A COMPARISON OF SURVEY DESIGNS FOR ESTIMATING CHILDHOOD IMMUNIZATION RATES

William D. Kalsbeek, Kristen A. Weigle, Norma J. Allred, Pao-Wen Liu
William D. Kalsbeek, University of North Carolina, Chapel Hill, NC 27599-7400

Key Words: measurement error, nonresponse bias, cost-efficiency

I. INTRODUCTION

There are seven childhood diseases that are preventable through the use of vaccines in preschool children: measles, mumps, rubella, diphtheria, tetanus, pertussis, and polio. However, the effectiveness of vaccines for measles and pertussis are partially hindered by inappropriate vaccination coverage. In North Carolina, measles evidently spread to preschool children between 1988 and 1989. Pertussis cases doubled in 1985 from 40 cases per year to 88 cases in 1986 and tripled to 128 cases in 1987 after being stable from 1974 to 1985. Most of the reported cases were in unimmunized children under the age of five (CLEMENTS et al., 1988).

The primary objective of our design feasibility study, funded by the National Centers for Disease Control (CDC) and the topic of this study, was to investigate the cost-effectiveness of four survey designs that were devised to estimate the rate of compliance with the recommended vaccination schedule for children in Wake County, North Carolina. The broader goal, however, was to evaluate the feasibility of the four designs for local application in other geographic locations. We studied the following four designs: School Record Abstraction, Provider Record Abstraction, Parental Mail Questionnaire, and Random Digit Dialing (RDD) Telephone Survey.

This approach used in this design study provided the opportunity to compare the relative cost-efficiency of the four designs, as well as to provide the opportunity to partition and compare certain bias components of the mean square error. Cost-efficiency was assessed by fixing the total direct cost of operation and then by comparing various measures such as selected sample size (N), respondent sample size (N_1), precision expressed as the expected half-width of a confidence interval (d), and relative root mean square error (RRMSE). Nonresponse and measurement components of bias were stratified on the following five dichotomous demographic subgroups: race of mother, education of mother, residence of mother (urban/rural), number of other living child in the household and marital status.

A number of recent studies illustrate the manner in which survey cost and errors are incorporated into design decision making. Cost in these studies is generally expressed in terms of the time and effort for data collection (KENT et al., 1982). In some settings, the object is to find a mathematical solution involving cost and variance in which one is minimized while fixing the other (CHOUDHRY et al., 1985). Other approaches specify similar cost and error models but seek empirical answer by choosing that design option which minimizes error while fixing cost (or vice versa) (GROVES, 1989).

II. DESIGNS FOR SAMPLING AND DATA COLLECTION

The four survey designs we considered are described below and summarized in Table 1:

School Record Abstraction:

School Record Abstraction was a retrospective study design. The data used were a child's 12 and 24 months old immunization status recorded at school entry. The sample was drawn from all children enrolled in the first grade in both public and private schools.

A stratified two-stage design was used to produce an approximately equal-probability sample of 859 first grade children. The sampling unit in the first stage was the school, and in the second stage it was the child. An approximately proportionate first stage sample of 32 public and 3 private schools was selected with probabilities proportional to size (in number of first graders), and samples of children within sample schools were chosen by systematic sampling from alphabetically sorted lists. A completed immunization record abstract was obtained for 859 of the 875 selected children (98.2 percent response).

Provider Record Abstraction:

The target population for the approximately proportionate time-stratified sample of 1,170 children in the Provider Record Abstraction design was all Wake County births to North Carolina residents, where the children would be 12 to 35 months of age at the time of sampling. After sorting the children by their date of birth, a systematic sample of 50 children was chosen per month from birth records for October 1986 through September 1988, thus yielding a total sample size of 1,200. After deleting deaths and adoptions, the final selected sample size was 1,170 children.

Compliance data under this design were abstracted by direct data entry from the medical records of all known public and private health care providers in the county including hospitals, public health clinics, and private practice clinics for pediatricians, and family physicians Data on compliance at 12 months of age were obtained by highly trained abstractors for 891 (76.2%) of the 1,170 children in the sample. Compliance data for 24 months of age were obtained for 446 of the 583 (76.5%) children in the 1986-88 cohort who had reached their second birthday.

Parental Mail Questionnaire:

The Parental Mail Questionnaire design used the same sample of 1,170 children as chosen for the Provider Record Abstraction design. Three data gathering protocols were tested under this design, with varying sample sizes in each, as seen in Table 2. All three protocols gathered data by mail but differed according to what combination of mail or telephone solicitation was used for up to three contact attempts, which were made at three to four week intervals.

Since the TMM protocol was expected to be most expensive, it was intended to select only 10 percent of the sample for this group. For each of the three protocols, a 10% sample of nonrespondents after the third attempt was contacted by home visit. Immunization compliance at 12 months of age could be measured for all 708 respondents to the sample, for an overall response rate of about 60 percent. Of the 583 children in the 1986-88 cohort who had reached their second birthday at the time of data collection, only 324 respondents had data for compliance at 24 months of age.

RDD Telephone Survey:

The sampling units used in the RDD Telephone Survey design were telephone numbers, and the sampling frame was a list of primarily residential numbers for the eight exchanges serving Wake County which was compiled by the A. C. Nielsen Company. A simple random sample of 2,000 numbers was chosen, and residential households with a child aged 12 to 35 months were considered eligible. A set of screening questions was devised to screen for eligible households, and the immunization coverage questionnaire was the same as used for the Parental Mail Questionnaire.

As expected, the screening rate for eligible children was very low. About 70 percent of the selected sample were working residential numbers, and among residential households, only 6 percent had a child aged between 12 to 35 months. In fact, the total number of children for whom immunization information was obtained by interview was only 49.

III. COST MODELS

Included and Excluded Costs:

A measure of the direct cost of data collection for each design in this feasibility study was determined by identifying specific components of this cost, making provision to monitor these cost components during data gathering in Wake County, and then compiling these costs in the form of a mathematical model that could be used to project costs for other survey settings. The cost model for each design consisted of a fixed component and one or more variable components, as illustrated for the School Record Abstraction Design. There, the two variable components are for school- and child-level costs. Each survey design had its own cost model. Some of these components in each model were "fixed" in the sense that they are not influenced by changes in sample size (e.g., purchases of computer hardware and software), while others were "variable" since sample size would influence their size (e.g., mileage , mailing materials, and postage). A portion of the fixed costs tied to the development of such things as questionnaires, computerized data entry screen, contact forms, and training manuals were seen as heavily dependent on who collected the data (e.g., survey research firm, state health department, or volunteer organization), and were therefore not included in our cost models. Other basic resource items, such as the cost of space to work, telephone lines for interviewing, and office supplies, were also not included.

School Record Abstraction (S)

Direct cost for the School Record Abstraction design, with two variable components school- and child-level cost, was modelled as:

$C_S = C_{S0} + M_S*C_{S1} + N_S*C_{S2}$ (1)

where

C_{S0} : fixed cost, contents cost of computers, software, diskettes, training abstractors, etc.

C_{S1} : unit/labor cost per school, such as cost of hours traveling to schools, travel reimbursement, contact and select school, etc.

C_{S2} : unit/labor cost per record abstracted, includes the cost of error correction forms, comments forms, data collection, data transfer, etc.

M_S : number of schools.

N_S : number of record abstracted.

Provider Record Abstraction (P)

Direct cost for the Provider Record Abstraction design, with separate variable components providers and individual records, was expressed as:

$C_P = C_{P0} + M_P*C_{P1} + N_P*C_{P2}$ (2)

where

C_{P0} : fixed cost, contents the cost of datatape, diskettes, computers, software, training abstractors, etc.

C_{P1} : unit/labor cost per provider, such as mailing letters to physicians, travel hours to providers' offices, travel reimbursement, etc.

C_{P2} : unit/labor cost per record abstracted, such as the cost of error correction forms, contraindication forms, data search and entry, data transfer, etc.

M_P : number of providers.

N_P : number of children per abstraction.

Parental Mail Questionnaire (M)

For the Parental Mail Questionnaire design direct cost, with three variable components for each contact attempt, was determined as:

$C_M = C_{M0} + N_{M1}*C_{M1} + N_{M2}*C_{M2} + N_{M3}*C_M$ (3)

where

C_{M0} : fixed cost, such as typing for questionnaire, search of deaths and adoptions, incentive, stamps pads, datatape, etc.

C_{MK} : unit/labor cost for contact K, K = 1, 2, 3 for first, second, and third contact, such as questionnaire, envelopes, stamps, keypunch, etc.

N_{MK} : number of children in contact K, K = 1, 2, 3 for first, second, and third contact.

RDD Telephone Survey (T)

Finally, the model of direct cost for the RDD Telephone Survey design, with only one variable component for unit cost, was the following:

$C_T = C_{T0} + N_T * C_{T1}$ (4)

where

C_{T0} : fixed cost, cost of training interviewers.

C_{T1} : unit/labor cost, includes cost of phone numbers, respondent contact forms, questionnaires, prefix selection, interviewer hours, etc.

N_T : number of phone numbers with eligible respondents.

IV. MODELS FOR COMPONENTS FOR THE MEAN SQUARE ERROR

To estimate a compliance rate, defined as the proportion (P) of children meeting a compliance standard, several models were needed for various components of the mean square error of the unweighted estimate (p_D) produced using actual data from a design.

Validity Standard:

To estimate these components for members of the sample who had been selected via birth certificates required that we develop a "validity standard measure" of compliance to immunization standards at 12 and 24 months of age . These "best" measures of compliance were obtained by reconciling the compliance data we had obtained from provider records, parent mail questionnaires and home visits (with parents was the basis for constructing the latter measure). Unweighted estimates of compliance rates produced using the validity standard measure were treated as the "true" rate of compliance in the population. Validity standard data on compliance at 12 months of age were available for 1,047 out of 1,170 children in the sample (89.5%). Comparable data for compliance at 24 months of age were available for 513 of the 583 children who reached their second birthday (88.0%). No specific adjustment was made for this nonresponse (due mostly to out-migration) in subsequent analysis.

Nonresponse Bias:

The nonresponse bias for the design-generated estimate of P (p_D), reflects the rate of nonresponse in the sample and the difference in "true" compliance rates between respondents and nonrespondents, was estimated as

$$(1-N_1/N)*(X_{1T}/N_1 - X_{0T}/N_0) \quad (5)$$

<u>Measurement Bias:</u>

The measurement bias for p_D was obtained simply, defined for respondents here as the difference between the design-generated estimate of the compliance rate and the "true" rate, was calculated as

$$X_{1D}/N_1 - X_{1T}/N_1 \quad (6)$$

<u>Net Bias:</u>

The net bias for p_D was considered to be the sum of nonresponse and measurement biases. This turns out to be the difference between the design-specific compliance rate for respondents and the "true" rate for the full sample (i.e., at least those for whom a validity standard measure could be produced).

$$N_0/N * (X_{1T}/N_1 - X_{0T}/N_0) + X_{1D}/N_1 - X_{1T}/N_1$$
$$= X_{1D}/N_1 - (X_{1T} + X_{0T})/N \quad (7)$$

where

1 : Respondents with a validity standard measure.
0 : Nonrespondents with a validity standard measure.
D : Using actual data produced by the design.
T : Using validity standard data (i.e., "true" value).
X : Number of sample records indicating compliance with immunization standards.
N : Number of sample records with compliance data.

The equations just presented were used to gauge the size of nonresponse, measurement, and net biases for compliance rates obtained for the population as a whole, as well as for biases associated with various demographic subgroups; e.g., defined by the mother's race, education, and marital status at the time of the child's birth.

<u>Mean Square Error:</u>

The following model was used for the mean square error of the estimate (p_D), based on actual data from a design:

$$MSE(p_D) = Var(p_D) + Bias^2(p_D) \quad (8)$$

where

$$Var(p_D) = Deff_D \; p_T(1-p_T)/N_1 \quad (9)$$

is the variance of p_D based on actual data from N_1 respondents, $Deff_D$ is the design effect expected based on the Wake County experience, p_T is the best estimate of P using validation standard data, and $Bias(p_D)$ is the net bias of p_D, as given in Equation (7).

V. COST-EFFICIENCY: STATISTICAL EFFICIENCY FOR FIXED DIRECT COSTS

Because of relatively simple relationships among sample size, variance, and cost in these designs, the comparison of cost-efficiency among designs was done by fixing the direct cost of data collection and then comparing design efficiency using several relevant statistical measures such as selected sample size (N), respondent sample size (N_1), precision level (d), and the relative root mean square error (RRMSE).

A portion of the findings from this cost-efficiency assessment for compliance at 24 months are presented in Table 3. Design effects ($Deff_D$) were estimated from the SUDAAN program (SHAH, 1990). The "response rate" (Resp Rate) for each design in this context is a production measure that was computed as the number of respondents to divided by the number of ultimate sampling units that were chosen. Entries in Table 3 for the selected sample size (N) and respondent sample size (N_1) were determined algebraically from the previously defined cost models (Equations 1-4). The precision level (d) was computed as 1.96 times the square root of $Var(p_D)$ from Equation (9), and the relative root mean square error (RRMSE), defined here as the square root of the mean square error, from Equation (8), divided by the validity estimate of compliance (p_T=0.696 for the compliance rate at 24 months of age). Sizes of d and RRMSE, relative to corresponding measure for the School Record Abstraction design, are presented in parenthesis.

Several additional comments are needed regarding the content of Table 3. First, data are only presented for compliance at 24 months of age in Table 3, since similar comparative findings emerged for both 12 and 24 months of age. Second, only data from the MMC protocol were selected to represent the Parental Mail Questionnaire design, since MMC had been previously found to be the most cost-efficient among the three protocols that were considered. Third, negative sample sizes precluded the possibility of entries for the Provider Record Abstraction design at a $1,000 and $2,000 cost level. And finally, the cost levels chosen for this table were those thought to be most likely for the local surveys to be funded by the study sponsor (CDC).

The findings in Table 3 reveal that at all cost levels and for both d and RRMSE the School Record Abstraction design is the best design, but the order of preference among the other designs varies somewhat by cost level. Using precision (i.e., variance) of estimates as the efficiency measure, the RDD Telephone Survey option is always least preferred, due mainly to the low screening rate for eligible children. However, because of its relatively small bias, it is preferred over the Parental Mail Questionnaire design for larger budgets with their larger samples since, when RRMSE is the efficiency criterion for comparison among designs and samples are larger, bias becomes a more important consideration in design evaluation. At the $5,000 level, the Provider Record Abstraction design was the third choice among the four designs when considering precision, but it was the least preferred choice when considering RRMSE.

VI. Bias Partitioning

A more detailed assessment of the bias of estimated compliance rates from the four design options was also possible in this study. The figures in Table 3 on net bias for compliance at 24 months of age reveal that the School Record Abstraction and Parental Mail Questionnaire designs both tended to overestimate compliance in this setting, especially the latter, while the Provider Record Abstraction design is likely to understate the rate. Net bias for the RDD Telephone Survey design was negligible.

Tables 4 and 5 contain separate estimates of biases due to nonresponse and measurement for total population estimates of compliance at 24 months of age, as well as for comparable estimates linked to various population subgroups. Since a validity measure could only feasibly be produced for the sample of birth certificates that were used to generate the sample for the Provider Record Abstraction and Parental Mail Questionnaire designs in Wake County, bias partitioning was only possible for these two designs. Table 4 contains the findings for the Parental Mail Questionnaire design, and Table 5 presents the results for the Provider Record Abstraction design. Subgroups were defined from information about the mother that was available on the child's birth certificate. Equations (5)-(7) were used to produce the bias figures in these tables. The proportion responding (PROP RESP) in Column (3) of this table is not the same as the "response rate" in Table 3, since for PROP RESP the denominator of the rate was the number of sample children with validity standard data. The percentage compliance rates presented for respondents, nonrespondents, and both groups combined in Columns (4), (5), and (9) of these tables were computed using validity standard data. The rates for design respondents presented in Column (7) of each table were obtained using the actual data generated by the design.

A number of mostly predictable findings for the Parental Mail Questionnaire design emerge in Table 4. For example, one would have generally expected that parents who had more fully complied with immunization requirements would be more likely to respond in the Parental Mail Questionnaire design, thus producing a positive nonresponse bias. Moreover, one might have assumed that measurement bias would be negative, due mainly to our reliance on the ability of the parent to recall the child's immunization history. Perhaps somewhat surprising was the magnitude of this negative bias, since one might conjecture that the tendency to forget immunizations would be partially offset by a certain amount of pressure to overreport immunizations to avoid the appearance of negligence insofar as the welfare of their child is concerned. This particular finding may indicate that parents did at least make a serious effort to provide accurate information. Finally, one might also have generally expected both biases to be higher for those subgroups that are more likely to have a higher percentage of the "disadvantaged" (i.e., non-white, low education, urban, mothers who had one or more living children, and single mothers), who are generally less likely to respond. The absence of a substantial differential in measurement error for the subgrouping according to the number of other living children may reflect the compensating tendencies of novelty and experience. The first-born may benefit from greater attention to the matter of immunization but less experience on the part of his parents in fully accomplishing the task, while those of high birth order may benefit from more experienced parents but also miss out because they pay less attention to the matter of immunization.

The patterns of biases for the Provider Record Abstraction design in Table 5 were somewhat different than those seen in Table 4 for the Parental Mail Questionnaire design. First, in contrast to the positive net bias with the parent-oriented design, the net bias for the provider-based design is clearly negative, caused mainly by the virtual elimination of nonresponse bias. Second, the negligible amount of nonresponse bias and minor differentials among subgroups can probably be explained by the more random nature of nonresponse and by consistently less nonresponse among all subgroups in the design. With seven of ten subgroup biases being positive, there was a slight hint of overstatement due to nonresponse, which is consistent to the findings from the Parental Mail Questionnaire design. Finally, measurement biases were still negative, although their magnitude was less and the patterns among subgroups differed from those previously seen in Table 4. Comparison of measurement bias among subgroups indicated consistent but opposite differentials involving the "disadvantaged" from what were observed for the Parental Mail Questionnaire design. The generally lower measurement bias for these subgroups may be due to the higher percentage of immunizations that their members received from public health clinics, where recordkeeping was more standardized and compliance histories more completely formed.

VII. DISCUSSION

While the findings of this study help to shed some light on the relative feasibility of the designs we considered, a number of significant limitations must of the study be noted. The first of these pertains to the notion that, based on these findings, the School Record Abstraction design is the most cost-efficient of the four designs, and it aptly illustrates how one cannot always simply take the numerical findings of a study like this one at face value. At issue here is that a school-based design, employing a sample of first graders, has a fundamental limitation in that it does not produce current immunization compliance information for 12 and 24 months of age. Instead, it produces measure of past compliance among the five and six year old children currently enrolled in school, since the data it uses are for immunizations that were administered several years earlier. The generally unknown effects of migration and the unpredictable implications of health policy change during the four year lag period call into question the feasibility of this design beyond our particular application, especially when both target population and immunization standards are fluid.

Also clearly at issue is the adequacy of the "validity standard measure" that was used in the study. While provider records and parents (by mail and by visit) were used as resources for arriving at this measure, not all three sources were available for the members of the original sample for whom a validation measure was produced. For example, 47 percent of the validation measures for 24-month compliance were available from only one of the three source (mostly provider records), which meant that for these children no reconciliation to better establish "truth" was possible. To illustrate the effect of this limitation, consider the 14.4 percent total population estimate of nonresponse bias for the Parental Mail Questionnaire in Table 4. There, the validated (or "true") compliance rate of 84.0 percent among respondents (in column 4) was computed from data on 324 children, 20 percent (64) of whom had validity data from only one source (mostly the parental questionnaire), and the validated compliance rate of 45.0 percent for nonrespondents (in column 5) was computed from 189 children, 93 percent (176) of whom had validity data from one source (mostly provider records). The likely implication on this particular estimate of bias would be for the validated estimates of the compliance rate for both respondents to be slightly underestimated and for nonrespondents to be more severely underestimated, thus making the 14.4 percent reported nonresponse bias too high. Other estimates of bias in Tables 4 and 5 would have been similarly affected, thus probably altering the magnitude of estimates but not necessarily our conclusions concerning comparisons among estimates.

Finally, an issue concerning the adequacy of the cost models used for cost-efficiency comparisons must also be raised. For example, the resource needs reflected by the these models were limited to the direct cost of the survey. A large part of the fixed costs (e.g., for planning, management, and analysis) was excluded. Had they been included, cost differentials among designs might have been less and the importance of cost in finding the "best" design might have diminished somewhat. A second limitation is that some of the more intangible difficulties faced in using the School Record Abstraction design (e.g., sampling school records) were not fully reflected in either the cost or error models. And finally, the cost models for general application were based on experience in Wake County, North Carolina. Development based on experience elsewhere might have led to different unit costs than used here.

In conclusion, findings from this study may provide a useful glimpse into the relative feasibility of four commonly used survey designs within the context of a particular health application. Costs and a variety of other statistical data enabled us to do a comparative assessment of cost-efficiency and to closely examine two of the major components of survey error. Yet, while a preference seems apparent, we come away with the sense that significant intangibles beyond those manifest in our models and calculations must be considered before making the final choice.

REFERENCES:

CHOUDHRY G. H., LEE H., and DREW J. D., Statistics Canada, "Cost Variance Optimization for the Canadian Labor Force Survey," ASA Survey Proc. 1985.

CLEMENTS D, WILFERT C, MACCORMACK JN, WEIGLE KA, DENNY F., "Pertussis Immunization in Eight Month Old Children in North Carolina", 1988.

KENT L. TEDIN and C. RICHARD HOFSTETTER, "The Effect of Cost and Importance Factors on the Return Rate for Single and Multiple Mailings," Public Opinion Quarterly Vol. 46:122-128, 1982.

GROVES R. M., Survey Errors and Survey Costs, John Wiley and Sons, New York, 1989.

SHAH, B.V., Software for Survey Data Analysis (SUDAAN) Version 5.30, Research Triangle Institute, P. O. Box 12194, Research Triangle Park, NC 27709, 1990.

Table 1. Summary of Four Survey Designs Used:

Design	Mode	Sampling Unit(s)	Source of Compliance Data
School Record Abstraction	Abstraction	School; Student	School Records
Provider Record Abstraction	Abstraction	Birth Certificate	Provider Records
Parental Mail Questionnaire	Mail	Birth Certificate	Parent
RDD Telephone Survey	Telephone	Telephone Number	Parent

Table 2. Comparison Data Collection Protocols for the Parental Mail Questionnaire Design

Proto-col	Selected Sample Size (%)	Contact			# of Respondents (by Age)[2]	
		1st	2nd	3rd	12 months	24 months
MMT	526(45%)	Mail & Postcard[1]	Mail	Telephone	332	150
MMC	527(45%)	Mail & Postcard[1]	Mail	Certified Mail	305	139
TMM	117(10%)	Telephone	Mail & Postcard[1]	Mail	71	35
	1,170(100%)	--	--	--	708	324

1 Postcard sent as follow-up reminder to initial mailing
2 Number of respondents with data, indicating either compliance or non-compliance.

Table 3. Comparison of Statistical Measures Among Four Survey Designs for Fixed Total Direct Cost: Estimating Compliance at 24 Months of Age*

Design	$Deff_D$	Resp Rate	Bias(p_D)	$1,000 N	$1,000 N_1	$1,000 d	$1,000 RRMSE	$2,000 N	$2,000 N_1	$2,000 d	$2,000 RRMSE	$5,000 N	$5,000 N_1	$5,000 d	$5,000 RRMSE
School Record Abstraction	1.42	0.978	0.021	251	246	0.0685	0.0586	976	955	0.0348	0.0395	3,151	3,083	0.0193	0.0333
Provider Record Abstraction	1.00	0.762	-0.028	--**	--	--	--	--	--	--	--	522	398	0.0452 (2.34)	0.0521 (1.56)
Parental Mail Questionnaire(MMC)	1.00	0.579	0.074	155	90	0.0950 (1.39)	0.1271 (2.17)	393	228	0.0597 (1.72)	0.1150 (2.91)	1,106	640	0.0356 (1.84)	0.1095 (3.28)
RDD Telephone Survey	1.03	0.025	-0.009	802	20	0.2046 (2.99)	0.1505 (2.57)	1,764	43	0.1395 (4.01)	0.1031 (2.61)	4,650	114	0.0857 (4.43)	0.0641 (1.92)

* Measure divided by comparable measure for School Record Abstraction design given in parentheses
** Not possible to collect data for this level of fixed cost
Resp Rate = Response rate N = Selected sample size N_1 = Respondent sample size = N * response rate
d = precision level = $1.96\sqrt{Var(p_D)}$, where $Var(p_D) = Deff_D(p_T(1-p_T))/N_1$
RRMSE = $\sqrt{(Var(p_D) + Bias^2(p_D))} / p_T$, where p_D = point estimate of compliance rate for actual data from design

Table 4. Biases for Estimated Rate of Compliance at 24 Months of Age (Parental Mail Questionnaire, MMC)

DEMOGRAPHIC SUBGROUPS	SAMPLE SIZE		NONRESPONSE BIAS				MEASUREMENT BIAS		NET BIAS	
	VALID. TOTAL	DESIGN: MMC	PROP RESP	RESP VALIDITY	NON-RESP VALIDITY	BIAS	RESP. MMC	BIAS	OVERALL VALIDITY	BIAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All Children	513	139	63%	84.0	45.0	14.4	77.0	-7.0	69.6	7.4
Mother's Race:										
white	371	115	71%	86.7	45.8	11.8	80.0	-6.7	74.9	5.1
non-white	142	24	42%	71.7	43.9	16.1	62.5	-9.2	55.6	6.9
Mother's Education:										
<= high school	202	37	46%	78.5	35.8	23.0	62.2	-16.3	55.4	6.8
> high school	311	102	74%	86.1	57.5	7.4	82.4	-3.7	78.8	3.6
Mother's Residence:										
rural	204	66	69%	84.4	42.9	12.8	75.8	-8.6	71.6	4.2
urban	309	73	59%	83.6	46.0	15.3	78.1	-5.5	68.3	9.8
# of Other Living Children:										
none	244	65	61%	87.3	50.0	14.4	80.0	-7.3	73.0	7.0
one or more	269	74	65%	81.0	40.0	14.5	74.3	-6.7	66.5	7.8
Mother's Marital Status:										
married	426	130	70%	84.9	45.3	11.9	78.5	-6.4	73.0	5.5
single	87	9	30%	73.1	44.3	20.2	55.6	-17.5	52.9	2.7

(1) Total No. of selected sample with validity standard data = N (2) Number of MMC respondents = N_1
(3) Proportion responding among those with validity data for Parental Mail Questionnaire = N_1 / N
(4) Validity rate among respondents = X_{11} / N_1 (5) Validity rate among nonrespondents = X_{01} / N_0
(6) Nonresponse bias = [1 - (3)] * [(4) - (5)] (7) Respondent compliance rate (MMC) = X_{1D} / N_1
(8) Measurement bias = (7) - (4) (9) Overall validity rate = $(X_{01} + X_{11}) / N$
(10) Net bias = (7) - (9)

Table 5. Biases for Estimated Rate of Compliance at 24 Months of Age (Provider Record Abstraction)

DEMOGRAPHIC SUBGROUPS	SAMPLE SIZE		NONRESPONSE BIAS				MEASUREMENT BIAS		NET BIAS	
	TOTAL	RESP	PROP RESP	RESP VALIDITY	NON-RESP VALIDITY	BIAS	RESP PROVIDER	BIAS	OVERALL VALIDITY	BIAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All Children	513	446	87%	69.5	70.1	-0.1	66.8	-2.7	69.6	-2.8
Mother's Race:										
white	371	320	86%	74.7	76.5	-0.2	71.3	-3.4	74.9	-3.6
non-white	142	126	89%	56.3	50.0	0.7	55.6	-0.7	55.6	0.0
Mother's Education:										
<= high school	202	185	92%	55.7	52.9	0.2	54.1	-1.6	55.4	-1.3
> high school	311	261	84%	79.3	76.0	0.5	75.9	-3.4	78.8	-2.9
Mother's Residence:										
rural	204	176	86%	72.2	67.9	0.6	69.3	-2.9	71.6	-2.3
urban	309	270	87%	67.8	71.8	-0.5	65.2	-2.6	68.3	-3.1
# of Other Living Children:										
none	244	210	86%	72.4	76.5	-0.6	70.0	-2.4	73.0	-3.0
one or more	269	236	88%	66.9	63.6	0.4	64.0	-2.9	66.5	-2.5
Mother's Marital Status:										
married	426	364	85%	73.1	72.6	0.1	70.1	-3.0	73.0	-2.9
single	87	82	94%	53.7	40.0	0.8	52.4	-1.3	52.9	-0.5

(1) Total No. of selected sample with validity standard data = N
(2) Number of respondents for Provider Record Abstraction = N_1
(3) Proportion responding among those with validity data for Provider Record Abstraction = N_1 / N
(4) Validity rate among respondents = X_{11} / N_1 (5) Validity rate among nonrespondents = X_{01} / N_0
(6) Nonresponse bias = [1 - (3)] * [(4) - (5)]
(7) Respondent compliance rate (Provider Record Abstraction) = X_{1D} / N_1
(8) Measurement bias = (7) - (4) (9) Overall validity rate = $(X_{01} + X_{11}) / N$ (10) Net bias = (7) - (9)