COVERAGE IN SCREENING SURVEYS AT WESTAT

David Judkins, Ralph DiGaetano, Adam Chu and Gary Shapiro, Westat
David Judkins, Westat, 1650 Research Blvd., Rockville, MD 20850

Key Words: Sample Design, Rare Domains

1. Introduction

The concept of undercoverage in sample surveys reflects the notion that some units in the intended sampling universe may have probabilities of frame inclusion less than one. These reduced probabilities of frame inclusion might result from living in nonstandard or hidden housing or an aversion to being reported as a member of households in conventional housing, among other factors. Overcoverage is also possible where some persons have multiple chances of frame inclusion such as snowbirds or children of separated parents.

Coverage is important for two reasons. First, coverage is an indirect indicator of quality – just like response rates. Users may interpret the results as applicable to the entire population when, in fact, a specific segment of the population may be missed or underrepresented (e.g., homeless persons). In addition, anticipating the degree of coverage for targeted domains is vital for survey planning. Coverage below anticipated levels will result in shortfalls in interview targets. Coverage above anticipated levels will cause budget overruns.

Measurement of coverage is difficult since it presumes the existence of a more accurate information system than the survey at hand. Furthermore, it can be complicated to separate nonresponse from frame undercoverage. For many demographic surveys, it is reasonable to estimate coverage by comparing estimates based on nonresponse-adjusted weights to the independent control totals obtained from major national surveys, often from those assembled by demographers at the U. S. Census Bureau. These control totals are based on information from the most recent Decennial census, adjusted for subsequent births, deaths, immigration, and emigration.[1] The quality of these control totals can vary by year, age range, and race and ethnicity.

For this paper, we have focused on four national household surveys with important screening components. Three of these surveys used area sampling with door-to-door screening and one was a dual-frame survey (area and RDD sampling) with both door-to-door and telephone screening. These surveys are the Continuing Survey of Food Intake by Individuals (CSFII), the Third National Health and Nutrition Examination Survey (NHANES III), the National Medical Expenditure Survey II (NMES II), and the National Survey of America's Families (NSAF). These surveys were all conducted between 1986 and 1997. The sampling, screening, and weighting procedures varied across these surveys. This variation presents both opportunities and obstacles. Variation in screening procedures is good in that the different surveys can be viewed as part of a large experiment. Variation in weighting procedures is unhelpful in that the measurements of coverage are to some extent inconsistent and thus introduce a confounding factor into the experiment. Still, we think that reporting on this set of four surveys does offer some clues about the likely performance of screening procedures on future surveys.

Two methods of screening for targeted populations are common. One method is to first establish a full household roster and then make a decision about whether to retain some or all household members for full interviews. For brevity, we refer to this method as "doorstep" rostering. The other method is to ask a simple focused question on the presence of eligible persons. This method is less expensive than doorstep rostering and may also facilitate a higher screener response rate since interviewers can make selection decisions very quickly under less than ideal circumstances. We refer to this approach as "overt focused screening."

The purpose of screening is to sample members of targeted domains at higher rates than other persons. The ratio of the sampling rate for a particular group to the smallest sampling rate across all domains is called the oversampling rate for the group.

The hypothesis we will investigate here is the assertion that coverage is inversely related to oversampling. That is, domains with higher oversampling rates would be expected to be undercovered while those with low sampling rates would be expected to be overcovered. Earlier papers have provided evidence for this hypothesis. For example, the National Immunization Survey (NIS), showed substantial undercoverage for the oversampled domain of interest when overt focused screening was used (Shapiro et al., 1996, Shapiro, 1987, and Camburn and Wright, 1996).

This hypothesis rests on the theory that either some potential sample persons may guess at the consequences of the screening process and seek to avoid being classified as eligible for the survey or some interviewers may attempt to shortcut the screening and interviewing process. In either case a quick, completed but inaccurate screener interview may be the result with no extended interview. According to this theory, overt focused screening can suggest to the respondent or interviewer which responses will lead to a quick termination of the interview. If doorstep rostering is

[1] See http://www.census.gov/population/methods/usmeth.txt.

used, then it becomes much more difficult for the screener respondent to guess which responses will lead to exclusion from the sample. However, if some domains are highly oversampled, then interviewers and survey materials (such as advance letters) might inadvertently telegraph to the screener respondents the categories of interest and thereby motivate them to leave certain persons off the household roster.

In addition, interviewers may also have motivation to find some households ineligible rather than uncooperative. Field interviewers are often graded based on their nonresponse rates, but eligibility rates are not held against them. Thus, if it is clear that a doorstep respondent is belligerent and unlikely to cooperate with the full interview, then the interviewer may be tempted not to record the presence of the person in the targeted domain. Interviewers may also switch to the overt direct screening method if they know that a particular household will only be selected if a member of a particular domain is present, thereby saving themselves the effort of a full roster.

We examine our historical coverage rates to see if they provide any information for or against this hypothesis.

2. NHANES III

NHANES III involved the collection of health and nutrition data about the civilian noninstitutional population through both personal interviews and physical examinations. Data collection extended over a six-year period (10/88–10/94). Screening was conducted with a paper and pencil household roster. Central computer algorithms had pre-designated sampling rules for each sample address (Waksberg and Mohadjer, 1991). As the interviewer was completing the roster, he/she knew the age-race-ethnicity domains that were eligible for interview at a particular household, but this was (at least in theory) a secret to the screener respondent. Substantial financial incentives ($50 per person) were available for examinations, but it is not known how often these were mentioned prior to completing the doorstep roster. The screener response rate was 100% because information from neighbors was used when the occupants of the sample address were not available (7% frequency). A small staff of just 27 interviewers did all the screening work, moving from PSU to PSU.

While persons of all ages were eligible for the survey and examination, the domains with the highest oversampling rates were babies under the age of 1 year, black girls between the ages of 1 and 5, Mexican-American boys and girls between the ages of 1 and 5, and Mexican-American men and women over the age of 59. Most households containing any of these groups was retained with certainty.

Coverage was estimated by comparing preliminary sample estimators of population totals with control totals from the Bureau of the Census adjusted for Census undercount, (Montaquila, et al., 1996). Selected highlights are shown in Table 1. Overall coverage was estimated to be 91.5 percent. This coverage rate compares favorably with coverage rates for major surveys conducted by the Bureau of the Census. For example, coverage in the October 1994 CPS was 92.1 percent.

Table 1. Coverage rates in NHANES III by race, ethnicity and age

Group	Coverage (%)
Overall	91.5
White and other	
Infants 2 to 11 months	87
All ages	90
Black non-Hispanic	
Infants and toddlers 2 to 35 months	105
Children 3 to 5 years	102
All ages	98
Mexican-American	
Infants and toddlers 2 to 35 months	107
Children 3 to 5 years	103
Men 60 years and older	82
Women 60 years and older	78
All ages	94

We might try to reconcile these results with the oversampling theory by speculating that the financial and intangible incentives noted above were more important to young black and Mexican-American families with children than to older families and young white families. Other theories could also be developed. One possibility that Montaquila and co-authors note is that interviewers were trained to record ethnicity differently than is customary on most other federal surveys. For NHANES, Hispanic Americans in the southwest were classified as Mexican-American if they did not self-identify with any detailed Hispanic country of ethnic origin.[2] In other surveys including the CPS, these Hispanics would most likely be classified as "other Hispanic" instead of "Mexican-American." Yet, if this was the total explanation, then one would be hard pressed to explain the poor coverage of the elderly Mexican Americans. Another possibility is that the 1990 population control totals for Mexican-Americans were not very good. Most of the overcoverage of Hispanics was concentrated in Phase 1 which was controlled to the 1990 CPS. Phase 2 was controlled to the 1993 CPS. Coverage ratios for Phase 2 looked closer to our expectations. A special analysis of 1993 (BLS, 1994) showed that the estimate of the Hispanic population in 1993 based on the same methodology as used in the 1990 CPS was 10 percent too small. We think this was due to an underallowance for undocumented Hispanic immigrants during the 1980s.

[2] It has been noted that some descendants of the pre-Mexican-American War Spanish settlers of the Southwest do not self identify with the Mexican-American label. These are sometimes referred to as Hispanos.

(See BLS, 1986, for description of how these were estimated.) The fact that there were just 44 PSUs in Phase 1 of NHANES III might also have contributed by causing high between-PSU variance on the Mexican-American estimates. However, there were another 45 PSUs in Phase 2, giving a total of 89, which is a fairly large number of sample PSUs.

3. Continuing Survey of Food Intakes by Individuals (CSFII)

The Continuing Survey of Food Intakes by Individuals (CSFII) 1994-96 sponsored by the U.S. Department of Agriculture collected two days of food intakes using face-to-face interviews for approximately 15,000 individuals over a three-year data collection period. Like the NHANES design, the CSFII used a national stratified multi-stage sample of PSUs, area segments, and dwelling units and persons within occupied households. The goal of the sample design was to select nationally representative samples of the noninstitutionalized U. S. population within each of 40 domains defined by sex, 10 age groups, and income level (below or above 130 percent of Federal poverty thresholds). In general, low-income persons, children 5 years of age or younger, and persons 70 years of age or older were sampled at the highest rates.

The selection of persons within households was done by the interviewers at the time of screening using a set of "sampling messages" that specified the characteristics of the persons to be included in the sample. The sampling messages, which were randomly assigned to households prior to the initial contact, were designed to achieve the target sample sizes for the specified subdomains (Tippett and Cypel, 1998). The specific form of the sampling messages ranged from being "all inclusive" (i.e., sample all persons in the household) to very focused (e.g., sample only low-income males 50-59 years of age). The number and structure of the sampling messages were revised periodically to reflect mid-course adjustments of the target sampling rates. To facilitate the selection of individuals, the interviewer first prepared a roster that included the sex and age of each household member. Upon completion of the household roster, the screener questionnaire then directed the interviewer to select the person(s) specified in the sampling message. Since the sampling messages were preprinted on the screener questionnaire, the interviewers were generally aware of the characteristics of the persons to sampled prior to screening.

To permit subsequent analysis of the survey data, sampling weights were calculated and assigned to each responding individual. The weight assigned to a respondent was equal to the inverse of the probability of selecting the person for the study and included an upward adjustment for nonresponse. The nonresponse adjustment was composed of two factors, one compensating for nonresponse in the screening interview and the other compensating for nonresponse in the food intake interview. Assuming that these adjustments adequately compensate for nonresponse, the weighted count of respondents using these weights provides an estimate of the number of persons in the U.S. population that are covered in the CSFII.

Table 2 summarizes the "coverage ratio" for selected domains, where the coverage ratio is defined to be the ratio of the CSFII weighted count to the corresponding Current Population Survey (CPS) estimate. The CPS estimates reflect adjustments for the Census undercount, and also include some households that are not covered in the CSFII (e.g., households on military bases). Taking account of these factors in the calculations will have the effect of increasing the ratios shown in Table 2.

In general, coverage was not as high as NHANES coverage. Coverage for adults was about 86 percent, coverage of preschool children was 84 percent, and coverage of school age children was 87 percent. The coverage of adults age 60 and older was also relatively high at 85 percent. The domains with the lowest coverage rates included the targeted domains with the highest oversampling rates: low income adults of all races (82%), children under age 3 (77%), and low-income school children (78%). Though not an oversampled domain, the coverage rate for black adult males was also low (75%).

Table 2. Coverage ratios in CSFII for selected domains

Domain	Coverage ratio
Males, 20+ years	0.84
20-39 years	0.82
60+ years	0.85
Nonblack	0.85
Black	0.75
Females, 20+ years	0.88
20-39 years	0.86
60+ years	0.86
≤130% poverty	0.82
≥131% poverty	0.89
Children, 5 years or younger	0.84
0-2 years	0.77
3-5 years	0.91
≤130% poverty	0.86
≥131% poverty	0.84
All persons, 6-19 years	0.87
6-11 years	0.88
12-19 years	0.86
≤130% poverty	0.78
≥131% poverty	0.90

4. National Medical Expenditure Survey II (1987)

The National Medical Expenditure Survey (NMES II) was a national stratified multi-stage area sample designed to collect data on health services used and corresponding expenditures, sources of payment, and health insurance. The target sample was 14,000 participating households after 4 rounds of interviewing over the course of a year. Many different domains were designated for oversampling: blacks, Hispanics, the poor and near poor, persons 65 or older, and persons limited in the activities of daily living (characterized here as the "functionally impaired"). Precision requirements varied by domain, and, in order to ensure a sufficient sample size for each domain, a large screening sample was undertaken several months prior to the first round of interviewing. Data collection was done by two organizations, NORC and Westat, each using its own national sample of PSUs. Details on the sample design are found in AHCPR (1991).

The screener interview was designed simply for determining household composition on several demographic characteristics. There was no obvious benefit to either the interviewer or the respondent for responding in one way or another since the final sampling rates had not even been worked out at the time of screening. It thus seems reasonable to believe that coverage would be similar to coverage in a general-purpose survey such as the CPS where there is no screening or oversampling.

Method for Evaluating Coverage

Coverage for domains of analysis of age, sex, and race/ethnicity can be evaluated by comparing control totals with NMES estimates based on tabulations of the screener sample with weights adjusted only for nonresponse. Such NMES tabulations may be found in Cohen and Potter (1990). Population control totals were recently resynthesized from 1986 and 1987 estimates from the National Health Interview Survey (NHIS), the Current Population Survey (CPS), and the monthly Employment and Earnings estimates provided by the Bureau of Labor Statistics.[3] Interpolation was used to obtain an estimate of the total civilian, noninstitutionalized population for roughly January 1, 1987. The coverage estimates are given in Table 3. Note that coverage is quite comparable to CPS coverage.

However, we are particularly interested in coverage rates for oversampled domains defined in terms of race, ethnicity, age and poverty status, rather than just marginal domains. At the time of this writing in 1999, tabulations of the NMES sample for these domains using nonresponse adjusted baseweights were not readily available. As the best feasible approximation, we used the unweighted distribution of the screener sample and compared it to the weighted CPS population distribution from March, 1985. Both distributions may be found in the Methodology report (Harper et al., 1991). The ratios of these two distributions are shown in Table 4. Note that these ratios do not indicate absolute magnitudes of coverage, only relative coverage. Also, there is some confounding with screener nonresponse, but we note that the screener response rate was 91.2 percent, so that impact of the confounding is probably small.

Results

Table 3 shows marginal coverage by age, sex, and race/ethnicity. These figures show generally high coverage for all domains (roughly 90 percent or more) except for blacks and Hispanics where coverage was in the neighborhood of 84 percent.

Table 3. Coverage totals are based on average NHIS totals from 1986 and 1987

Domain	Estimated coverage rate
Age	
0-4	0.929
5--14	0.919
15-24	0.936
25-34	0.896
35-44	0.949
45-54	0.905
55-64	0.905
65+	0.916
Total	0.919
Race/ethnicity	
Hispanic	0.832
Black, not Hispanic	0.845
Other	0.939
Total	0.919
Sex	
Male	0.914
Female	0.924
Total	0.919

Despite the qualifications discussed above for Table 4, the unweighted joint Westat-NORC NMES data do allow us to consider the hypothesis specifically for sample domains. As we see from in Table 4, most of the oversampled groups are actually covered better (relative to the CPS) than the remainder group. The only oversampled domain to show lower coverage than the CPS was blacks under age 65. This was also the most intensively oversampled group, so this provides some mild supporting evidence for the hypothesis.

The NMES II data suggest that reasonable coverage can be obtained using doorstep rostering. They also suggest that one need not expect particularly low coverage rates for domains with high rates of oversampling. Other than the "Black, under 65" domain, all of the domains oversampled relative to the "Remainder" stratum" had sample proportions at or above the corresponding CPS proportions.

[3] The control totals used in the NMES II weighting were no longer readily accessible. A variety of sources had to be used because of differences in age domains tabulated in the various reports.

Table 4. Comparisons of Full NMES II screener sample and CPS distributions across sample domains

Demographic grouping	NMES screener sample distribution	CPS estimated population distribution	Ratio: sample to CPS distribution
Black,<65	.1023	.1095	0.934
Black, 65+	.0103	.0095	1.078
Hispanic, 65+	.0038	.0035	1.084
Hispanic,<65	.0687	.0688	0.998
White, 65-79	.0856	.0808	1.059
White, 80+	.0251	.0189	1.329
White,<65, poor or near poor	.1078	.0973	1.108
Other races, at or below poverty	.0068	.0055	1.238
Remainder	.5897	.6061	0.973
Total	1.0000	1.0000	

5. National Survey of America's Families (NSAF)

The NSAF was a dual-frame survey with area and RDD components. In the area component, there was door-to-door screening for nontelephone households. In the RDD component, there was screening for households containing at least one person under age 65 with oversampling of households with children and low-income households. In this paper, we focus primarily on the RDD screening. Synthetic Census-CPS estimates of the telephone population were the standard of comparison.[4] The NSAF weights used for the data shown here are preliminary weights that include the inverse of the probability of selection for a household and an adjustment for screener nonresponse. The weights do not include later weighting steps of poststratification for households and of several stages of weighting at the person level.

The first screening question concerned the presence of at least one person under the age of 65. If none were reported, the interview was immediately terminated. The second screening question concerned the presence of children. Interviews were immediately terminated for a subsample of the households without children. The third screening question concerned the income of the household. Interviews were immediately terminated for a subsample of the households that reported income above 200 percent of poverty. See Judkins, et al. (1999) for more details.

Table 5 shows estimated coverage rates by household composition among households with telephones. We experienced overcoverage of both households with children and elderly-only households. The overcoverage of the elderly-only households fits in well with the hypothesis. Since these households were immediately released from the interview, there was an incentive for the respondent to state that everyone was 65 or older. Although the screening question was asked in a manner that it is not obvious to a respondent whether the survey is interested in households with all older persons or in other households, about half the households received an advance letter prior to the RDD telephone call about the survey. From this letter, one could surmise that an answer that there are no young people would result in a shorter interview. However, there are also other plausible explanations for the overcoverage of the elderly-only households. One explanation for this overcoverage might be that the screener nonresponse rate was lower for households containing only older people. Since this group only had to answer one screening question to be considered complete, there is less opportunity for break-off's (where the respondent hangs up the phone) than for households with some younger members. Also, the elderly are often found at home and may use answering machines less.

The overcoverage of households with children, on the other hand, argues against the hypothesis and is very difficult to understand. We subsampled households without children and the interviews were about an hour, including some uncomfortable topics. Nonetheless, the prime focus of the sample (households with children) was overcovered.

Turning attention briefly to the door-to-door screening for nontelephone households in the area component of the survey, NSAF determined a lower rate of nontelephone households than the synthetic estimate based on the Decennial Census and the 1990 and 1997 CPS discussed above. Although there is considerable uncertainty about the correct number of nontelephone households, we believe that the NSAF estimate is probably closer to the correct level than either the Census or the CPS (Cunningham, et al., 1999).

Table 5. Coverage by age-composition of household in the RDD component of NSAF

Households with children	105%
Adult-only households with at least one member under age 65	88%
Households with only elderly members	125%
All telephone households	100%

[4] The control totals used in this calculation were obtained by adjusting estimates of the telephone population by household composition forward from the 1990 Decennial Census. The adjustment factors were based on ratios of CPS estimates in 1990 and 1997. This procedure was used because of large differences between the Decennial Census and the CPS in the size of the telephone population.

6. Discussion

For NMES II, NHANES III and CSFII, the doorstep roster approach to screening was used. The subsampling operation was done differently for NMES II than for the other two surveys. For NMES II, screening and interviewing were two entirely separate operations so that the screening interviewers had no interest in the outcome of the screening. For NHANES III and CSFII there was an immediate segue to extended interviews if the household was inducted. For both these surveys, the interviewer did have the capability of discovering what roster pattern would bring the household into sample. NHANES III and NMES II coverage was quite similar to CPS coverage. CSFII coverage was generally lower. It is not surprising that NMES II coverage is similar to CPS coverage given that the screener was essentially a general-purpose survey just like the CPS. Given the similarity of screening and subsampling techniques in NHANES III and CSFII, an explanation for the difference in coverage might be either the financial incentives in NHANES or the attraction of the NHANES physical exams.

The overcoverage of the excluded group of elderly households for the NSAF lends some support to the theory that low sampling rates (in this case, zero) can lead to overcoverage. However, the overcoverage of households with children which had relatively high sampling rates in the RDD component of NSAF was inconsistent with this notion.

The evidence from the other surveys is also not very clear. NHANES and CSFII are typical "screen and go" surveys where there is an immediate segue from screening to interviewing for selected eligible households. NHANES had a mixed pattern with low coverage for two of the most oversampled domains and overcoverage for the other most oversampled domains. Given this inconsistent pattern, it appears that NHANES cannot be used to support or refute the theory.

The CSFII also presented ambiguous evidence regarding the second hypothesis. Against the hypothesis, coverage was above or near average for the population 60+ that includes the oversampled domain of persons 70+. Consistent with the hypothesis, there were oversampled domains with low coverage: low-income adults, children under age three, and low-income school children. However, the results on coverage of the low-income populations could easily be due to income classification errors rather than to rostering errors. Thus, the most convincing evidence is in the low coverage of children under age three. CSFII achieved 77 percent coverage for these children.

In conclusion, it appears that a meta-analysis over a larger set of surveys will be required to shed much light on the secrets of achieving high coverage. To permit such an analysis, coverage estimates will be needed preferably with standard errors and documentation of survey particulars. Indeed survey data quality would be improved with the routine development of coverage estimates and corresponding estimates of variability.

7. References

AHCPR (1991). *National Medical Expenditure Survey: Sample Design of the 1987 Household Survey: Methods 3,* (Primary authors S. B. Cohen, R. DiGaetano, J. Waksberg), U.S. DHHS, PHS, AHCPR Pub. No. 91-0037.

BLS (1986). *Employment and Earnings.* February, 1986. Vol. 33, No. 2.

BLS (1994). *Employment and Earnings.* February, 1994. Vol. 41, No. 2.

Cohen, S. B. and Potter, D. E. B. (1990). Data collection organization effects in the National Medical Expenditure Survey. *Journal of Official Statistics.* pp. 275-294.

Camburn, D. P. and Wright, R. A. (1996). Predicting eligibility rates for rare population in RDD screening surveys. Paper presented at the Annual AAPOR Conference, Salt Lake City, May 1996.

Cunningham, P., Shapiro, G., and Brick, J. M. (1999). In-person survey methods for the 1997 NSAF. *NSAF Methodology Report,* No. 5, The Urban Institute.

Harper, T., Berlin, M., DiGaetano, R., Walsh, D. and Ingels, J. (1991). *National Medical Expenditure Survey: Household Survey: Final Methodology Report,* Westat Final Report

Judkins, D., Shapiro, G., Brick, M., Flores-Cervantes, I., Ferraro, D., Strickler, T., and Waksberg, J. (1999). 1997 NSAF Sample Design. *NSAF Methodology Report*, No. 2, The Urban Institute.

Montaquila, J. M., Mohadjer, L., Waksberg, J., and Khare, M. (1996). A detailed look at coverage in the third National Health and Nutrition Examination Survey (NHANES III, 1998-1994), *ASA SRMS Proceedings.* pp. 532-537.

Shapiro, G., Battaglia, M., Hoaglin, D., Buckley, P., and Massey, J. (1996). Geographic variation in coverage of households with telephones in an RDD survey, *ASA SRMS Proceedings.* pp. 491-496.

Shapiro, G. (1987). Interviewer-respondent bias resulting from adding supplemental questions. *Journal of Official Statistics*, pp. 155-168.

Tippett, K. and Cypel, Y. (eds.), 1998, Design and Operation: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, 1994-96, U.S. Department of Agriculture, Agricultural Research Service, Nationwide Food Surveys Report, No. 96-1.

Waksberg, J., and Mohadjer, L. (1991). Automation of within-household sampling, *ASA, SRMS Proceedings.* pp. 350-355.