course using such a sampling scheme has the consequence of inducing a design effect that the proportionate probability sample would not induce.

The floor constraint does not induce a large design effect. It implies over-sampling the small shelters or housing facilities, but this results in only a small number of over-sampled cases (because of the small number of women at each such site), and these are over-sampled only to a modest extent (because in this case, sampling with certainty is not a large increase from the average sampling rate). On the other hand, the presence of a ceiling induces a larger design effect in this case. The ceiling forces us to under-sample the larger shelters. This applies to a larger number of cases in the sample, and for the largest shelters, involves relatively substantial under-sampling. Therefore, individuals in larger sites have a smaller probability of being sampled.

Figure 1 shows a plot of the number of women to be sampled versus the size of a site, using our sampling mechanism. This plot also compares our sampling mechanism with the proportionate-to-size (PPS) sampling mechanism. The deviation between the straight line ("proportionate sampling") and the piece-wise line ("disproportionate sampling") represents the degree of over- or under- sampling involved. The magnitude of this deviation, in combination with the proportion of completes corresponding to a given region of the *x*-axis, determines the design effect induced by our sampling mechanism.

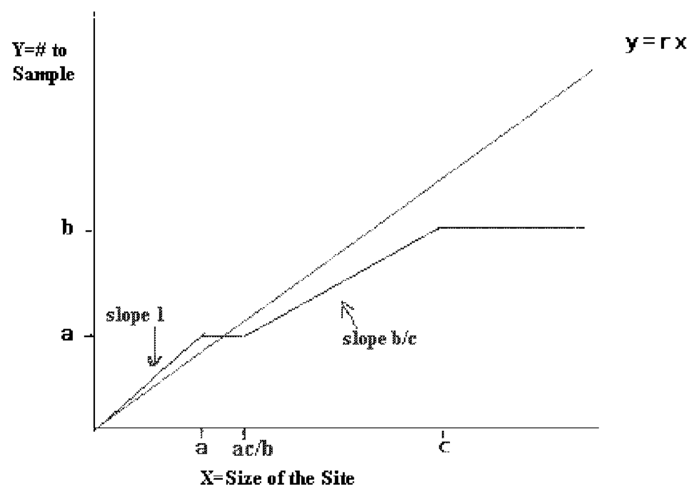

Figure 1. Rule for selecting women

From a technical point of view we can describe the sampling mechanism in the following way. The following description refers to shelters, but it applies also to the sample of units in Section 8 buildings. For shelter $i = 1, \ldots, k$, where k is the total number of eligible shelters, let Y_i be the number of women to be sampled at that shelter and X_i the number of women (expected to be) present at such a shelter.

The rule for selecting Y_i is the following.

$$Y_i = \begin{cases} X_i & \text{if } X_i < a \\ \max(a, \frac{bX_i}{c}) & \text{if } a \le X_i < c \\ b & \text{if } X_i > c \end{cases} \quad (1)$$

Figure 1 provides a graphical representation of this sampling mechanism. Given the scheme above, we can compute the probability of selecting a woman or the probability of including a woman in our sample. This probability is the same for all the women in a given shelter of size X.

For shelter i the probability of inclusion Y_i / X_i is given by the following equation.

$$\frac{Y_i}{X_i} = \begin{cases} 1 & \text{if } X_i < a \\ \max(\frac{a}{X_i}, \frac{b}{c}) & \text{if } a \le X_i < c \\ \frac{b}{X_i} & \text{if } X_i > c \end{cases} \quad (2)$$

Figure 2 shows the relationship between the probability of inclusion and the site size.

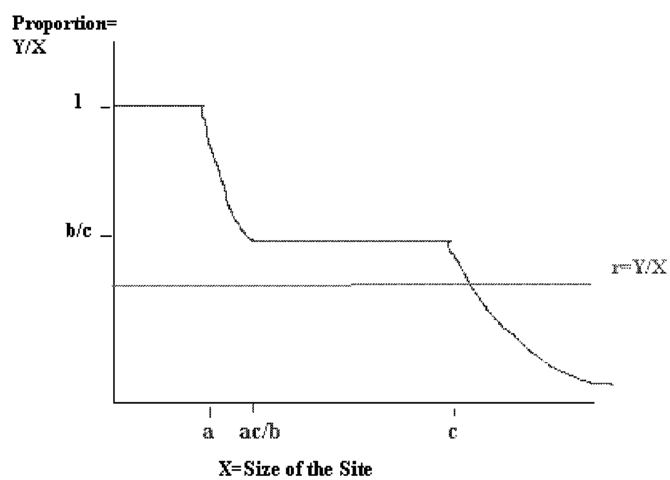

Figure 2. Proportion sampled

2.4. Setting sampling parameters

Because final measures of size (MOS) for each site could not be obtained a priori, we used two measures of size for two different purposes. An initial estimated MOS was used only for the purposes of estimating optimal sampling parameters. A second, final MOS was used as X_i in (1) upon visiting the actual shelter or Section 8 building to conduct sampling.

A simulation study was conducted using the initial (estimated) MOS to determine optimal values for the parameters *a; b; c*. This process was conducted independently for the sheltered homeless and low-income housed samples. In particular, we solved for the sampling proportion that minimized DEFF for a number of selected combinations of floor and ceiling values. We then examined the trade-offs between increasing field efficiency (high floors), decreasing field burden (low ceilings), and increasing statistical efficiency (low DEFF) to make our parameter selections. These selected parameters in turn resulted in a field sampling table that translated site MOS into sample size.

2.5. Determining the measure of size (MOS)

In what follows we describe how we obtained such an estimate both for shelters and housing facilities.

Shelters: We telephoned the managers of the shelters and asked them to provide an estimate of the number of women present in the shelter on the night before the call. From previous studies of these same shelters we knew that the managers have some tendency to overestimate how many women are in the shelters, especially for small shelters. Therefore, the estimates provided by the manager were shrunk by a multiplicative factor. Two different sets of factors were used: one for transitional shelters and one for all the other shelters. Table 1 reports the shrinking factors, which were derived from the experiences of Sumner et al. (2001).

Table 1. Shrinking factors as function of the estimated MOS

Estimated MOS	Transition shelters	Other shelters
6-9 estimated	0.41	0.30
10-14 estimated	0.48	0.35
15-20 estimated	0.72	0.53
21 + estimated	0.96	0.71

The final MOS was defined as the number of women present at the shelter, the night that the shelter was visited.

Housing facilities: In order to determine the size of the housing facilities, we called the managers and asked how many units were occupied. If the manager was not reachable, we relied on the information provided to us by HUD. Again the actual MOS was determined on the sampling and initial interviewing day. For example, unoccupied units were crossed off the list prior to sampling.

2.6. Sampling women within housing units

For the sheltered homeless case, the sampling scheme described above essentially gives a simple random sample (SRS) within each site. For the low-income housing case instead, the mechanism gives a sample of individual residential units (i.e.: apartments), some of which might contain more than one eligible woman. When this was the case, the eligible woman with the most recent birth-date was asked to participate in the study. This requires weighting housed cases proportionately to the number of eligible women residing in the sampled unit.

In essence for the housed women we used a two-stage sample:

(a) Stage 1: within each site/stratum, we sample units (disproportionately with respect to MOS);

(b) Stage 2: within each unit we take a SRS of one.

3. Results

3.1. Response rates

In this section we report the overall response and refusal rate for the study. We define the two rates in the following way.

$$\text{Response Rates} = \frac{\text{Completes}}{\text{Completes + Refusals + Other Eligibles}}$$

$$\text{Refusal Rates} = \quad 1\text{- Completion Rate} = \quad 1\text{-}\frac{\text{Completes}}{\text{Completes + Refusals}}$$

Note:

- **Ineligibles** are defined as ineligible plus Error and plus Repeater
- **Other Eligible** are defined as unlocatable.

As it can be seen in Table 2, the overall response rates were 77and 86%, with an 8% refusal rate in shelters and a 19% refusal rate in low-income housing units.

Table 2. Response and refusal rates

	Sheltered	Housed
Completes	460	438
Refusals	40	104
Other eligible	36	29
Ineligible	48	362
Response rate	86%	77%
Refusal rate	8%	19%

3.2. Sampling probabilities, weights and design effect

There are two sources that generate a design effect in this type of survey. One source is the sampling mechanism that we adopted: disproportionate sampling. The other source is non-ignorable (disproportionate) non-response. These factors make the probabilities of inclusion in the sample unequal. We compensate for the varying probability of inclusion in the final set of completes via inverse probability weighting.

The probability of being included in the sample is therefore given by the probability of being selected (2) times the probability of completing the survey given selection.

$$p(completion) = p(selected)p(completion/selected) \qquad (3)$$

The first term in the equation above is a known, predetermined quantity given by (2). The second term depends on the unknown population response rate, which may vary by respondent characteristics, and which can only be estimated through data.

In this situation, the information on non-respondents available to model non-response is very limited, so that non-response probabilities can be modeled only by site. If such a model indicates no evidence of true variation in the response rate by site, site-specific response probabilities would result in unnecessary variation in weights and would thus needlessly lower efficiency through an unnecessary increase in DEFF. Under such circumstances, a mean probability of completion given sampling would be superior to the very noisy empirical proportions by site (especially noisy for small sites).

In order to assess whether the variation among these rates was greater than that which could be attributed to chance, we ran two logistic regressions: one for the shelters and one for the housing facilities. With these logistic regressions we want to model the probability of being a responder as a function of site. In this case, both logistic regressions indicate that the response rate varies by site. The p-values for both logistic regressions are significant (< 0.001) and the c-statistics are somewhat high (0.685 and 0.709 for the shelters and housing facilities samples respectively). The F-statistic for the shelter sample is 2.53 and for the housing facilities sample is 1.78, indicating that the between units variability is 2.53 and 1.78 as large as the within unit variability.

As can be seen in Table 3, the total design effect from sample design and non-response is modest (1.26-1.31) and mostly attributable to sample design. Note also that sample design weights and non-response weights (and the corresponding underlying probabilities) are negatively correlated here.

Table 3. Sample design and non-response DEFF

	Size	Total DEFF	Design DEFF	Non-Response DEFF
Sheltered	460	1.26	1.28	1.15
Housed	438	1.31	1.29	1.06

For comparison purposes we computed the DEFF we would have obtained if we had used a clustered design instead of a stratified one. The number of eligible shelters and Section 8 buildings was 51 and 66 respectively. If we had sampled 20 institutions and completed 23 interviews per shelter and 22 per Section 8 building we would have obtained the same sample sizes of our study (460 and 438 cases in shelters and Section 8 buildings respectively). However, if we assume an intra-class correlation coefficient of 0.05, this would have implied a DEFF of $1+(23-1)*0.05=2.1$ and 2.05 for the sheltered and housed sample respectively, which would have resulted in an effective sample size only one half as large as that for a SRS of the same size and substantially smaller than the one obtained with our design.

3.3. Bias/Variance trade-off

Table 4 presents twelve month prevalence estimates for physical violence, sexual violence, illegal drug use, alcohol use, and alcohol bingeing among the sheltered homeless and low-income housed women populations, both unweighted and with final weights. The reduction in bias through weighting is approximated by the difference between weighted and unweighted prevalence estimates. Variance inflation is based on design effects and excludes effects attributable to changes in the estimated prevalence. The ratio of squared bias to change in variance allows one to assess whether the estimated bias reduction exceeds the variance inflation, which is indicated by ratios greater than 1.0 greater than

1.0 in the final row of Table 4. In such cases (three of ten cases in Table 4), the mean squared error of the prevalence estimates is improved by weighting.

Table 4. 12-month prevalence estimates for various outcomes, with and without weights

Sheltered	Physical violence	Sexual violence	Drug use	Alcohol use	Alcohol bingeing
Unweighted prevalence	0.343	0.084	0.478	0.604	0.310
Weighted prevalence	0.338	0.079	0.479	0.619	0.307
$Bias^2 \times 10^4$	0.202	0.238	0.028	2.440	0.120
$Var \times 10^4$	1.270	0.429	1.410	1.340	1.210
$Bias^2$ / Var	0.158	0.554	0.020	1.815	0.099
Housed					
Unweighted prevalence	0.134	0.006	0.146	0.648	0.162
Weighted prevalence	0.130	0.004	0.163	0.655	0.175
$Bias^2 \times 10^4$	0.171	0.036	3.000	0.484	1.660
$Var \times 10^4$	0.826	0.042	0.940	1.630	1.000
$Bias^2$ / Var	0.206	0.872	3.201	0.297	1.652

4. CONCLUSIONS

By treating residential units as strata (rather than clusters) and by imposing minimum and maximum sample sizes per residential unit, we achieved a sample design that was statistically efficient, cost-effective, and which allowed relatively unburdensome access to low-income housing structures and homeless shelters. This same approach is applicable to a variety of two-stage settings with similar constraints. In three of ten instances involving prevalence estimates, adding weights improved the MSE, correcting biases of 1.3-1.7 percentage points. While non-response varied significantly by site, the relatively moderate F-statistic for Section 8 housing units suggests that these non-response weights may be overly variable; future work should consider empirical Bayesian techniques to shrink estimated response probabilities toward the grand mean (Carlin and Louis, 2000).

REFERENCES

Buddin, R.J., Gresenz, C.R., Hosek, S.D., Elliott, M.N., and Hawes-Dawson, J. (1999). An Evaluation of Housing Options for Military Families. RAND. MR-1020-OSD

Burnam, M.A. and Koegel, P. (1988). Methodology for Obtaining a Representative Sample of Homeless Persons. Evaluation Review, 12, 117-152.

Carlin, B.P. and Louis, T.A. (2000). Bayes and Empirical Bayes Methods for Data Analysis, (2nd edition). Boca Raton, FL: Chapman and Hall/CRC.

Koegel, P., Burnam M.A., and Morton, S.C. (1996). Enumerating Homeless People. Alternative Strategies and Consequences. Evaluation Review, 20, 378-403.

Rossi, P.H., Fisher, G.A., and Willis, G. (1986). The Condition of the Homeless in Chicago. Amherst: Social and Demographic Research Institute, University of Massachusetts.

Sumner, G., Andersen, R., Wenzel, S., and Gelberg, L. (2001). Weighting for Period Perspectives in Samples of the Homeless. American Behavioral Scientist, 45, 80-104.

Tucker, J.S., Wenzel, S., Elliott, M.N., Hambarsoomian, K., and Golinelli, D. (Submitted, 2003). Patterns and Correlates of HIV Testing Among Sheltered and Low-Income Housed Women.

Wenzel, S. (1999). Drug Abuse, Violence and HIV/AIDS Among Impoverished Women (R01 DA11370). Rockville, MD: National Institute on Drug Abuse, National Institutes of Health.

Wenzel, S., Tucker, J.S., Elliott, M.N., Hambarsoomian, K., Perlman, J., Golinelli, D., Becker, K., and Kollross, C. (Submitted, 2003). Prevalence and Co-Occurrence of Violence, Substance Use and Disorder, and HIV Risk Behavior: A Comparison of Sheltered and Low-Income Housed Women In Los Angeles County.