Approximation of Variable Costs for the National Health Interview Survey

Iris Shimizu, Feng Lan, National Center for Health Statistics
Iris Shimizu, NCHS, Room 915, 6525 Hyattsville, Maryland 20782

Key Words: Cost model, Survey design research

1. Introduction[1]

The National Center for Health Statistics (NCHS) conducts the National Health Interview Survey (NHIS) to produce general health information about the resident civilian non-institutional population of the U.S.A. The U.S. Bureau of the Census (BOC) is the primary data collector for the NHIS. A new sample design is implemented following each decennial census. The NCHS is conducting research for the next sample redesign which is currently targeted for implementation in 2005-2014. Efficient sample designs require information about the variable survey costs - that is, costs which increase with increases in sample size at each sampling stage. This paper discusses the data and processes used to approximate annual variable costs for the 1999 NHIS which can be used to evaluate alternative sample design options for the 2005-2014 sample.

An objective of the redesign is to improve the reliability of population subdomain estimates. Hence, there is interest in continuing to over sample the targeted populations by use of a screening subsample in which all housing units (HUs) are screened but only those occupied by a member of the targeted population are retained in the sample. Thus, the goals of the cost model include the estimation of the ratio of costs for a screened-out HU to the cost of a screened-in HU.

2. An overview of NHIS survey design and operations which affect costs

The NHIS uses a multistage sample of housing units. The primary sampling units (PSUs) in the 1995-2004 design are counties (county equivalents) or groups of counties. The secondary sampling units (referred to as segments, hereafter) are clusters of housing units in either an area frame or a permit frame. In areas where government units issue and maintain building permits, dwelling units built since April 1, 1990, are subject to sampling only from a permit frame and are out-of-scope in the area frame. All other HUs (including HUs built in non-permit areas since April 1, 1990) are included in an area frame. The ultimate sampling unit consists of a cluster containing four HUs or equivalent, based on the 1990 Census information. The expected number of HUs included in NHIS annually from any segment is 8 or 12 for

---

[1] The opinions expressed in this paper are those of the authors and not necessarily those of the National Center for Health Statistics.

the area frame and 4 for the permit frame. In 1999, the field work for NHIS included 358 PSUs, 6,025 area segments, 1,121 permit segments, and 72,487 HUs.

The BOC groups the sample segments into weekly assignments for their field representatives (FRs). To minimize FR travel required to collect an assignment's data, each assignment typically includes 2-3 segments and about 20 HUs within the same county, where possible, and otherwise in the same basic PSU component (an area which belongs to the same PSU for all surveys fielded by the BOC). The assignments are distributed across FRs and across the weeks in the year (to permit quarterly estimates). FRs are given 15 days, starting with the Monday of the assignment week, to complete each assignment. When feasible, the BOC attempts to avoid giving assignments in consecutive weeks to an individual FR. Assuming an average of about 20 HUS per assignment, there were an estimated 3,476 assignments in the 1999 NHIS.

Annually, FRs visit permit office to obtain lists of new construction. They also visit about 30 percent of the area segments to list or update the addresses in those segments. Sampled HUs from these lists are targeted in subsequent months of NHIS data collection.

A screening sample was added to oversample black and Hispanic persons. Prior to data collection, but after the address listing operation, some of the addresses from each area segment are randomly assigned a code of S (screening) and the rest are assigned an I (interview) code by the BOC staff. The FRs attempt to conduct the complete NHIS interview with households at every address that is assigned an I code. For every household at addresses with a code of S which also contains a black or Hispanic person, the household is retained in the sample and the FR completes the remainder of the interview. If an S sample household contains neither a black nor Hispanic person, then the household is not retained in the sample, and the FR does not complete the interview.

3. An overview of cost model format for NHIS

Because the NHIS uses a multistage sample, a simplified overall cost model for the NHIS may be expressed in the form:

$$C = C_0 + \sum_h C_h\, n_h \qquad (1)$$

where

C = Total survey cost,

C_0 = Overhead (fixed costs),

h is a subscript that denotes a sampling or operation

stage ($h = 1, 2, 3$ correspond to PSU, segment, and HU, respectively),

C_h = Cost of including an additional unit in the sample at the h-th stage. When h is a cluster (PSU or segment), this cost excludes the cost of units sampled within the cluster.

n_h = number of sample units selected at the h-th selection stage.

The C_hs are the variable costs which we attempt to estimate in this report. The fixed costs in C_0 are ignored, here.

The cost C_h for each level may consist of several classes of costs so that it may be formulated as the sum:

$$C_h = \sum_i C_h(i) ,\qquad (2)$$

where $C_h(i)$ is that amount in the i-th cost class which is allocated to the h-th sampling stage. In the NHIS, the costs associated with each sampling stage do indeed come from multiple sources. Some of the cost classes that are typical for simple three stage sample surveys are as follows:

For PSU level costs:
- Costs of retaining and maintaining a sufficient number of FRs to collect the data in individual large PSUs and an average number of FRs per PSU to collect the data in smaller PSUs that do not have enough data collection assignments to keep a FR busy.
- Cost of travel between PSUs for survey work. (Here, travel between PSUs is defined as travel between consecutive trip stops, at least one of which is a sample HU or segment, that is not in the same PSU.)
- Cost of FR time for the travel between PSUs.

For segment level costs:
- Amortized cost of the initial listing of HUs (old construction) within the segments.
- Average cost per segment for address listing and sampling of addresses done annually.
- Cost of travel inside a PSU which is between segments. (Here, travel between segments is defined as travel between consecutive trip stops, at least one of which is a sample HU, that is in the same PSU but in different segments.)
- Cost of FR time for travel between segments inside a PSU.

Segment level costs exclude cost of travel between PSUs.

For HU (case) level costs:
- Costs of travel inside a segment which is between consecutive trip stops (at least one of which is a sample HU) that are both in the same segment.
- Cost of FR time for travel inside a segment.

- Cost of communication with the HU occupants (includes cost of postage for mailing introductory letters and phone calls).
- Cost of FR time to interview the contacted occupants.
- Miscellaneous costs - cost of all other FR time for activities related to data collection and reporting, including: Computer use (to gather data such as phone numbers and ownership of sampled HUs), completing signed permission forms (documenting respondents' permission to conduct the interviews), cleaning up computer files, reporting refusals or other special HU cases to supervisors as needed, and camping out (work time spent waiting for HU occupants to return home).

Except when an HU is the only one remaining incomplete in an assignment, HU costs exclude travel between segments and PSUs.

The simple cost model just described assumes that data from each PSU are collected during one visit to the PSU. However it is noted that multiple FR assignments are made annually in individual NHIS PSUs and those assignments are spread across the year to facilitate the production of quarterly estimates. Hence, an alternate cost model was considered for the NHIS in which FR assignments replace PSUs.

4. Data available for approximating cost components

Most of the cost data for the model came from administrative records. Most of those came from reports generated for the 1999 NHIS from the BOC's Cost and Response Management Network (CARMN), a system which was first implemented in late 1998. CARMN reports costs separately for individual surveys conducted by the BOC. The FRs are requested to report to the Network every day that they work on a BOC-conducted survey and give their hours, miles and other expenses. For each survey, the CARMN provides cost data separately by BOC's regional offices and by survey task. The listed tasks for NHIS are initial training, re-interview and listing check, observation, SFR conference refresher training, interviewing, listing and segmenting, office work, and "other task codes." For each task, CARMN gives details in terms of hours (with associated salaries and benefits), miles and mileage expenses, per diem, other reimbursements, phone, and "other costs."

In addition to CARMN data, the BOC data also provide figures for the annual cost of laptop computers used by FRs in the NHIS, which is a computer assisted personal interview (CAPI) survey. BOC also reported its National Processing Center's cost for preparing folders (with maps and instructions) for listing addresses in segments. Because an advance letter is sent to sampled addresses, we added 1999 postage (at $0.32 each) costs for those mailings.

Some other office and headquarters costs (such as costs associated with selecting the samples of area segments

within sampled PSUs) should probably be included in the variable costs, but BOC was unable to break out such costs separately. Hence, such costs are omitted from this study. Also ignored, here, are the costs of recruiting field representatives (FRs) for NHIS because the BOC includes those expenses in their overhead due to the fact that the BOC hires and trains FRs for potential work on all of the BOC-conducted surveys and not specifically for the NHIS.

Some anecdotal information was used in estimating costs. Limited information about field activities that contribute to costs was obtained via a few field trips in which FRs were shadowed as the FRs worked at NHIS data collection. Information on what is "typical" in field operations was obtained from the shadowed FRs and field operation supervisors at headquarters.

5. Allocating variable costs

Those known costs which were associated entirely with a single survey level were allocated first. Starting with the costs items not kept in the CARMN system, the cost of the FRs' laptop computers was allocated to the assignment level because it is an expense of equipping the FRs and because the number of FRs working on the survey is probably correlated somewhat with the number of assignments. The National Processing Center's expenses for segment listing and sampling were assigned to the segment level costs. The postage costs were case level expenses.

Turning to the CARMN data, the total expenses for some CARMEN listed tasks could be allocated to a single survey level. The tasks for initial training, reinterview and listing check, observation, SFR conference, and refresher training are all aimed at training, supervising, and nurturing the FRs. Hence, the expenses for those five tasks were allocated entirely to the assignment level. The expenses for the "office work" and the "other expenses" tasks were allocated to "fixed costs" with the exception of costs for office staff time spent on segment listing activities. It was noted that for each segment listed in the year, BOC budget estimates assumed that 10 minutes of office staff time would be spent checking in and reviewing segment listings returned by FRs. Hence, we allocated the cost of 10 minutes of office labor to the segment level for each segment listed.

The expenses for the field work (the interviewing and the listing and segmenting) tasks were split among the three survey levels included in our cost model. For that splitting, we used the detailed expenses listed in CARMN for these tasks. All of the per diem costs were allocated to the assignment level because per diem means the FR was probably traveling away from the PSU in which the FR resided.

The CARMN item "other reimbursements" was reported to be used mostly for purchase of special survey or promotional materials, none of which is given to respondents. Hence, the expenses for this item belong to

the PSU or assignment level.

Regarding phone costs, it is known that FRs use phones to complete some interviews. It is also reported that FRs use their phone lines with laptop modems to download and upload assignment information. Toll-free lines are not used for those transmissions. Most regional offices (ROs) do have toll-free lines for FRs to use when consulting the ROs, but some do not yet have such numbers. Thus, in CARMN's interview task, phone expenses belong to both the case and assignment levels. Because no information is available for dividing that cost, it was decided to allocate half to each level in the cost model. In CARMNS's "listing and segmenting" task, the phone costs were allocated entirely to the segment level because all activity is at the segment level and typically does not involve contacts with HUs.

The costs of mileage and labor in CARMN's "listing and segmenting" task were allocated to the segment level because is it is reported that the FRs typically do their listing work when they are in the area of their listing for other business. They probably do not travel between PSUs just for the purpose of segment listing.

The miles and labor for the interviewing task were assumed to be split among the three survey levels because travel is required to reach the units at each level. This allocation was done in three steps:

- First, the miles were allocated to the three levels.
- Next, the travel hours for each survey level were estimated by dividing the miles allocated to each level by driving speeds assumed for the level. These time estimates ignore the unknown amount of time spent using mass transit, such as air travel. It is known, however, that the amount of time spent in mass transit was small relative to driving time.
- All non-travel time was then allocated to the case level for such activities as knocking on doors, conducting interviews, data editing, transmitting data to regional offices, and so forth.

Because there are no data readily available on the distances between survey units and the numbers of trips made to those units in the current survey, all distances and trip numbers had to be estimated for the cost model. Starting with the miles traveled between cases and segments, it was assumed that FRs attempted every case in their assignment before revisiting any case, and that first circuit of cases was accomplished on a single trip to the assignment. After that first circuit, it was further assumed that FRs visited all cases not completed on the first attempt within a given segment before making third visits to any case in that segment, and so fourth for subsequent rounds of visits to that segment's sample cases. In hopes that simple numbers would keep model calculations easy without being completely unrealistic, the distances between

adjacent sample HUs within area segments were assumed to be one mile and two miles in self-representing (SR) and non self-representing (NSR) PSUs, respectively, with an average of 1.36 miles over all PSUs. The SR and NSR PSUs were used as surrogates for urban and rural areas because sample counts were readily available by SR and NSR status while such counts were not known for urban/rural status. It is known that SRs are generally more densely populated and, thus, distances between dwellings in SR PSUs are generally shorter than in NSR PSUs.

Because cases in permit segments are not clustered like the cases within area segments, we assumed that the distances between permit cases were the same as the distances between area segments and that the units in these combined groups (area segments and permit cases) were scattered evenly throughout the assignment's PSU component. Under that assumption, the average shortest distance between adjacent sample units was estimated by

$$\sqrt{\frac{Area\ in\ PSU\ component}{No.\ of\ permit\ cases\ +\ no.\ of\ area\ segments}},$$

where, for simplicity in absence of area size information, we used 1,000 and 2,000 square miles to be the average area per component in SR and NSR PSUs, respectively. The resulting estimates were 18 miles and 27 miles in SR and NSR PSUs, respectively.

Distances between incomplete cases in a segment increase as cases are completed and removed from the later visit rounds. Hence, to estimate total miles traveled per case, we assumed a distribution for the sample cases by the number of attempts required to complete them. The cumulative percent distribution of cases and its complement are shown in Table A. Information obtained in an earlier study by Kalsbeek et al (1994) on numbers of attempts required to complete cases by outcome type was not readily available. Hence, to derive the case distribution in Table A, we: (1) Assumed that about 4 percent of the "completed interview" cases in NHIS required more than nine visits to complete, as was found in the 1997 NHIS (Sangster and Chiu, 2001); (2) Adapted percent distributions by follow-up levels which were assumed for cases having each survey outcome in budget estimates for the 1999 NHIS; (3) Assigned numbers of attempts which seemed appropriate to each level of follow-up, and (4) Assumed the distribution of cases by survey outcome that was found in the 1999 NHIS. Using the proportions shown in Column (3) of Table A for cases remaining in each of nine rounds of visits, average increases in distances per trip relative to the distance for the first trips were calculated from the formula:

$$\sqrt{1/(percent\ of\ cases\ that\ are\ incomplete)} - 1.$$

It can be seen, for example, in Column 4 of Table A that

the second trips between cases within segments were estimated to be about 31 percent longer than the first trip. A case not completed after nine attempts was assumed to be the only one left in the assignment for travel to the tenth or later attempt. Hence, for simplicity, the round trip distance for each tenth or later trip was estimated to be twice the square root of the area for that assignment's PSU component, ignoring the possibility that FRs may work on other assignments on the way to or from those tenth or later attempts. Summing the increasing distances traveled over all attempts per case and averaging over cases, an estimated total of 10 and 97 travel miles at the case level were required to complete each case in area and permit segments, respectively.

For distances traveled between segments, Judkins and Waksberg (1990) assumed five to six trips per segment in their cost model. An average of five trips between area segments was adopted in the present model because BOC staff felt that the originally proposed 5.5 trips were too high for the current NHIS. Thus, 89 and 135 miles (five times the average miles per trip estimated above) were allocated for each area segment in SR and NSR PSUs, respectively, with an overall average of 105 miles.

The remaining miles for the interviewing task were allocated to the assignment level. Within the assignment level, miles traveled between NSR PSUs were first estimated and then the remaining miles were distributed equally among all assignments. For NSR travel, BOC staff estimated that about 50-70 percent of assignments in NSR PSUs are located in PSUs without resident FRs and about a third of such assignments require per diem. Hence, for modeling purposes, 60 percent of NSR assignments were assumed to require non-resident FRs. For every assignment requiring per diem, it was assumed that each assigned FR completed his/her work in a single trip to that assignment's PSU and traveled an average of 183 miles for the round trip to that PSU. For each NSR assignment without per diem, it was assumed that FRs made four trips to the assignment's PSU with round trips to the PSUs averaging an estimated 45 miles each. The estimated distribution of miles to the various survey levels is presented in Table B.

Table B also presents the distribution of interview task time resulting after FR travel times were estimated by dividing the allocated miles by the following assumed driving speeds:

Survey unit	SR PSUs	NSR PSUs
Cases in area segments	10	20
Cases in permit segments	25	35
Area segments	25	35
Assignments	45	50

Almost a third of the time was estimated for travel at the case level. The remaining travel was estimated to use about 20 percent of the task time. That leaves about half of the task time for non-travel activities, all of which was allocated to the case level.

Applying the average cost per hour and mile to the hours and miles allocated to the different survey levels and summing over all known costs gives a total cost for each survey level in the cost model. These totals were averaged over the units in the final sample at each survey level. As opposed to the fielded cases, the final cases were those 39,340 which yielded a "completed interview" and which were, thus, retained in the 1999 NHIS data files.

Table C presents the costs per unit relative to that of a case in the final sample for each survey level. It can be seen that the cost per segment and cost per assignment are about 2.2 and 7.3 times that for a final case. If the cost per final case were $1.00, the simple model for variable costs would become:

$$
\begin{aligned}
\text{Variable costs} = \ &\$1.00 \ (\text{number of final cases}) \\
&+ \$2.17 \ (\text{number of segments}) \\
&+ \$7.33 \ (\text{number of assignments}) \ .
\end{aligned}
$$

6. Screening sample costs

As mentioned earlier, the NHIS oversamples subgroups of the population by including a screening sample in which interviews are completed only in those HUs occupied by members of the targeted universe. The remaining HUs that would be eligible for the NHIS interview if they were in the regular (interview) sample are "screened-out." Because screening is costly, there is interest in knowing the cost of a screened-out case relative to that of a screened-in case (completed interview). The costs of travel and labor for getting to the HU, contacting the occupants, and reporting the case results are the same between the two types of outcomes. The only difference in survey operations between a completed interview and a screened-out case is the number of minutes spent in the actual interview. Because NHIS is a CAPI survey, computer "interview" minutes were recorded by the FR laptops for each case. The average "interview" minutes per case are presented in Table D together with estimated cost ratios for each survey outcome in area segments (the source of the frame for the screening sample). For each outcome, the ratio numerator is the estimated cost per case with that outcome and the denominator is the cost per case for a completed interview outcome. Column 1 in Table D reflects only case level costs. Because numbers of assignments and area segments (averaging 21 and 11.3 cases each, respectively, in the 1999 NHIS) must be increased to permit a screening sample, segment and assignment costs per case were included in the calculations for the Column 2 cost ratios. Based on the current model, the case-level and total costs for a screened-out outcome are about 69 and 84 percent of those costs, respectively, for a complete interview outcome

in the screening sample. Table D also shows that the average total costs per fielded and final case in the screening sample are 0.8 and 3.5 times, respectively, those for a completed interview outcome. The total cost per final case in the screening sample was also estimated to be about 2.6 times the total cost of a final case in the interview sample (not shown in Table D).

In the 1999 NHIS sample only 24 percent of the screening sample was screened in. To estimate the cost per case that would result from a different proportion for the screened in cases while everything else in the sample and survey operations stays the same, the total costs relative to that for a screened-in case may be formulated as

Relative total cost per fielded case = 0.16 R + 0.76
Relative total cost per final case = 0.16 + 0.76 / R

where R is the proportion of screened-in cases.

7. Summary

A simple model to approximate the annual variable costs for the 1999 NHIS was developed using primarily administrative data. The model is subject to the accuracy of assumptions made about details of survey operations which are typically not needed in administrative records. However, the model appears adequate for study of costs due to screening samples included in the NHIS. Among other observations from the model, it appears that the housing unit-level costs for a screened-out outcome are about 65-70 percent of those for a completed interview outcome and that ratio becomes roubhly 85 percent when total costs, including segment and assignment level costs, are considered.

References

Sangster RL and Chiu P (6/7/2001). "Interagency activities to address nonresponse in household surveys (part 2)." Paper presented at seminar for Washington Statistical Society. Washington, D.C.

Judkins D and Waksberg J. (1991). "Cost functions for NHIS and implications for survey design." *1990 Proceedings of the Section on Survey Research Methods*. American Statistical Association. pp 34-43.

Kalsbeek WD, Botman SL, Massey JT, and Liu P (1994). "Cost-efficiency and the number of allowable call attempts in the National Health Interview Survey." *Journal of Official Statistics*. pp 133-152.

Table A: Assumed cumulative distributions of cases by completion attempt and distances traveled per attempt relative to first attempt: 1999 NHIS

	Percent of cases		
Attempt Number (1)	Already completed (2)	Not completed (3)	Percent increase in distance (4)
1st	0%	100%	0%
2nd - 4th (ave. 3.0)	42%	58%	31%
5th - 9th (ave. 6.5)	72%	28%	91%
10th plus (ave. 10.5)	97%	3%	Not applicable

Table B: Derived distributions for FR miles and hours in CARMN's interviewing task: 1999 NHIS

Activity/Unit	Miles		Hours	
All	100%		100%	
Travel	100%		52%	
Cases	39%			32%
In area segments		24%		26%
In permit segments		15%		6%
Area Segments	23%			9%
Assignments	38%			11%
Non-travel			48%	

Table C: Estimated ratios of cost per survey unit to cost of a final case: 1999 NHIS

Unit	All	Subtotals
Final Case (n = 39,340)	**1.00**	
Fielded Case (n = 72485)		0.54
In area segment		0.50
In permit segment		1.16
Segment	**2.17**	
Area segment		2.25
Permit segment		1.78
Assignments	**7.33**	

Table D: Estimated averages per case in the screening sample for relative costs and interview minutes: 1999 NHIS.

	Expense level(s)		
Survey outcome	Case only	Case, segment, & assignment	Average interview minutes
Relative cost per case[1]			
Screening sample			
Final (n=7,363)	2.93	3.51	
Fielded (n=30,531)	0.71	0.85	19.3
Completed interview	1.00	1.00	57.5
Screened-out but occupants are NHIS eligible	0.69	0.84	8.0
Eligible occupants but no data collected	1.10	1.05	10.8
HU exists but no NHIS eligible occupants	0.44	0.70	4.5
Not an eligible HU	0.25	0.61	5.6

[1] Relative cost = cost per case with specific survey outcome ÷ cost per case with a completed interview.