

SPECIAL COMMUNICATION

A Method for Obtaining an Unbiased Sample of Family Medicine Patients for Research Purposes

Everett E. Logue, PhD, and Claire C. Bourguet, PhD

Introduction: Obtaining a representative patient sample for research purposes can be challenging. Classic probability sampling can be trusted, but these approaches are not always feasible; yet alternatives may introduce bias. We summarize relevant literature, the need for new approaches, and illustrate a practical hybrid approach that could consistently produce representative patient samples.

Methods: Valid approaches shift sampling decisions from fallible interviewers to less fallible reproducible processes. In lieu of the interviewer's inclination to select particular patients, we used the day of the week, the appointment time, and the sequence of the patient's last name in the alphabet to select a sample for a consent process and a survey.

Illustrative Use: Characteristics of the study sample ($n = 225$) were compared with the characteristics of the population ($N = 1964$) that had an office visit during the recruiting period. The data suggested that the study sample was highly representative of the population in this illustrative case.

Discussion: A hybrid sampling approach, in the context of a brief consent process, and a nonthreatening interview produced a representative study sample, but formal evaluation via simulation is needed to validate the hybrid approach. Convenience samples of consecutive patients should be avoided to minimize bias. (J Am Board Fam Med 2011;24:583–588.)

Keywords: Health Services, Primary Health Care, Research Methods, Sampling Studies, Selection Bias, Survey Methods, Sampling Bias

Obtaining an unbiased or a representative sample of primary care patients for cross-sectional research purposes is challenging. Resources for patient recruiting are usually limited, the research process must not interfere with patient care, and research is rarely a priority for patients. The classic procedures for obtaining an unbiased simple random sample from a patient population involve constructing a sampling frame (or list) of eligible patients. However, simple random sampling is rarely used in

large-scale surveys because data collection costs, sample sizes, and statistical power must also be considered.¹ Thus, compromises are made and combinations of sampling techniques (simple random sampling, stratified sampling, systematic sampling, or multistage cluster sampling) are used to identify the patients who will be recruited for a given cost and to achieve a specified sample size.¹ In primary care, construction and adherence to the sampling frame for a simple random sample may be compromised by real-life office practice. Unscheduled drop-in patients may be seen each day and no-shows occur, thus compromising a list based on patients scheduled the previous day. Convenience samples are used because they are “convenient” and the costs are low, but they are likely to yield biased results.

A key word search (survey sampling and primary care, surveys and primary care, sampling methods, and primary care) of the PubMed and Google Scholar databases (6/24/2010) produced a large number of hits, but the first listed papers were generally not relevant to the issue of methods for

This article was externally peer reviewed.
Submitted 8 March 2010; revised 7 December 2010; accepted 20 December 2010.

From the Department of Family Medicine (EEL), Summa Health System, Department of Family and Community Medicine (CCB), Northeast Ohio Medical University Akron.

Funding: Department of Family Medicine, Summa Health System, Summa Foundation, Northeast Ohio Medical University, Akron.

Conflict of interest: none declared.

Corresponding author: Everett Logue, PhD, Department of Family Medicine, Summa Health System, 525 E Market St, Suite 290, Akron, OH 44309-2090 (E-mail: LogueE@summahealth.org).

selecting patients or visits for cross-sectional research in primary care. We also searched family medicine journal web sites. A dozen reports over a 16-year period were found.^{2–13} Four studies compare probability and nonprobability sampling.^{2,6,10,11} The research of Cumming supports the proposition that a systematic sample can yield valid results, but the reported study was done in shopping centers rather than in primary care.² Smith et al report that a convenient treatment-specific sample (analgesia for chronic pain) yielded a demographic profile similar to that produced by a general practice population sampling scheme.¹⁰ However, the patients were offered only hypothetical rather than actual participation in a randomized trial, which changes the context. Rothemich et al describe a self-administered, office-based survey in great detail and report that the nonprobability waiting room sample was similar to the target population; however, the sample was quite large (844/1085) relative to the population, which undercuts the rationale for sampling.⁶ Feild et al report that probability sampling and nonprobability sampling of hemodialysis patients and their spouses yielded similar demographic profiles, but probability sampling was more efficient. However, the setting was outpatient specialty care, in which the patients are more homogeneous and scheduling processes are likely to be more orderly.¹¹ Woolf et al document that the consent process for a survey may introduce moderate to large selection bias, depending on the demographic characteristic under consideration.⁷ Other literature addresses issues peripheral to this paper.^{3–5,8,9,12,13} This relatively sparse literature does not adequately address hybrid sampling techniques in primary care. We illustrate a practical hybrid systematic and multistage patient sampling and recruiting scheme that adds to this sparse literature on sampling in primary care.

Description of Sampling Method

The method combined features of a systematic and a multistage sampling approach. In the context of patients and office visits, a systematic sampling scheme might involve selecting a random patient on the first day of the study (a random start) and then attempting to recruit every 5th or 10th patient (a fixed interval related to the desired sample size) across all subsequent study days.¹⁴ This assumes that there is a natural way that patients can be counted across days such as on a single sign-in

sheet. A multistage (clustered) sampling scheme might involve randomly selecting half-days when patients are seen, then appointment times within the half-days, and then all patients within the selected appointment times.¹⁴ The method used in the current study was a hybrid “systematic” and “multistage” sampling approach that was based on the day of the week, clock time, the patient appointment schedule, and the alphabet. Briefly,

- Index clock times were first selected depending on whether the day of the week was Monday, Wednesday, or Friday (Set 1) or Tuesday, Thursday, or Saturday (Set 2).
- The index clock times for Set 1 were the hours (9:00, 10:00, 11:00, etc) and the half hours (9:30, 10:30, 11:30, etc).
- The index clock times for Set 2 were 15 minutes past the hour (9:15, 10:15, 11:15, etc) and 45 minutes past the hour (9:45, 10:45, 11:45, etc).
- After taking note of the day of week (Set 1 or Set 2), the interviewer checked the time, determined from the patient appointment schedule which (adult) patient had the appointment for the next index time, and approached that patient for a brief informed consent process and an interview.
- If there was more than one patient scheduled for the index time the interviewer selected the patient whose last name came first in the alphabet. Any arbitrary alphabet rule could be used to break ties, as long as the rule is followed. If the patient uses a non-Latin alphabet, a discrete roll of dice in a cup (or a table of random numbers) could be used as a tie breaker. We assumed that an alphabet rule was more convenient than dice or a table of random numbers to break appointment ties.

Because our family medicine center patients are usually given 15-, 30-, or 60-minute appointments starting at the hour, 15 minutes past, 30 minutes past, or 45 minutes past the hour, the proposed sampling scheme fits with the flow of patients through our office. Other scheduling practices such as 10-, 20-, or 40-minute appointments could be accommodated by the proposed method. If the dominant pattern is six 10-minute appointments per hour, then half could be sampled on Monday-Wednesday-Friday and the other half could be sampled on Tuesday-Thursday-Saturday. If there

is a concern about a hidden selection bias related to specific combinations of clock times and days of the week, the index clock times (on the hour or 15 minutes after) can be alternated between Set 1 and Set 2 every other week. Post hoc statistical modeling can also be used to account for specific features of the study design.

Illustrative Use

The illustrative site was an urban family medicine resident training clinic affiliated with a multihospital health system and a medical school. Forty percent of the registered adult patients are African-American, 70% are female, 75% are unmarried, 40% have Medicaid insurance, 10% receive organized charity care, and 6% have no insurance. There has been a recent (after the study) influx of patients from North Central Asia with limited English proficiency who use a non-Latin alphabet.

Data Collection Method Used in This Example

The population and sample data described in this paper were part of a study focused on sleep habits and obesity.¹⁵ The study was conducted from mid-May to mid-July, 2009, when a student research assistant was available. Eligible patients (adults) were identified in the family medicine center waiting room. The informed consent process and the interview were conducted either in a semi-private corner of the waiting room, in the examination room, or in a nearby conference room. The local institutional review board approved the use of a one-page statement of research rather than a multipage consent document. The interview concerned an innocuous topic (sleep habits), which is reported elsewhere.¹⁵ The questionnaire was administered by the interviewer because of the relatively low educational or literacy levels of some patients.

The target population was all patients who were actually seen on days when the survey data were collected. The data for this sampling example were obtained from the electronic medical record. Prevalent characteristics are emphasized because the sample estimates are more stable.

The proportional distributions of the demographic and clinical variables for the sample and the population were calculated along with 95% (adjusted Wald) confidence intervals for each sample subgroup.¹⁶ We also used single-sample *z*-tests, based on the normal approximation to the binomial distribution and exact binominal tests when the

confidence intervals and *z*-tests yielded contradictory results. As a final check, we performed a global χ^2 goodness-of-fit test to compare the entire distribution (of age, ethnicity, etc) in the sample with the corresponding population distribution, in which the population distribution is assumed to be fixed.

Table 1 displays the demographic and clinical characteristics of the target population and the sample patients who were selected and interviewed under the clock and appointment-based the sampling protocol. In every instance except two, the adjusted Wald 95% confidence interval included the population parameter and the single-sample *z*-test yielded a large *P* value ($P > .05$), supporting the conclusion that the sample estimates and the population parameters were statistically identical. The exceptions were a possible underestimation of the frequency of normal weight patients (14.6% relative to 20.1%) and a possible overestimation of the frequency of organized charity care patients (18.8% relative to 12.5%). Snedecor and Cochran point out that the normal approximation to the binomial test can be misleading if the population proportion is 20% or less and the sample size is less than 50, which was the case for the two exceptions.¹⁷ Moreover, exact binomial tests for identity between the sample proportions and the population proportions yielded large *P* values ($P > .05$) for the two exceptions. Finally, we used χ^2 goodness-of-fit tests to ascertain whether the sample distributions were identical to the population distributions, while assuming that the population distributions are fixed. Each test yielded a large *P* value ($P > .05$), again suggesting that the sample and the population are statistically identical. The student research assistant reported that few patients (< 10%) refused to participate in the nonthreatening (sleep habits) interview. These data suggest that the clock and alphabet-based sampling approach (in the context of a nonthreatening interview) achieved a sample that is representative of the target population without increasing costs, which were fixed by the time commitment of the student research assistant to the project.

Discussion

The goal under any sampling scheme is to obtain a representative sample of the target population. Standard approaches such as attempting to recruit

Table 1. Population and Sample Characteristics

Characteristic	Population Parameter (N = 1964) Percentage	Sample Estimate* (N = 225) Percentage	95% Confidence Interval for the Sample Estimate*
Age group, years			
18 to 39	42.4	48.0	41.6, 54.5
40 to 64	44.2	41.9	35.5, 48.3
65 to 84	12.0	10.5	6.5, 14.4
85+	1.4	1.3	-0.2, 2.8
Sex			
Female	71.2	73.8	68.1, 79.5
Male	28.8	26.2	20.5, 31.9
Race/ethnicity			
European-American	53.3	55.9	49.5, 62.3
African-American	41.3	40.2	33.8, 46.5
Other	5.4	4.8	2.0, 7.6
Marital status			
Married	24.5	25.3	19.7, 31.0
Divorced/legally separated	10.7	8.3	4.7, 11.9
Widowed	2.6	3.5	1.1, 5.9
Other	62.3	64.6	58.4, 70.8
BMI group (kg/m ² interval)			
Underweight (< 18)	1.0	1.4	-0.2, 3.0
Normal (18 to 24)	20.1	14.6	9.8, 19.3†
Overweight (25 to 29)	26.2	26.3	20.4, 32.2
1° obesity (30 to 34)	21.0	20.2	14.8, 25.6
2° obesity (35 to 39)	14.7	19.7	14.4, 25.1
3° obesity (40+)	16.9	21.6	16.1, 27.1
Hypertension			
Yes	40.3	38.9	32.6, 45.2
No	59.7	61.1	54.8, 67.4
Insurance type			
Medicaid	43.2	40.6	34.3, 47.0
Commercial	18.2	18.3	13.3, 23.4
Medicare	20.1	17.9	12.9, 22.9
Organized charity	12.5	18.8	13.7, 23.8‡
No insurance	6.1	7.0	3.7, 10.3

*Adjusted Wald estimates.

†Normal approximation to binomial, $P = .02$; exact binomial, $P = .49$.

‡Normal approximation to binomial, $P = .01$; exact binomial, $P = .33$.

BMI, body mass index.

all successive eligible patients may not be practical if there are only a few patient interviewers available, the inclusion criteria are broad (all adult patients who have an office visit during some time interval), and the office is busy. Moreover, giving the interviewer latitude to select patients is problematic because friendly, healthy patients will be oversampled and grouchy, ill patients will be undersampled. The illustrative procedure deliberately limits interviewer decision-making and produced a representative, unbiased sample for about the cost of a convenience sample.

The method we illustrate shares features with both systematic sampling and multistage sampling.¹ Selecting the first patient according to an approximate random process (clock time) and then selecting patients at fixed (30-minute) intervals thereafter is analogous to standard descriptions of systematic sampling.¹⁴ However, selecting patients according to day of the week (1 of 2 possibilities), clock times within days (1 of 12 possibilities), and position of name in the alphabet (1 of N possible patients at each clock time) is analogous to multistage sampling. Based on our single global compar-

ison between the sample and the population shown in Table 1, the proposed method should produce less bias than convenience samples, but possibly more bias than true probability sampling, albeit at a lower cost than true probability sampling. A formal evaluation of the alternative sampling methods, associated data collection costs, and the final sample size that can be obtained for a given design is standard procedure for well funded large-scale surveys.^{12,13} These issues are beyond the scope of this paper. Sampling textbooks and review papers address these issues in more detail.^{3,14}

For our cross-sectional research, the theoretical sampling frame for a simple random sample was the list of eligible patients who made at least one office visit during the data collection period. However, the complete list of these patients could not be known until after the data collection period was over. Thus, a simple random was not feasible. A more complex probability design might be a multistage, clustered sampling scheme based on randomly selected weeks, days, half-days, clock times, and patient characteristics. A formal multistage clustered probability design was not attractive because it apparently would increase data collection costs and reduce the final sample size under a small fixed budget. Our nonprobability approach shares features (day, time, patient characteristic) with a multistage probability design. Investigators could chose to include these design features in their post hoc data analysis models to adjust for possible bias. We chose not to model these theoretical design effects in our substantive analyses of sleep habits because the sample appeared to be highly representative and similar to previous samples from our population.

Previous authors have addressed sampling schemes that could be used for selecting patients or visits for office-based research.²⁻¹³ Our study adds to this sparse literature on the pros and cons of alternative ways of sampling patients from primary care. More experience with practical but valid hybrid samples in primary care is needed, especially in the context of a greater emphasis on evidence-based primary care. There are no perfect sampling solutions for all circumstances. Compromises between practicality, costs, statistical power, and a tolerable level of (selection) bias are usually made.^{12,13}

The ultimate success of any sampling approach depends not only on sampling mechanics but also on inherent study risk, study appeal, the complexity

of the consent process, the length or content of any research interviews, patient compensation, and the clinical context within which the data collection occurs. Thus, we have described the clinical context around data collection for our study in sufficient detail so that readers can draw their own inferences. More work is needed on practical and valid procedures for selecting primary care samples for research and factors that encourage and discourage participation. Traditionally accepted probability sampling methods should be used wherever possible. We have illustrated a clock-based systematic and multistage sampling procedure that yielded a representative sample, and may be preferable to other nonrandom procedures after further testing.

References

1. Aday LA. Deciding who will be in the sample. In: Aday LA, editor. *Designing and Conducting Health Surveys: A Comprehensive Guide*. San Francisco: Jossey-Bass; 1996:112-42.
2. Cumming RG. Is probability sampling always better? A comparison of results from a quota and a probability sample survey. *Community Health Stud* 1990;14:132-7.
3. Mainous AG III, Hougland JG Jr. Survey sampling issues in primary care research. *Fam Med* 1991;23: 539-43.
4. Graubard BI, Korn EL. Modelling the sampling design in the analysis of health surveys. *Stat Methods Med Res* 1996;5:263-81.
5. Pfeiffermann D. The use of sampling weights for survey data analysis. *Stat Methods Med Res* 1996;5: 239-61.
6. Rothemich SF, Woolf SH, Johnson RE, Marsland DW. Evaluating the administration of an office-based health survey in a primary care practice. *J Clin Epidemiol* 2000;53:1002-12.
7. Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection bias from requiring patients to give consent to examine data for health services research. *Arch Fam Med* 2000;9:1111-8.
8. Roach KE. A clinician's guide to specification and sampling. *J Orthop Sports Phys Ther* 2001;31: 753-8.
9. Lee ML, Yano EM, Wang M, Simon BF, Rubenstein LV. What patient population does visit-based sampling in primary care settings represent? *Med Care* 2002;40:761-70.
10. Smith BH, Hannaford PC, Elliott AM, et al. The 'number needed to sample' in primary care research: comparison of two primary care sampling frames for chronic back pain. *Fam Pract* 2005;22:205-14.
11. Feild L, Pruchno RA, Bewley J, et al. Using probability vs nonprobability sampling to identify hard-to-

- access participants for health-related research: costs and contrasts. *J Aging Health* 2006;18:565–83.
12. Slora EJ, Thoma KA, Wasserman RC, Pedlow SE, Bocian AB. Patient visits to a national practice-based research network: comparing pediatric research in office settings with the National Ambulatory Medical Care Survey. *Pediatrics* 2006;118:e228–34.
 13. Binns HJ, Lanier D, Pace WD, et al. Describing primary care encounters: the Primary Care Network Survey and the National Ambulatory Medical Care Survey. *Ann Fam Med* 2007;5:39–47.
 14. Levy PS, Lemeshow S. *Sampling for Health Professionals*. Belmont, CA: Lifetime Learning Publications; 1980.
 15. Logue E, Bourguet C, Scott ED, et al. Sleep adequacy and obesity in primary care. *Obesity* 2010;18:(Suppl 2):S145.
 16. Agresti A, Coull B. Approximate is better than “Exact” for interval estimation of binomial proportions. *Am Stat* 1998;52:119–26.
 17. Snedecor GW, Cochran WG. *Sampling from the Binomial Distribution*. In: Snedecor GW, Cochran WG, editors. *Statistical Methods*. Ames, Iowa: Iowa State University Press; 1974:199–227.