SAMPLE DESIGN OF THE NATIONAL PREGNANCY AND HEALTH SURVEY

Ralph DiGaetano, John Ward, Diane Cadell, Lana Ryaboy, Westat, Inc.
Ralph DiGaetano, Westat, Inc., 1650 Research Boulevard, Rockville, Maryland 20850

KEY WORDS: sample design, oversampling, self-weighting, design effect, intraclass correlation.

Introduction

This paper describes the sample design of the National Pregnancy and Health Survey (NPHS) sponsored by the National Institute on Drug Abuse (NIDA). It also discusses some of the important issues that were addressed in the implementation of the survey. The purpose of the survey was to provide information on the extent to which mothers used licit or illicit substances (such as cigarettes, alcohol, marijuana, and cocaine) over the course of their pregnancies.

1. Target Population, Design Requirements and the Population Surveyed

The population of interest for this study was mothers of newborns in the United States. Estimates were desired at the national level and for several subgroups of mothers: blacks, Hispanics, non-Hispanic whites, and those giving birth in metropolitan areas. The sample sizes were determined so that prevalence estimates as low as 5 percent would have relative standard errors (RSEs) of no more than 30 percent for all groups of interest except Hispanics, for whom the RSEs were not to exceed 35 percent.

In order to make the survey feasible to administer, some portions of the target population were not covered. The population actually surveyed were mothers of newborns who:

1. Spoke either English or Spanish; and
2. Gave birth in a nonfederal or Indian Health Service hospital in one of the contiguous 48 states, where the hospital experienced over 200 births annually (based on 1989 data).

We estimate that the NPHS covered roughly 91 percent of the target population. The 91 percent is the cumulative effect of approximately 99 percent of births occurring in hospitals, the sample frame of hospitals representing approximately 94 percent of births in hospitals in the nation, and roughly 98 percent of the general population speaking either English or Spanish well.

2. The Sample Design

The sample design was a two-stage selection procedure within strata. Hospitals were selected at the first stage and mothers of newborns at the second.

Hospitals were grouped into three strata:

1. Metropolitan area hospitals with high Hispanic populations (Metro-High Hispanic);
2. Other metropolitan area hospitals (Metro-Other); and
3. Nonmetropolitan area hospitals (Non-metro).

The first stage of selection within strata was that of hospital clusters, called "pseudo-hospitals". A pseudo-hospital consisted of one or several nearby hospitals grouped together to achieve a predesignated minimum number of annual births in total, that would provide an efficient interviewer workload. Pseudo-hospitals were selected with probability proportionate to number of births. The second stage of selection was mothers of newborns, to be interviewed in the hospital within a day or two after giving birth. The rates of selection within hospitals were determined to achieve an approximately self-weighting sample within strata.

The basis for the sample frame of hospitals was the 1990 SMG Hospital Market Database, with number of births in hospitals in 1989 serving as the measure of size for sample selection. Within the 48 contiguous states, the frame covered approximately 97 percent of births in metro area hospitals and 83 percent of births in nonmetro area hospitals.

Stratum	Sampled number of pseudo-hospitals	Targeted sample size of mothers within pseudo-hospitals
Metro - High Hispanic	12	85
Metro - Other	21	85
Nonmetro	4	56
Total	37	XX
Sampled hospitals	60	XX
Sampled mothers	3,386	XX

The minimum size for a metro area pseudo-hospital was 1,600 births annually. For a nonmetro one, it was 1,300. These minimum sizes were based on an expected length of stay for an interviewer of 20 days in metro pseudo-hospitals and 17 in nonmetro. Even with the difference in minimum size and the exclusion of hospitals with under 200 births from the sample frame, there was an average of 2.5 hospitals per nonmetro pseudo-hospital and 1.5 hospitals per metro area hospital.

More mothers were selected in metro area hospitals than in nonmetro for reasons of cost and time

in gaining participation in the study and actually conducting the survey. The extra hospital per pseudo-hospital in nonmetro areas was quite costly in terms of hospital negotiations, travel, and training of hospital staff. Moreover, available data suggested that drug use in metro areas would be higher than in nonmetro areas. The relatively large sample sizes within pseudo-hospitals (85 and 56) were chosen because the intraclass correlation of mothers reporting drug use was expected to be small. Estimates of intraclass correlation coefficients are presented later.

A disproportionately large (in terms of numbers of births) sample of hospitals was selected from the Metro-High Hispanic stratum to ensure that an adequate sample of Hispanic mothers would be obtained. A disproportionately small sample of nonmetro hospitals in terms of number of births was selected to reduce survey costs. If, as expected, drug use was greater in metro than in nonmetro hospitals, the overall precision of survey estimates may not have been reduced, and may even have been increased as a result.

3. Data Collection and Validity: Plans and Concerns

There were five sources of data collected:

1. Interviewer Administered Questionnaire (IAQ) Data;
2. Self-Administered Questionnaire (SAQ) Data;
3. Urine Specimen test data;
4. Hair Specimen test data; and
5. Data from mother and infant medical records.

The IAQ covered background demographic information while the SAQ dealt with the questions of substance use. There were two major reasons for this distinction. First it was thought that a respondent would feel more at ease to respond accurately if the questionnaire was self-administered. Second, some states require that knowledge of drug use by expectant mothers be reported. This division into two questionnaires protected the interviewers from a potential legal conflict. The urine specimens were collected for validation purposes, the hair samples to evaluate hair as a potential validation tool in future surveys. Data from medical records were obtained for a secondary objective: investigating the association of substance use with infant outcome measures.

The general data collection plan within a hospital was to:

1. Gain hospital permission to carry out the survey;
2. Train hospital staff to carry out the survey;
3. Construct a list of mothers who just gave birth from hospital delivery logs on a daily basis;
4. Randomly select the mothers to be included in the survey; and
5. Collect the data.

There were, of course, many concerns with respect to the implementation of this plan. Concerns expressed by sampled hospitals included:

1. The burden placed on the mothers;
2. The burden placed on the hospital staff, particularly nurses; and
3. Testing urine specimens without the explicit consent of the mother.

Obtaining data from new mothers also introduced operational and other survey difficulties.

1. The brevity of hospital stays for mothers requires almost immediate contact (many hospitals have policies of discharging women who experienced no complications within 24 hours after giving birth).
2. Asking a mother to respond to a survey perhaps as soon as 6 hours after delivering.
3. The legal and social implications of asking substance users to complete the survey;
   a. Could they be made to feel protected by the steps taken to ensure confidentiality of respondents?
   b. If so, would they respond?
   c. Even if they responded, would they give accurate information?

To promote cooperation and accurate reporting, a variety of steps were taken including: the use of self-administered questionnaires; concealment of responses from interviewers; providing certificates of confidentiality to respondents; and the absence of any personal identifiers from the questionnaires or record abstracts.

Because of the concerns regarding the accuracy of self-reported data involving sensitive matters such as drug use, we explored alternative toxicological measures that might serve as validation tools. The chief measures considered were hair, meconium, and urine.

It was decided to employ urine testing because of its ease of implementation and the reliability of the measure as an indicator of drug use. However, many illicit drugs can be detected in the urine for no longer than about three days after use. Thus, a series of questions regarding substance use during the last 3 days before delivery were incorporated into the SAQ.

There were two separate issues to be addressed in considering the validation of the self-reports:

1. Were those mothers who chose not to respond different from those who did in terms of substance use; and
2. Did those mothers who responded, respond accurately?

To address these two issues, an approach developed by the Centers for Disease Control (CDC) for HIV surveillance purposes was adapted for this survey. The collected urine specimens were divided into two portions: a "blinded" portion to help address the issue of whether survey nonrespondents were different from

respondents; and a "nonblinded" portion from which test results would be compared to questionnaire responses. For the "blinded" portion, procedures were established to ensure the anonymity of the sampled woman, and her consent was not sought. All sampled mothers, regardless of whether or not they participated in the survey, were included in the "blinded" study. The identifiers of the sampled women were removed, and four demographic items (age, race, ethnicity, and method of hospital payment) were obtained for analysis purposes. The test results from these specimens cannot be linked to the questionnaire data or the respondent.

With respect to the "nonblinded" portion of the urine specimen, a request was made of the sampled mother to permit the testing of her urine specimen as part of the conduct of the survey. Where consent was obtained and the SAQ was completed, results were included in the survey database for analysis.

4. Response Rates

The response rates include hospital participation, mothers' completion of the two questionnaires, and urine specimen testing.

At the hospital level, 37 pseudo-hospitals were selected, consisting of 60 hospitals in all. Thirty nine of these hospitals agreed to the interview and nonblinded testing components of the study, a 65 percent response rate. Thirty four agreed to the full study, including the "blinded" component, about 57 percent of all hospitals selected and 87 percent of those agreeing to the interview and nonblinded components.

For those hospitals which declined to participate, the following substitution procedure was used. If the hospital represented a large proportion of the births in a pseudo-hospital (roughly at least one-third of the births), it was replaced with a comparable hospital (matching on size, state, location within state, and percent minority in the five-digit ZIP code area where the hospital was located). If a hospital was relatively small, the targeted number of mothers was obtained from the remaining hospitals in the pseudo-hospital. Of the 21 hospitals declining to participate, 13 were replaced and 8 were not. The 8 accounted for about 5 percent of the sampled mothers. This substitution rule had the potential for introducing a small amount of bias, as relatively small hospitals were not replaced while larger ones were. It was employed because of the very high cost of negotiating with selected hospitals to gain their participation in the study.

There were 3,386 mothers selected of which 46 were identified as ineligible (mainly language problems; if a child was not liveborn, the mother was also ineligible). Of the 3,340 remaining, we contacted 3,007 or 90 percent. The others were missed, either because they left the hospital before we were able to contact them, or the hospital barred access to them for health reasons.

Of those approached, about 90 percent agreed to the IAQ (or about 81% of the eligible mothers). Of those completing the IAQ, close to 97 percent completed the SAQ, or about a 78 percent response rate among all eligible mothers sampled. Table 1 shows response rates and dispositions for the IAQ and SAQ.

In the 47 hospitals participating in the "blinded urine" component of the survey, there were 2,926 sampled mothers for whom urine specimens were collected and tested. Specimens were tested for 1,622 of these specimens (55.4 percent).

The "nonblinded urine" component of the survey was carried out in all 52 sample hospitals. Of the 2,613 respondents to the SAQ, 2,461 consented to having their urine tested (94.2 percent) and 1,178 specimens were actually tested (47.9 percent of those giving consent).

The major reasons for the nontesting of urine specimens were: women unable to void; specimen not collected by hospital staff; and specimen inadvertently discarded by the hospital staff. Other reasons included a low volume of urine (if there was only enough urine for one study component, the specimen was assigned to the "blinded" component) and clerical errors by interviewers.

5. Estimated Design Effects and Intraclass Correlation Coefficients

Design effects associated with clustering (after accounting for the design effect associated with differential sample weights) and intraclass correlation coefficients were estimated for several substances (any illicit drug, cocaine, marijuana, alcohol, and cigarettes) over several different time periods (including the previous year, the three trimesters of pregnancy, and the last three days before giving birth.) These estimates appear in Tables 2 and 3, respectively, and serve as an indication of the extent to which mothers who report use of a substance are clustered within pseudo-hospitals (there was an average of 1.4 hospitals per sample pseudo-hospital).

The largest estimated design effect and intraclass correlation coefficient were 2.76 and .025, respectively, for alcohol use some time during the year prior to giving birth. These relatively small values suggest that those who report substance use are not highly clustered within pseudo-hospitals. It also should be noted that the estimated design effects for cocaine and cigarettes are all less than one (and thus the estimated intraclass correlation coefficients for cocaine and cigarettes are all negative). This indicates the high degree of variability associated with these estimates, resulting from the limited number of degrees of freedom associated with a national sample of 37 PSUs (pseudo-hospitals). However, it provides further evidence that the extent to which reported substance users are clustered within pseudo-hospitals is low.

Table 1. NPHS questionnaire response rates

	Number	As a percentage of total eligible	As a percentage of eligible subjects contacted	As a percentage of IAQ completes
Total sampled	3,386			
Ineligibles identified	46			
Total eligible	3,340	100.0		
Total eligible subjects contacted	3,007	90.0	100.0	
Interviewer administered questionnaire (IAQ)				
Complete	2,701	80.9	89.8	100.0
Refused	223	6.7	7.4	
Missed (not contacted)	296	8.9	–	
Hospital barred access (not contacted)	37	1.1	–	
Language problem	28	0.8	0.9	
Other nonresponse	55	1.6	1.8	
Total nonresponse--IAQ only	639	19.1	10.2	
Self -administered questionnaire (SAQ)				
Complete	2,613	78.2	86.9	96.7
Refused	63	1.9	2.1	2.3
Other nonresponse	25	0.7	0.8	0.9
Total nonresponse -- SAQ only	88	2.6	2.9	3.3

Table 2. Estimated design effect associated with clustering

Drug type	Used in last year	Used at 3 months before pregnancy	Used at any time during pregnancy	Used during 1st trimester	Used during 2nd trimester	Used during 3rd trimester	Used 3 days before birth
Any illicit drug	1.84	1.58	1.70	2.66	1.50	1.14	0.82
Cocaine	0.76	0.59	0.60	0.62	0.75	0.37	0.53
Marijuana	1.43	1.43	2.22	2.18	0.95	1.03	0.79
Alcohol	2.76	2.71	1.93	1.55	1.16	1.91	0.49
Cigarette	0.90	0.80	0.93	0.92	0.96	0.82	0.57

Table 3. Estimated intraclass correlation coefficients for mothers within pseudo-hospitals

Drug type	Used in last year	Used at 3 months before pregnancy	Used at any time during pregnancy	Used during 1st trimester	Used during 2nd trimester	Used during 3rd trimester	Used 3 days before birth
Any illicit drug	0.012	0.008	0.010	0.024	0.007	0.002	-0.003
Cocaine	-0.003	-0.006	-0.006	-0.005	-0.004	-0.009	-0.007
Marijuana	0.006	0.006	0.018	0.017	-0.001	0.000	-0.003
Alcohol	0.025	0.024	0.013	0.008	0.002	0.013	-0.007
Cigarette	-0.001	-0.003	-0.001	-0.001	-0.001	-0.003	-0.006