Optimal Sample Design Procedures for National Hospital Ambulatory Medical Care Survey

Feng Lan, U.S. Postal Service and Iris Shimizu, National Center for Health Statistics
Iris Shimizu, NCHS, 3311 Toledo, Rm 3106, Hyattsville, MD 20782

KEY WORDS: Sample design, complex sample, optimization, Lagrange method

1. Introduction[1]

The National Hospital Ambulatory Medical Care Survey (NHAMCS) is actually two surveys which share a common hospital sample. Both produce statistics about ambulatory visits made to hospitals for health care, but one is for visits made to outpatient departments (OPDs) while the other is for visits made to emergency departments (EDs). While the questionnaires for the two surveys do contain a number of items in common, the questionnaire for each survey also includes items which are unique to the visit population eligible for that survey. For example, the ED visit questionnaire asks about the use of triage procedures on the sampled visit.

The NHAMC uses a four stage probability sample of ambulatory visits made for health care to hospitals. The primary sampling units (PSUs) consist of counties, groups of counties or metropolitan statistical area (MSAs) in the 50 states and the District of Columbia. Within PSUs, non-institutional and non-Federal general and short stay (average length of stay is less than 30 days) hospitals are selected. Within hospitals, a stratified sample of service units is selected with strata defined by department (OPD and ED). Within departments having fewer than six service units, all of the service units are taken into the sample with certainty. Otherwise service units are subsampled. The service unit (or clinic) samples within OPDs are stratified by six clinic specialty groups: general medicine, surgery, pediatrics, OB/GYN, substance abuse, and other. The ultimate sampling units are ambulatory visits made to the EDs and the OPDs to receive health care; visits made solely to deposit samples or pay bills are excluded from the survey.

The PSU sample and most of the sample hospitals have been in NHAMCS since the survey's inception 1991-92. Concern has been expressed about the effect on response rate which may result from increasing hospital longevity in the survey. Hence, in the event that funds become available to replace the aging sample, research was begun on designs for a replacement sample. This paper outlines the procedures being used to develop and evaluate four-stage samples that are optimum for NHAMCS. At the time of this writing, however, verification of the resulting designs is incomplete so that design numbers included for illustration may not be final.

Section 2 presents objectives for the new sampling design for NHAMCS. Sections 3 and 4 discuss the methodology being used to develop sampling designs for NHAMCS while Section 5 covers procedures aimed at evaluating the resulting designs for potential fielding. Section 6 summarizes the research.

2. Objectives and assumptions

Because of tight budgets, the primary objective for the new sampling design for NHAMCS is to minimize rel-variances for a fixed cost with rel-variances being minimized for national statistics from both the ED and the OPD surveys comprising NHAMCS. It is required that the same hospital sample be used for both the ED and the OPD surveys in any design considered for the NHAMCS. The new sample is also to be cost neutral so that the only cost increase in the new sample will be due to inflation.

In this paper, it is assumed the new design will have the same four sampling stages that are present in the current design with geographic PSUs, hospitals, service units within hospitals, and visits being selected at the respective stages. The six specialty strata for clinics within OPDs are also to be retained in the new design to minimize the between-clinic sampling variances due to clinic specialization when clinics are not selected with certainty.

Because estimates by region (Northeast, South, Midwest, and West) and MSA status (is or is not a MSA) are desirable by-products of the survey, it is assumed that the sampling designs will be stratified by region and MSA status. No attempt, however, was made to optimize the samples with regard to regional estimates or MSA status estimates.

For research purposes, it is assumed that the NHAMCS field operations and, hence, the variable costs (survey expenditures which rise or fall with the numbers of sample cases included at each sampling stage in the survey) would be similar to those in 1999, the most recent year for which data were available at the time research was begun on the new sample design. Shimizu and Lan (2002) developed the cost model adopted in the optimization research. Those costs estimates, divided by the estimated cost per sampled visit, are included in Table 1. (Confidentiality restrictions forbid presentation of the actual estimates.) Relative costs are adequate for optimization purposes.

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

Estimates assumed to be of interest for optimization purposes were selected from among items which are common between the ED and OPD survey questionnaires. An attempt was made to include a range of estimated magnitudes in addition to including a variety of estimates for major patient demographic characteristics known to be associated with clustering in the surveyed visit populations. The variables selected for the research (11 for each of the two surveys) are shown in Table 2 together with their 1999 estimates and relative standard errors (RSEs).

For the optimization process, the sampling at all stages is assumed to be without replacement.

3. Methodology for optimizing sample sizes

Formulae for the required samples sizes at each stage were derived using the method of Lagrange multipliers as presented by Hansen, Hurwitz, and Madow (1996). The procedure obtains the optimum sample sizes that maximize precision for a fixed survey cost or minimize the cost for a fixed level of precision. For the NHAMCS, the cost was fixed while precision was maximized. The function that must be solved for optimizing values may then be formulated as:

$$F = \text{Rel-var}(X) + \lambda\ [\, f(c) - C\,] \qquad (1)$$

where:

- $$\text{Rel-var}(X) = \frac{1}{X^2}\Bigg[\sum_k A_k^2\left(\frac{1}{a_k}-\frac{1}{A_k}\right)S_{1k}^2 + \sum_k \frac{A_k^2\overline{M}_k^2}{a_k}\left(\frac{1}{\overline{m}_k}-\frac{1}{\overline{M}_k}\right)S_{2k}^2 + \sum_k \frac{A_k^2}{a_k}\frac{\overline{M}_k^2}{\overline{m}_k}\sum_i\sum_j \overline{N}_{kij}^2\left\{\left(\frac{1}{\overline{n}_{kij}}-\frac{1}{\overline{N}_{kij}}\right)S_{3kij}^2 + \frac{\overline{V}_{kij}^2}{\overline{n}_{kij}}\left(\frac{1}{\overline{v}_{kij}}-\frac{1}{\overline{V}_{kij}}\right)S_{4kij}^2\right\}\Bigg] \qquad (2)$$

where:

- o k denotes PSU strata,
- o i denotes hospital department (ED or OPD),
- o j denotes service unit strata within department; j = 1 in EDs and j = 1, …, 6 in OPDs.
- o A_k, $\overline{M}_k$, $\overline{N}_{kij}$, and $\overline{V}_{kij}$ represent estimated counts of PSU's, hospitals, service units, and visits respectively, and a_k, $\overline{m}_k$, $\overline{n}_{kij}$, and $\overline{v}_{kij}$ represent the corresponding sample sizes by strata.
- o S_{1k}^2, S_{2k}^2, S_{3kij}^2, and S_{4kij}^2 are the estimated population variances of X between units aggregated at each stage of sampling. These variances were approximated by first estimating the population variance components for visit means in a four-stage sampling design. Then the first, second and third stage components for means were multiplied by the estimated average numbers of visits per PSU, per hospital, and per service unit, respectively, to yield estimates for the corresponding components for visit aggregates. The details of methodology used to estimate the variance components may be obtained from the authors. The variance estimates were based on 1999 NHAMCs data.
- λ is the Lagrange multiplier
- f(c) is the cost function consisting of the costs associated with each stage of the survey

$$f(c) = \sum_k a_k c_1 + \sum_k a_k \overline{m}_k c_2 + \sum_k a_k \overline{m}_k \sum_i \sum_j \overline{n}_{kij}(c_3 + \overline{v}_{kij} c_4) \qquad (3)$$

where the c_ℓ (ℓ = 1, 2, 3, 4) is the variable cost associated with each respondent sample unit at the ℓ-th sampling stage (shown in Table 1, column 3).

- C is the fixed total for variable costs.

The function F was differentiated with respect to each unknown sample size and λ, resulting in equations which were set equal to zero and solved simultaneously. The general derived formulae for sample size allocation are:

$$\overline{v}_{kij} = \overline{V}_{kij}^2\sqrt{\left(c_3/c_4\right) S_{4kij}^2 \Big/ \left(S_{3kij}^2 - \overline{V}_{kij} S_{4kij}^2\right)} \qquad (4)$$

$$\overline{n}_{kij} = \overline{N}_{kij}\sqrt{\frac{c_2}{c_3}\frac{\left(S_{3kij}^2 - \overline{V}_{kij} S_{4kij}^2\right)}{S_{2k}^2 - \sum_{i'}\sum_{j'} \overline{N}_{ki'j'} S_{3ki'j''}^2}} \qquad (5)$$

$$\overline{m}_k = \overline{M}_k\sqrt{\frac{c_1}{c_2}\frac{(S_{2k}^2 - \sum_i\sum_j \overline{N}_{kij} S_{3kij}^2)}{S_{1k}^2 - \overline{M}_k S_{2k}^2}} \qquad (6)$$

$$a_k = \left(A_k / b\right)\sqrt{\left(S_{1k}^2 - \overline{M}_k S_{2k}^2\right)\Big/ c_1} \qquad (7)$$

where, letting the term "k in SR" denote PSU strata in which PSUs are self-representing (selected with certainty) and letting "k in NSR" denote strata in which PSUs are non self-representing (not selected with certainty),

$$b = \frac{1}{C - \sum_{k \text{ in SR}} c_1}\left(d + e\right) , \quad (8)$$

$$d = \sum_{k \text{ in NSR}} \frac{A_k}{\sqrt{c_1}} \sqrt{S_{1k}^2 - \overline{M}_k S_{2k}^2}\left(c_1 + m_k g_k\right) ,$$

$$e = \sum_{k \text{ in SR}} \frac{M_k g_k}{\sqrt{c_2}} \sqrt{S_{2k}^2 - \sum_i \sum_j \overline{N}_{kij} S_{3kij}^2} \text{ , and}$$

$$g_k = c_2 + \sum_i \sum_j \overline{n}_{kij}(c_3 + \overline{v}_{kij} c_4) .$$

The general solutions in (4) - (7) apply except in two situations:

a) EDs in NSR PSUs tend to have only one service unit so that, for NSR PSUs: $\overline{n}_{k,ED} = 1$ and

$$\overline{v}_{k,ED} = \overline{V}_{k,ED} \sqrt{(c_2 / c_3) S_{4k,ED}^2 / \left(S_{2k} - \overline{V}_{k,ED} S_{4k,ED}^2\right)} . \quad (9)$$

b) The SR PSU strata are comprised of single PSUs so that for SR PSUs: $a_k = 1$ and

$$\overline{m}_k = \left(M_k / b\right) \sqrt{\left(S_{2k}^2 - \sum_i \sum_j \overline{N}_{kij} S_{3kij}^2\right) \Big/ c_2} . \quad (10)$$

4. Adjustments to Lagrange methods

The requirement to produce separate estimates for non-overlapping visit populations of EDs and OPDs demands some modifications to the Lagrange method. In the following, these altered procedures are described for OPD statistics to facilitate discussion with the understanding that these procedures also apply to ED statistics, unless otherwise stated.

When optimizing the sample for an OPD statistic, the hospital level cost c_2 in (3) must reflect the cost of adding OPDs to the sample as opposed to the cost of adding hospitals in general, because some hospitals do not have OPDs and are, thus, ineligible for OPD surveys. Let c_2(OPD) and c_2(HOS) denote the hospital level costs for respondent OPDs and hospitals, respectively. These two costs differ (see Table 1) because hospital frame data are not sufficient to permit identifying and removing the hospitals which have no OPDs from sampling frames for OPD surveys.

Also, when producing statistics for OPDs, data from EDs are not required and, hence, the optimum sizes $\overline{n}_{k,ED}$ and $\overline{v}_{k,ED}$ in (4) and (5) are zero when samples are optimized for OPD visit statistics. When the sample is void of ED units, the Lagrange multiplier assumes all of the variable costs are spent on producing OPD statistics, leaving no funds with which to collect ED visit data. To force compliance with the objective of producing statistics for both EDs and OPDs from the NHAMCS sample, modifications were made to the costs used in the Lagrange multiplier function in equation (1) and in resulting samples.

a) Service unit and visit level costs for ED data within hospitals were added to hospital level costs to produce MOD-c_2(OPD) which was then substituted for c_2(OPD) in (3) when samples were optimized for OPD survey samples. Use of this artificially increased c_2 cost in the Lagrange process reserves some funds to collect ED data by forcing a reduction in hospital sample sizes and, hence, in the actual hospital level costs for the new sample. The MOD-c_2(OPD) was derived by first adding to total hospital level costs the ED portion of total costs for the service unit and visit level samples and then dividing the resulting sum by the number of OPDs which participated in the NHAMCS. Using data from Table 1, MOD-c_2(OPD) = \$653.70 [=(\$119.4K + \$17.0K + \$21.2K)/241]. Similarly, when optimizing samples for ED statistics, MOD-c_2(ED) = \$488.80 [=\$119.4K + \$34.5K + \$29.9K)/376}.

b) When optimizing samples for OPD statistics, the fixed total variable cost (C) targeted in the Lagrange multiplier function (1) was decreased by subtracting the ED portion of costs for the third and fourth stage samples. Using data from Table 1, the fixed total costs targeted when optimizing samples for OPD statistics became MOD-C(OPD) = \$305.8K (=\$344.0K- \$17.0K - \$21.2K) where K = 1,000. Likewise, MOD-C(ED) = \$279.6K (= \$344.0 - \$34.2K - \$29.9K). Targeting the reduced MOD-C(OPD) when optimizing samples for OPD statistics reserves some additional funds for collecting data about ED visits, and so forth.

c) The samples optimized for estimating the variable X for OPD visits were augmented with service unit and visit samples from EDs where the added service units per hospital and visits per service unit were those yielded by the Lagrange process when samples were optimized for estimating the corresponding variable X for ED visits. That is, if $a_k(\text{OPD})$ and $\overline{m}_k(\text{OPD})$ denote the optimum sample sizes for PSUs and hospitals when samples are optimized for OPD statistics and $\overline{n}_{k,ED}(\text{ED})$ and $\overline{v}_{k,ED}(\text{ED})$ denote the optimum sample sizes for service units and visits when samples are optimized for ED statistics, then from the k-th PSU stratum, the number of ED service units and visits added to samples optimized for OPD statistics may be formulated as:

$$n_{k,ED}(\text{OPD}) = a_k(\text{OPD}) * (0.77/0.49)\overline{m}_{k,ED}(\text{OPD}) * \overline{n}_{k,ED}(\text{ED}) \quad (11)$$

$$\overline{v}_{k,ED}(\text{OPD}) = n_{k,ED}(\text{OPD}) * \overline{v}_{k,ED}(\text{ED}) \quad (12)$$

where 0.77 and 0.49 are the proportions of the total 1999 NHAMCS hospital sample which had participating EDs and OPDs, respectively, and the product $(0.77/0.49)\overline{m}_k(\text{OPD})$ is an estimate for the number of respondent EDs expected from the k-th stratum in samples that are optimized for OPD statistics. Similarly, the number of OPD service units and visits added to samples optimized for ED statistics may be formulated as

$$n_{k,OPD}(\text{ED}) = a_k(\text{ED}) * (0.49/0.77)\overline{m}_k(\text{ED}) * \overline{n}_{k,OPD}(\text{OPD}) \quad (13)$$

$$v_{k,OPD,j}(\text{ED}) = n_{k,OPD,j}(\text{ED}) * \overline{v}_{k,OPD,j}(\text{OPD}) \quad (14)$$

Other modifications were applied when the Lagrange process yielded un-reasonable sample sizes. If sample sizes exceeded the population sizes at any stage, the excessive sample size was replaced with the corresponding population size. To assure the ability to approximate variances, samples sizes that were less than two in a stratum at any stage were set equal to the minimum of two or the universe size in that stratum and stage. Also, fixes were made during computations to prevent the computer from stopping when negative or zero terms appeared under radical signs. For example, if the difference $S_{3kij}^2 - V_{kij}S_{4kij}^2$ in (4) and (5) was negative, $\overline{v}_{kij}$ and $\overline{n}_{kij}$ could not be computed. In these cases $\overline{v}_{kij}$ was set equal to minimum (30, $\overline{V}_{kij}$) and $\overline{n}_{kij}$ was set equal to minimum (2, $\overline{N}_{kij}$). The 30 visit minimum for $\overline{v}_{kij}$ is arbitrary but it yields about one sampled visit per day in the hospital's four-week data collection period and, thus minimizes the chance of hospital staff forgetting to sample visits and abstract their data for NHAMCS.

The optimum sample sizes yielded by the Lagrange technique are the numbers of responding sample units needed to minimize the rel-variances for the fixed cost. To assure that those optimum numbers of respondents are obtained the total sample fielded must be inflated sufficiently to allow for ineligible and non-respondent units that are likely to be included in the sample. At the hospital level, that inflation consisted of dividing the optimum hospital numbers by the proportions of the hospital sample that had OPDs and EDs which participated in the 1999 NHAMCS. As mentioned above, these proportions were 0.49 and 0.77 for OPDs and EDs, respectively.

Of the remaining (non-hospital) sampling stages, only the PSU sample sizes were inflated. It is unlikely that geographic out-of-scope PSUs (PSUs with no hospitals) will be sampled because the PSU sampling frame will be restricted to those containing hospitals according to the hospital universe list. However, a PSU becomes a non-respondent if none of its eligible hospitals participate in the survey. The potential for PSU non-response is greatest in those PSU strata where the hospital universe size $\overline{M}_k$ per PSU is less than the total hospital sample size needed per sample PSUs to account for hospital non-response and ineligibility. In such PSU strata, the total PSU sample number was derived by dividing the total hospital sample needed across PSUs from that stratum by the PSU average $\overline{M}_k$. At the service unit and visit sampling stages, few, if any, out-of-scope units are selected. While some non-response does occur at the service unit and visit levels, inflation for non-response at these two stages was not done in this research under the assumption that such inflation increases the risk of hospital non-response due to response burden.

5. Design evaluations

Because samples were optimized for eleven research variables for each of two hospital department types (ED and OPD), 22 potential designs were developed for the next NHAMCS sample. Evaluations of the costs and the expected precision levels for those samples are desired to aide in choosing which of those designs may be considered for fielding.

One goal of the optimum design is to make the new sample cost neutral, which means that if the new sample were implemented in 1999, it would cost about the same as the 4-stage sample which was actually fielded in 1999. Samples which cost more than the 1999 NHAMCS sample will be dropped from further consideration. Hence, the total variable costs for each of the derived samples were estimated and compared with the targeted cost. The costs of the different samples will vary from the targeted total cost because of modifications required in the Lagrange procedures to assure that each sample could produce estimates for both hospital department types and because of adjustments made to correct unreasonable sample sizes initially yielded by the procedures.

The variable costs for the sample optimized to estimate the variable X for hospital department type i (i = ED, OPD) can be formulated as:

$$C(i,X) = f(c) + \sum_k \left[c_3 n_{k,i}(\tilde{i}) + c_4 v_{k,i}(\tilde{i}) \right] \qquad (15)$$

where

f(c) is given in (3) when the un-augmented optimum sizes (before inflation for non-response and ineligibles) for that sample are used together with the costs per unit given in Table 1, and $c_2(\mathrm{OPD})$ and $c_2(\mathrm{ED})$ are used in place of the hospital level cost c_2 for samples that are optimized for OPD and ED visit statistics, respectively.

$n_{k,i}(\tilde{i})$ and $v_{k,i}(\tilde{i})$ ($\tilde{i}$ = ED, OPD and $\tilde{i} \neq i$) are the total sample sizes described in Section 4 for service units and visits that were added to the original samples which were optimized to estimate X for hospital department type i. For samples optimized to produce OPD statistics, these numbers are $n_{k,ED}(\mathrm{OPD})$ and $v_{k,ED}(\mathrm{OPD})$ from (11) and (12), respectively. For samples optimized to produce ED statistics, these numbers from (13) and (14), respectively, are the sums (over clinic strata): $n_{k,OPD}(\mathrm{ED}) = \sum_j n_{k,OPD,j}(\mathrm{ED})$ and $v_{k,OPD}(\mathrm{ED}) = \sum_j v_{k,OPD,j}(\mathrm{ED})$.

Table 3 presents costs ratios for preliminary sample designs where the ratios are derived by dividing the estimated costs from (15) by the 1999 variable cost total. If the preliminary sample sizes were verified to be the final sizes, it can be seen that certain of the samples would be dropped from further consideration on the basis of costs. The low (less than 0.95) cost ratios for all of the samples that were optimized for ED statistics suggests that the cost modifications made in the Lagrange process over-compensated for the cost of sampling OPD service units and visits within sample hospitals.

Another goal for the new sample is to maximize the precision for the fixed cost. To evaluate precision levels that could be expected from each of the derived samples, that sample design was used to approximate the relative standard error (RSE) for the estimate of each research variable Y. The expected RSE for $\hat{Y}$ may be formulated as:

$$\mathrm{RSE}_X(\hat{Y}) = \sqrt{\mathrm{Rel}-\mathrm{var}_X(\hat{Y})} \qquad (16)$$

where the formula for $\mathrm{Rel}-\mathrm{var}_X(\hat{Y})$ is given in equation (2) when the variate X is replaced by the estimate $\hat{Y}$, the population variance components are those for the variate Y, and the sample sizes are the optimum (before inflation for non-response and ineligibles) sizes derived for estimating the variate X.

Computer code which implements (16) to compute the expected RSEs has been written but the values generated, thus far, appear un-realistically small and, hence, are not presented with this writing. After this issue is resolved, RSEs will be computed and analyzed to determine which of the potential samples yield the best expected RSEs for all of the research variables.

After considering costs and RSEs, there may still be multiple designs that are potential for the new NHAMCS sample. If that happens, the choice of design may be based on hospital sample sizes under the assumption that hospital numbers probably have the greatest influence on the NHAMCS survey costs when all else is equal. That is, among the designs not eliminated on the basis of costs and RSEs, the design chosen may be that one for which the hospital sample size is closest to the hospital sample fielded annually in the current NHAMCS design. If the number of hospitals in the new sample is close to the number currently fielded annually, it is likely that the new sample will be close to cost neutrality despite potential for flaws in the available cost model.

6. Summary

An altered Lagrange technique was used to develop potential sample designs for the National Hospital Ambulatory Care Survey (NHAMCS) because NHAMCS requires the use of a single hospital sample to produce separate estimates about visits to outpatient departments (OPDs) and emergency departments (EDs). When being used to optimize samples that produce estimates for one of the two visit populations in the hospitals, the Lagrange method assumes that all of the survey's expenditures are made to collect data for that population alone, leaving no funds to collect data from the other population. By making certain modifications to the cost model used in the Lagrange technique and then augmenting the Lagrange results, preliminary sample designs for NHAMCS were developed which permit separate estimates about OPD and ED visits. Subject of verification of calculations, the designs will be evaluated in terms of their estimated fielding costs and their expected precision levels to facilitate decisions about which to consider for fielding.

References

Hansen MH, Hurwitz WN, and Madow WG. (1996). *Sample Survey Methods and Theory (Vol. 1&2)*. New York: John Wiley & Sons, Inc.

Shimizu I and Lan F (2002). "Approximation of Variable Costs for the National Hospital Ambulatory Medical Care Survey." 2002 Proceedings of the American Statistical Association, Survey Methods Research Section [CD-ROM], Alexandria, VA: American Statistical Association. 3197-3202.

Table 1: Estimated variable costs relative to that of a sample visit by sampling stage: 1999 NHAMCS[1]

Sampling stage	Sample numbers (1)	Relative costs: Total (in 1,000s) (2)	Relative costs: Per unit (3)
		Costs relative to that of a sample visit	
All		$345.2K	
PSU (resp.)	108	$123.3K	$1,141.72
Hospital (resp.)	403	$119.4K	$296.24
ED (resp.)	376		$488.80
OPD (resp.)	241		$653.70
Service units	1,405		
From ED	463	$17.0K	
From OPD	942	$34.5K	
Responding	1,261	$51.5K	$40.85
Visits	51K	$51.0K	$1.00
From ED	21K	$21.2K	
From OPD	30K	$29.9K	

[1] Source: Shimizu and Lan (2002)

Table 2: Research variables and their relative standard errors (RSEs) by hospital department type

Variable	ICD-9-CM codes	ED: Estimates in 1000s (RSEs[1])	OPD: Estimates in 1000s (RSEs[1])	ED: Estimated Percent	OPD: Estimated Percent
All visits		102,765 (4.4%)	84,623 (8.3%)	100.0%	100.0%
Less than 18 years old		27,350 (6.0%)	22,778 (10.6%)	26.6%	26.9%
Greater than 64 years old		15,659 (4.1%)	12,461 (8.7%)	15.2%	14.7%
Women		54,219 (4.4%)	51,050 (8.5%)	52.8%	60.3%
White		78,581 (5.1%)	63,542 (9.5%)	76.5%	75.1%
Black		21,119 (5.6%)	17,809 (9.4%)	20.6%	21.0%
Injury & Poisoning	800-999	29,586 (4.5%)	5,304 (12.7%)	29.3%	6.3%
Mental Disorder	290-319	2,903 (7.4%)	5,476 (16.8%)	2.9%	6.5%
Respiratory Disease	460-519	12,991 (5.9%)	10,664 (10.9%)	12.9%	12.6%
Circulatory Disease	390-459	4,397 (6.2%)	5,496 (12.1%)	4.4%	6.5%
Digestive Disease	520-579	5,947 (6.2%)	3,213 (13.2%)	5.9%	3.8%

[1] $RSE(X) = \sqrt{Variance(X) / X^2}$

Table 3: Total PSU and hospital sample sizes and cost ratio for preliminary optimized samples: NHAMCS

		Hospital department type					
		ED		OPD		ED	OPD
		PSU	Hospital	PSU	Hospital		
		Total sample sizes (inflated for non-response)				Cost ratios	
All visits		65	178	69	353	0.48	0.71
Less than 18 years old		115	323	71	463	0.83	0.96
Greater than 64 years old		106	334	78	426	0.87	1.07
Women		87	340	64	344	0.83	0.92
White		89	261	76	321	0.81	0.96
Black		126	340	92	250	0.93	0.93
Injury & Poisoning	800-999	81	255	91	604	0.68	1.15
Mental Disorder	290-319	92	287	85	425	0.75	1.02
Respiratory Disease	460-519	91	283	73	332	0.74	0.91
Circulatory Disease	390-459	84	293	78	343	0.76	0.93
Digestive Disease	520-579	106	318	86	303	0.87	0.96