Contract No.: 282-98-0021
MPR Reference No.: 8860-220

2003 Health Care Survey of DoD Beneficiaries:

Child Sample Report

May 2003

Final

Nancy Clusen
Esther Friedman

Submitted to:

TRICARE Management Activity
5111 Leesburg Pike, Suite 810
Falls Church, VA 22041
(703) 681-4263

Task Order Officer:
LTC Michael Hartzell, DVM MPH

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
(202) 484-9220

Project Director:
Eric Schone, Ph.D.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Contents

Chapter **Page**

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Appendices

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Executive Summary

The annual Health Care Survey of DoD Beneficiaries (HCSDB) is comprised of two surveys. One is an adult survey of active-duty military personnel, retirees and their family members eligible for care under the military health system (MHS), and the other is a survey of beneficiaries younger than 18. Both surveys measure health care status as well as access to care, use of, and satisfaction with care in the MHS. The child survey was conducted in the 1996, 1997, 1999, 2000, and 2002 HCSDBs. The child survey was not fielded in the 1998 HCSDB.

Data collection for the 2003 Child HCSDB will occur during the third quarter of 2003. This report documents the procedures used to design and select the sample of child beneficiaries for the 2003 Child HCSDB.

The 2003 Child HCSDB has a stratified sample design and involves the selection of 35,000 child beneficiaries. The sample selection process included six steps: (1) constructing the sampling frame for use in sample selection, (2) determining sampling strata based on analytic purposes, (3) assigning the sample sizes to strata to satisfy the precision goals established for the study using an optimal allocation algorithm, (4) selecting the samples for the survey using a systematic sample selection algorithm, (5) creating sampling weights that reflect the probability of selection, and (6) checking results to ensure that sampling and weighting were implemented as specified.

The major features of the sample design for the 2003 Child HCSDB are:

- The sampling frame consists of the approximately 1.8 million beneficiaries younger than 18 years old who are eligible for military health care benefits as of February 28, 2003. The sampling frame is limited to beneficiaries who live in the United States.
- Sampling strata are based on the cross of two types of TRICARE Prime enrollment statuses by three geographic areas and by three age groups. Types of TRICARE Prime enrollment status include enrolled in TRICARE Prime and not enrolled in TRICARE Prime. The geographic areas are New Region with regions 1, 2, and 5, Mature Region with regions 6, 9, 10, 11, 12, and 16, and Other Region with regions 3, 4, 7, and 8. The three age groups are less than 6 years, 6 through 12 years, and 13 through 17 years.
- The goals for the precision of survey estimates are half-lengths of 95 percent confidence intervals for a percentage of size 50. For all sampling strata, the goal is half-lengths of 5 percentage points. For estimates for the three regions described above, the goal is half-lengths of 2 percentage points. In addition, for the precision estimates for the MHS as a whole, the precision goal is half-lengths of 1 percentage point.
- The response rate for the 2003 survey is expected to be 31 percent.
- The precision requirements and expected response rates resulted in a sample of about 31,800 beneficiaries for the 2003 Child HCSDB. However, we have the resources to allow us to sample 35,000. We used an optimal allocation algorithm to allocate the 3,200 additional child beneficiaries.
- A systematic sample selection algorithm was used both to ensure proportional representation of the various substrata in the sample and to ensure that we did not select multiple children from the same family.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Chapter 1

Introduction

Congress mandated that the Department of Defense conduct an annual survey of its active-duty personnel, retirees, and eligible family members to measure accessibility, usage, and satisfaction in a range of health care services. The first of these surveys, referred to as the 1995 Health Care Survey of DoD Beneficiaries (HCSDB), was completed in the spring of 1995. The 1995 HCSDB surveyed the health care experiences of adult beneficiaries (that is, beneficiaries 18 or older). In 1996, the HCSDB expanded to include topics related to health care for children. The 1996 survey consisted of two separate questionnaires: Form A for adults and Form C for children's topics. The child survey was repeated in 1997, 1999, 2001, and 2002. This report summarizes and documents the procedures used to select the child samples for the 2003 Child HCSDB.

Chapter II of this report describes the construction of files required to select the samples of child beneficiaries for the 2003 Child HCSDB. Next, Chapter III discusses the stratification scheme used to draw the samples. This is followed in Chapter IV by a discussion of the derivation of sample sizes required to meet specified precision requirements. Finally, Chapter V describes the procedures used to draw the samples and summarizes the results of the sampling process.

The appendices include tables and SAS programs. Appendix A lists DEERS variables provided by Defense Manpower Data Center (DMDC). Appendix B includes population counts and sample counts (weighted and unweighted) tabulated for all sampling strata as part of the sample verification process. Appendix C includes all variables delivered to National Research Corporation (NRC) after the sample was selected, and Appendix D contains all SAS programs used for the 2003 sample design and sample selection. All technical arguments and related formula used in determining the sample sizes are presented in Appendix E.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Chapter 2

Construction of the Sampling Frame

The first step to select a sample that represents the target population is to create a sampling frame that lists all eligible members of the target population. The sampling frame for the 2003 Child HCSDB was based on a DEERS extract file and constructed as follows:

- A DEERS extract file was requested for sampling purposes.
- The sampling frame was constructed from the DEERS extract file for the reference date of February 28, 2003.
- The variables required for sampling were constructed.
- Population counts were calculated for potential stratification cells defined by the cross-classification of geographic area, beneficiary type, and enrollment status.

This chapter describes these operations.

A. REQUESTING THE DEERS EXTRACT FILE

The first step in building the frame was to prepare specifications that TMA could use to create the DEERS population extract. The extract file contained data for more than 8 million DoD health care beneficiaries (adults and children) as of February 28, 2003, and included information needed for sample selection as well as address/locator information for mailing the survey questionnaires. The variables in the extract file are listed in Appendix A.

Because we planned to use in-house SAS programs for sampling, we converted the extract file to a SAS data set. Beneficiaries in the DEERS extract file can be uniquely identified by a constructed variable, SSNSMPL, which contains confidential data.[1] We created a nonconfidential identification variable (MPRID) by randomly and uniquely assigning values from 1 to 1,978,410 to each child beneficiary in the extract file. The SAS-converted extract data file incorporates MPRID as the identification variable and excludes SSNSMPL. For historical purposes, we retained a crosswalk file that includes SSNSMPL and MPRID. The crosswalk file allows us to link frame records to the DEERS database if needed. Appendix D includes the SAS programs we used to create the crosswalk file and to transform the data set to a SAS data set.

To safeguard the security of the DEERS extract file, we used the procedures outlined in the following sources: The Guide to Understanding Configuration Management in Trusted Systems (Orange Book); DoD 5200.28; Appendix III to OMB Circular No. A-130-Security of Federal Automated Information Resources; the Computer Security Act of 1987; and the Privacy Act of 1974. We also maintained a secure data storage facility and a C2-compliant local area network, and we set up a chain of custody procedures. The original extract was returned to TMA four weeks after we received the data.

---

[1] SSNSMPL is formed by three DEERS variables: the nine-digit Social Security number (SSN), the one-digit family sequence number (FSN), and the two-digit DEERS dependent suffix (DDS).

B. DETERMINING ELIGIBLES FOR THE SAMPLING FRAME

The sampling frame for the 2003 Child HCSDB included beneficiaries listed in DEERS if they were:

- Younger than 18 on February 28, 2003 and resided only in the United States
- Eligible for military health care benefits as of February 28, 2003

Using the DEERS variables DHSRGN and DAGEQY, we determined that about 1.8 million children were eligible for health care benefits at the time of sample selection.[2]

The child sample was selected from this sampling frame after the constructed variables needed for sampling were added.

C. CONSTRUCTING THE VARIABLES REQUIRED FOR SAMPLING

Because the 2003 Child HCSDB uses a stratified sample design, variables for stratification had to be included in the sampling frame. Beneficiaries for the child survey were stratified by a combination of enrollment status, geographic area, and age group. (The stratification process is described in Chapter III.) For sampling purposes, some variables had to be created using the information from the DEERS extract files. These variables, along with the input DEERS variables used to construct them, appear below.

- **MPRID (nonconfidential identification number).** This variable corresponds uniquely to SSNSMPL so that units in the frame can be linked back to information from the extract file.
- **BGCSMPL (beneficiary group)**. This variable carries an extension of 1, 2,or 3 that denotes the following beneficiary groups: 1 = active duty, [3] 2 = active-duty family members, and 3 = retirees and family members. DEERS variables = PATCAT.
- **ENLSMPL (enrollment status of a beneficiary, stratification variable).** This variable carries an extension of 1 or 2 that denotes the following groups: 1 = enrollee, 2 = nonenrollee. DEERS variables = PCM.
- **REGSMPL (geographic region).** This variable carries an extension of 1 through 12 and 16 that denotes the following groups: 1 = Northeast, 2 = Mid-Atlantic, 3 = Southeast, 4 = Gulf South, 5 = Heartland, 6 = Southwest, 7 = Central, 8 = Central 9 = Southern California, 10 = Golden Gate, 11 = Northwest, 12 = Hawaii, and 16 = Alaska. DEERS variables = DHSRGN.
- **SUPREG (geographic area-stratification variable).** This variable was constructed from REGSMPL and carries an extension of 1, 2, or 3 that denotes the following: 1 = New Region (regions 1, 2, 5), 2 = Mature Region (regions 6, 9, 10, 11, 12, 16), 3 = Other Region (region 3, 4, 7, 8).
- **SVCSMPL (branch of service).** This variable carries an extension of 1, 2, 3, 4, 5, or 6 that denotes the following: 1 = Army, 2 = Navy, 3 = Air Force, 4 = Marine Corps, 5 = Coast Guard, 6 = Other.[4] DEERS variable: SVCCD.

---

[2] The SAS program used to build the sampling frame of eligible child beneficiaries is given in Appendix D.

[3] The active duty category also includes academy students, reservists, and people recently separated from military service who are covered for the mandated transition period after separation.

[4] If the Branch of service was unknown or otherwise could not be determined from the information in the DEERS file, we assigned a SVCSMPL code of 6.

- **SEXSMPL (sex of the beneficiary).** This variable carries an extension of 1 or 2 that denotes the following: 1 = male or missing or unknown, [5] 2 = female. DEERS variable: PNSEXCD.
- **AGESMPL (age group-stratification variable).** This variable carries an extension of 1, 2, or 3 that denotes the following: 1 = younger than 6, [6] 2 = 6-12 years old, 3 = 13-17 years old. DEERS variable: DAGEQY.
- **FAMCODE (family group indicator).** This variable was constructed from the sponsor Social Security number (SSNSMPL). Each child with the same sponsor Social Security number is grouped into the same family group indicator. The value of this variable is unique for an individual sponsor and was constructed so that it does not reveal confidential information about that sponsor.

---

[5] For sampling purposes, the “unknowns” were grouped with males.

[6] For sampling purposes, the “unknowns” were grouped with younger than 6.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Chapter 3

Construction of Sampling Strata

The 2003 Child HCSDB sample was independently selected within strata defined by geographic area, age group, and enrollment status. In this chapter, we describe the construction of sampling strata for the 2003 Child HCSDB.

A. STRATIFICATION VARIABLES

The 2003 Child HCSDB was stratified using a scheme similar to the 2002 Child HCSDB. The stratification variables are: (1) TRICARE Prime enrollment status, (2) super region based on geographic regions, and (3) age group.

1. Enrollment Status

We determined enrollment status by first dividing the target population into two groups: (1) enrolled in TRICARE Prime and (3) not enrolled in TRICARE Prime.

Enrollment status was determined using the DEERS variables for the PCM code (PCMCODE). By the definition of PCMCODE values, all beneficiaries with PCMCODE = MTF or CIV were assigned to the enrolled group. All others were assigned to the non-enrollment group.

2. Super Region

The definition of geographic area depends on a beneficiary's regional assignment (REGSMPL). The REGSMPL variable was constructed with a DEERS variable, the derived location military health service region code (DHSRGN). Related SAS codes are found in Framec.sas program of Appendix D. Super region variable (SUPREG) was then constructed by grouping values of REGSMPL:

- New Region (SUPREG = 1) consists of REGSMPL values of 1 = Northeast, 2 = Mid-Atlantic, and 5 = Heartland
- Mature Region (SUPREG = 2) consists of REGSMPL values of 6 = Southwest, 9 = Southern California, 10 = Golden Gate, 11 = Northwest, 12 = Hawaii, and 16 = Alaska
- Other Region (SUPREG = 3) consists of REGSMPL values of 3 = Southeast, 4 = Gulf South, 7 = Central, and 8 = Central

3. Age Group

Beneficiaries were assigned to one of three age groups: (1) younger than 6 years old, (2) 6 through 12 years old, and (3) 13 through 17 years old.

4. Family Group

To permit sampling procedures that eliminate multiple selections within the same household, all children with the same family code were grouped into one stratum. The assignment used a procedure to randomly assign children within the same household to one stratum. Each child within a family group was assigned a random number. The random numbers were then sorted within the families. All children were assigned to the stratum associated with the characteristics of the child with the smallest random number.

B. STRATIFICATION RESULTS

There are 18 sampling strata, which can be uniquely specified using the following variables: SUPREG (geographic site), ENLSMPL (enrollment status), and AGESMPL (age group).

The final step before selecting the sample was to generate stratum-level population counts to allocate the sample to meet predetermined precision rules for various domains. Chapter 4 discusses sample size allocation.

Chapter 4

Sample Sizes

The sample size appropriate to meet precision goals for each stratum and available resources determined the total sample size for the 2003 Child HCSDB. Because the strata are also important analytic domains, this strategy ensures that samples drawn from each stratum will be large enough to meet our precision requirements. In addition, stratification with approximately optimum allocation to strata can be effective in reducing the sampling errors. In this chapter, we present the procedures used for sample size allocation to meet precision requirements for the 2003 Child HCSDB. We discuss the requirements themselves, response rates, and how the sample sizes were finally determined.

A. PRECISION REQUIREMENTS

Precision requirements and expected response rates were the basis for determining stratum-level sample sizes. These requirements were defined to ensure adequate precision for constructing 95 percent confidence intervals. We are interested in estimating the proportion of beneficiaries with a certain attribute for a particular domain of interest. When the sample size is large enough, we can assume that estimated proportions will follow approximate normal distributions according to the Central Limit Theorem (Skinner et al. 1989). The resulting 100(1-α) confidence interval for a proportion of interest P is based on the standard formula:

$$p \pm z_{1-\alpha/2} \sqrt{V(p)} = p \pm HL \tag{4.1}$$

where p is an estimate of P, $z_{1-\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile point from the standard normal distribution with a mean of zero and standard deviation one, and *HL* is the half-length of the two-sided 95 percent confidence interval, or $HL = z_{.975}\sqrt{V(p)}$.

For the 2003 Child HCSDB, precision requirements specified that the *HL* of the 95 percent confidence interval in (4.1) for a given estimate should be less than or equal to a specified value. Because the maximum *HL* value occurs for $P = 0.5$, the precision requirements for the *HL*s were set for *P* values of 0.5; this guaranteed that *HL*s for all estimates would be less than specified values. The precision requirements for estimated proportions derived from the 2003 Child HCSDB are as follows:

- For sampling stratum-level estimates, the HLs are less than 0.05 (or 5 percentage points).
- For geographic area-level estimates, the HLs are less than 0.02 (or 2 percentage points).
- For estimates for the entire population, the HLs are 0.01 (or 1 percentage point) or less.

B. RESPONSE RATES

After calculating the desired number of eligible respondents needed to achieve the precision requirements specified above, we inflated the resulting sample sizes to account for survey nonresponse. The expected 2003 Child HCSDB response rate was assumed to be 31 percent across all strata.

C. SAMPLE SIZE COMPUTATION

In this section, we describe the key algorithms used in determining sample sizes and summarize how each precision requirement affected the total sample size. The technical presentation in Appendix E is the basis for the sample sizes we developed to meet the precision requirements for the 2003 Child HCSDB. Appendix D includes the in-house SAS programs we used in determining sample sizes.

1. Stratum-Level Sample Sizes

The first step in determining the final sample size was to calculate initial sample sizes, say n_h^0, for all strata using formula (E.4) in Appendix E to achieve the required half-length confidence interval of 0.05 or less for stratum-level proportion estimates. So, we substituted the value of 0.05, the half-length of the two-sided 95 percent confidence interval, for *B* in (E.4). In summary, the following specifications must be put into (E.4) to determine initial stratum sample sizes:

- *B* = 0.05 for all strata
- Stratum-level population counts for *N*

2. Sample Sizes to Meet Geographic Area-Level Precision Constraints

The next step was to adjust the initial sample sizes, n_h^0, to meet the precision goal for the geographic areas. First, we calculated formula (E.7) in Appendix E. We needed values for stratum-level population size (POPSIZE) N_h, and domain-specific population size (DSUM1) N_d. The summation in the formula occurs over all strata within the domain *d* geographic areas. Input values for (E.7) were:

- N_h : POPSIZE
- $N_d = \sum_{h \in d} N_h$: DSUM1
- n_h : n_h^0 obtained from (E.4)
- *B*: 0.02

We then compared the calculated variances using (E.7) with the predetermined values $V_{d,0} = B_d^2 / 3.8416$ for all geographic areas. If calculated values for all domains were less than or equal to the predetermined values, then the final stratum-level sample sizes, n_h^F, were the same as the initial sample sizes, n_h^0, for all strata within those geographic areas. Specifically, if all geographic area-level variances using formula (E.7) were less than or equal to $V_{d,0}$, then we skipped all steps in this section and considered the precision requirement for enrollment status group-level estimates, as discussed in the next section. However, the initial sample allocation did not meet the geographic area-level precision requirements for all areas. To satisfy these requirements, we increased the strata sample sizes using the optimization algorithm described in Appendix E.

The optimal geographic area-level sample sizes were calculated using formula (E.9) in Appendix E for all geographic areas. Here, N_d, N_h, and $V_{d,0}$ are the same as defined above, and the summation in the formula occurs over all strata within domain *d*. The output is denoted by n_d. With the optimal geographic area-level sample sizes, n_d, stratum-level sample sizes were also optimally allocated for all strata. Input values for (E.11) in Appendix E are the same as defined for (E.9) above. The resulting sample sizes at this step are denoted as n_h^{opt}.

We then compared optimal stratum-level sample sizes, n_h^{opt}, with initial sample sizes, n_h^0, from (E.4). If $n_h^{opt} \geq n_h^0$ (or $n_h^{opt} \leq n_h^0$) for all strata, then we took n_h^{opt} (or n_h^0) as tentative sample sizes for all strata and went to the next step to consider the precision requirement for the total sample estimates. However, in some strata, $n_h^{opt} < n_h^0$ or otherwise. Instead of unnecessarily inflating the total sample by taking the larger values from these two sample sizes, we repeated the optimal allocation algorithms to minimize the total sample size while meeting the stratum- and geographic area-level precision requirements.

3. Final Sample Sizes

Because resources allow us to field a larger sample than is required under the precision requirements, we used an optimal allocation algorithm for the surplus. Because of the surplus sample we can be assured that the subsequent quantity satisfies the precision rule *B* = 0.01 for the entire population.

After finalizing strata sample sizes for eligible respondents, we incorporated the expected response rates to obtain the final sample sizes. The final sample sizes were then calculated as:

$$n_h^F = \frac{n_h}{R}$$

for each stratum *h* where *R* = 0.31. The total number of beneficiaries to be selected for the 2003 Child HCSDB was determined to be 35,000.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Chapter
5

Selecting the Sample

The 2003 Child HCSDB sample was selected with a stratified, systematic sample design. The sampling was independently performed within the strata (see Chapter 3) based on the sample size allocation (see Chapter 4). Within each stratum, beneficiaries were sorted using hierarchical serpentine sorting[7] by the following analytic variables: family group (FAMCODE), region (REGSMPL), age (AGE_N), and sex (SEXSMPL). After beneficiaries were sorted, we sampled them using a systematic sample selection procedure. This procedure ensures that the various substrata, defined by the sorting variable, contain sample sizes proportional to their population sizes. Systematic sampling provides greater control over the distribution of the sample and, with appropriate choices for the sorting variables, can produce a stratified systematic sample that improves on the precision of a stratified simple random sample. However, systematic sampling can result in potential systematic biasing effects because only one random selection is used.

Beneficiaries were sampled at varying rates depending on the sampling stratum. The algorithm used to draw the sample automatically selected beneficiaries to yield the predetermined stratum sample size. In the discussion that follows, we describe systematic selection procedure, the development of the sampling weight, and how we checked sample size and weights to evaluate the selection procedure. Appendix D contains the SAS program for the 2003 Child HCSDB sample selection.

A. SYSTEMATIC SELECTION PROCEDURES

Details of the systematic selection procedure we used can be found in Cochran's *Sampling Techniques* (1977). We used the corresponding SAS procedure SURVEYSELECT, which has an option for systematic selection (SAS Institute Inc. 1999). Our sample selection process was based on a stratified sample design and predetermined stratum sample sizes. The population was stratified by the 18 strata resulting from the cross of the three stratifying variables (enrollment status, super region, and age group). Independent samples were drawn from each stratum separately.

Using the stratum sample size n_h for each stratum (h = 1, ..., 18), we used a systematic sample selection procedure with hierarchical serpentine sorting instead of simple random sampling to avoid the possibility of extreme concentrations of the selected sample in a few analytic domains. In selecting the sample size n_h from the N_h total beneficiaries in the h^{th} stratum, we sorted all beneficiaries by family group (FAMCODE), region (REGSMPL), age (AGE_N), and sex (SEXSMPL). We then selected a beneficiary with an equal probability of being selected within each stratum to define a random starting point for sample selection. Beginning with the randomly selected beneficiary and treating the stratum list as circular, we used a systematic procedure and

---

[7] Hierarchical serpentine sorting is a technique in which the sort order is reversed as each boundary is crossed for higher-level sort variables, to preserve the similarity of adjacent children in the sorted list. This procedure ensures that the substrata, defined by the sorting variables, contain sample sizes proportional to their number in the population.

selected every n^{th} beneficiary to yield the desired stratum n_h sample. We incorporated the SAS algorithms and wrote a custom program for the sample selection (Appendix D).

To limit the burden for any one family, we designed our sample to select only one child per household. While there are 1,828,020 child beneficiaries in the sample frame, there are only 1,008,827 families. Therefore, child beneficiaries were grouped into households according to their adult health benefits sponsor (FAMCODE). The selection interval used in systematic sampling is, in this case, larger than the number of children in any family. Therefore, within the same stratum no more than one child will be selected from each family. After sample selection, we checked whether multiple children from the same family were selected and found none.

B. SAMPLING WEIGHT

The last step in sample selection was to compute the base sampling weight (BWT) for each record and add this variable to the sampling file that was delivered to the data collection contractor, NRC. We constructed the sampling weight on the basis of the sample design and selected the sample with differential probabilities of selection across strata. Established precision requirements determined the sample sizes. The sampling weights, which reflect these unequal sampling rates across strata, were defined as the inverse of the beneficiary's selection probability, or $BWT_{hi} = N_h/n_h$, where BWT_{hi} is the sampling weight for the i^{th} sampled beneficiary from the h^{th} stratum, N_h is the total number of beneficiaries in the h^{th} stratum, and n_h is the number of sampled beneficiaries from stratum h. The sum of the sampling weights over selections from the h^{th} stratum equals the total population size of the h^{th} stratum or N_h.

C. CHECKS FOR THE SELECTED SAMPLE

After drawing the sample and creating the sampling weight, we evaluated the selection procedure by checking sample sizes and weighted sums for all strata. Appendix B contains the following frequency tables with unweighted and weighted counts:

- The number of sampled records for the stratification variable for geographic area (SUPREG) by stratification variable for enrollment group (ENLSMPL) by stratification variable for age group (AGESMPL)
- The number of sampled records for region (REGSMPL) by enrollment group (ENLSMPL) by age group (AGESMPL)
- The number of sampled records for region (REGSMPL) by sex (SEXSMPL)
- The weighted count of sampled records for the stratification variable for geographic area (SUPREG) by stratification variable for enrollment group (ENLSMPL) by stratification variable for age group (AGESMPL), where the weight is equal to BWT
- The weighted count of sampled records for region (REGSMPL) by enrollment group (ENLSMPL) by age group (AGESMPL), where the weight is equal to BWT
- The weighted count of sampled records for region (REGSMPL) by sex (SEXSMPL), where the weight is equal to BWT
- The population counts for the stratification variable for geographic area (SUPREG) by stratification variable for enrollment group (ENLSMPL) by stratification variable for age group (AGESMPL)
- The population counts for region (REGSMPL) by enrollment group (ENLSMPL) by age group (AGESMPL)
- The population counts for region (REGSMPL) by sex (SEXSMPL)

The sample counts after selection must be the same as the predetermined sample sizes for each stratum (SUPREG*ENLSMPL*AGESMPL). Also, the weighted sample counts must be the same as the population counts for each sampling stratum (SUPREG* ENLSMPL*AGESMPL). For analytic domains such as REGSMPL*ENLSMPL*AGESMPL and REGSMPL*SEXSMPL, sample count distributions were checked against the corresponding population distributions to ensure that no operational errors occurred and that the sample appeared to be reasonably balanced. Because the sampling rates used in the selection process varied, the weighted distributions for the analytic domains do not exactly match the population distributions.

After completing the sample checks, we attached the data elements that are used in the survey mailing and operations to each record in the sample extract file. We received addresses, as of February 28, 2003, for all beneficiaries in the DEERS extract file on April 16, 2003.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

REFERENCES

Cochran, W. *Sampling Techniques.* 3rd edition. New York: John Wiley, 1977.

SAS Institute Inc. *SAS/STAT® User's Guide, Version 8*. Cary, NC: SAS Institute Inc., 1999.

Skinner, C.J., D. Holt, and T.M.F. Smith. *Analysis of Complex Surveys*. New York: John Wiley, 1989.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

APPENDIX A

DEERS VARIABLES REQUESTED BY MPR

TABLE 1. LIST OF VARIABLES

	Variable Description
B.1.	Beneficiary's family sequence number
B.2.	Beneficiary's DEERS dependent suffix
B.3.	Age of beneficiary in years, representing the difference between the date-of-birth and the date of the reference date
B.4.	Beneficiary's date of birth
B.5.	Beneficiary's record type
B.6.	Beneficiary's marital status
B.7.	Beneficiary's race/ethnicity
B.8.	Beneficiary's sex
B.9.	Beneficiary's Medicare flag
B.10.	Beneficiary's primary care manager code
B.11.	Beneficiary's provider code
B.12.	Beneficiary's alternative care value
B.13.	Beneficiary's enrolled DMIS
B.14.	Beneficiary's first name
B.15.	Beneficiary's middle name
B.16.	Beneficiary's last name
B.17.	Beneficiary's generation
B.18.	Beneficiary's residential address – line 1
B.19.	Beneficiary's residential address – line 2
B.20.	Beneficiary's residential address – line 3
B.21.	Beneficiary's residential address – city
B.22.	Beneficiary's residential address – state
B.23.	Beneficiary's residential address – zip
B.24.	Beneficiary's residential address – zip extension
B.25.	Beneficiary's residential address flag – 0 if no res. Address available, 1 if there is a residential address
B.26.	Beneficiary's residence address – region
B.27.	Beneficiary's residence address – dmis code
B.28.	Beneficiary's residence phone number
B.29.	Beneficiary Type coded as one of these four groups: (1) active duty; (2) active-duty dependents; (3) Retirees and their dependents less than 65; or (4) Retirees and their dependents 65 and over
B.30.	Beneficiary's prime enrollment status as one of these three cases: (1) enrolled as a military Primary Care Organization; and (3) not enrolled
B.31.	Beneficiary's senior prime enrollment status coded as: (1) Senior Prime enrollee; or (2) Nonenrollee
B.32.	Beneficiary's Catchment area from the consolidation of: (i) the list of MTFs for Prime enrollees with military Primary Care Organization; (ii) the list of catchment areas for Prime enrollees with a civilian Primary Care Organization; and (iii) the list of service areas for non-enrollees
B.33.	Beneficiary's TRICARE region based on the constructed Catchment area assignment
S.1.	Sponsor's social security number
S.2.	Sponsor's duty status
S.3.	Sponsor's group code
S.4.	Sponsor's marital status
S.5.	Sponsor's pay grade
S.6.	Sponsor's race/ethnicity
S.7.	Sponsor's rank abbreviation
S.8.	Sponsor's service
S.9.	Sponsor's total dependents counted
S.10.	Sponsor's total active federal months of service
S.11.	Medical privileges of sponsor
S.12.	Sponsor's eligible dependents counted
S.13.	Number of dependents reported for sponsor

	Variable Description
S.14.	Sex of the sponsor
S.15.	Age of the sponsor
S.16.	Sponsor first name
S.17.	Sponsor middle name
S.18.	Sponsor last name
S.19.	Sponsor generation name
S.20.	Sponsor’s unit address – street 1
S.21.	Sponsor’s unit address – street 2
S.22.	Sponsor’s unit address – street 3
S.23.	Sponsor’s unit address – city
S.24.	Sponsor’s unit address – state/with asterisks to distinguish foreign vs. domestic addresses
S.25.	Sponsor’s unit address – zip
S.26.	Sponsor’s unit address – zip extension
S.27.	Sponsor’s unit address flag – 0 if no unit address is available, 1 if there is a unit address
S.28.	Sponsor’s unit region
S.29.	Sponsor’s unit address – dmis code
S.30.	Sponsor’s residential address – line 1
S.31.	Sponsor’s residential address – line 2
S.32.	Sponsor’s residential address – line 3
S.33.	Sponsor’s residential address – city
S.34.	Sponsor’s residential address – state
S.35.	Sponsor’s residential address – zip
S.36.	Sponsor’s residential address – zip extension
S.37.	Sponsor’s residence phone number
D.1.	Dependent SSN
D.2.	Dependent’s relationship to sponsor
D.3.	Dependent first name
D.4.	Dependent middle name
D.5.	Dependent last name
D.6.	Dependent generation name

APPENDIX B

TABLES FOR SAMPLING CHECK

TABLE B.1 Unweighted Values from Selected Sample Dataset
SUPREG by ENLSMPL by AGESMPL

Super Region	Enrolled			Not Enrolled			
	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
1	1,984	2,322	1,786	1,677	1,985	1,897	11,651
2	2,233	2,339	1,734	1,649	1,936	1,801	11,692
3	1,997	2,371	1,781	1,624	1,943	1,941	11,657
Total	6,214	7,032	5,301	4,950	5,864	5,639	35,000

TABLE B.2 Weighted values from Selected Sample Dataset
SUPREG by ENLSMPL by AGESMPL

Super Region	Enrolled			Not Enrolled			
	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
1	139,596	159,269	103,800	54,955	95,033	98,615	651,268
2	154,855	157,659	96,039	30,802	55,588	56,836	551,780
3	140,619	165,372	110,953	41,471	78,443	88,116	624,973
Total	435,070	482,300	310,792	127,228	229,064	243,567	1,828,020

TABLE B.3 Population Counts from Sample Frame
SUPREG by ENLSMPL by AGESMPL

Super Region	Enrolled			Not Enrolled			
	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
1	140,870	159,555	101,497	54,085	94,561	99,830	650,398
2	153,368	158,777	98,900	29,896	53,037	58,137	552,115
3	139,318	165,044	114,288	41,539	77,853	87,465	625,507
Total	433,556	483,376	314,685	125,520	225,451	245,432	1,828,020

TABLE B.4 Unweighted Values from Selected Sample Dataset
REGSMPL by ENLSMPL by AGESMPL

	Enrolled			Not Enrolled			
Region	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
Northeast	648	882	663	717	819	765	4,494
Mid-Atlantic	868	896	701	462	591	516	4,034
Southeast	774	938	719	533	676	747	4,387
Gulfsouth	383	460	399	391	440	440	2,513
Heartland	468	544	422	498	575	616	3,123
Southwest	833	896	753	811	880	781	4,954
Central	840	973	663	700	827	754	4,757
Southern California	610	572	387	367	439	424	2,799
Golden Gate	158	166	147	128	183	186	968
Northwest	307	372	264	241	314	306	1,804
Hawaii	211	199	109	62	62	61	704
Alaska	114	134	74	40	58	43	463
Total	6,214	7,032	5,301	4,950	5,864	5,639	35,000

TABLE B.5 Weighted values from Selected Sample Dataset
REGSMPL by ENLSMPL by AGESMPL

	Enrolled			Not Enrolled			
Region	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
Northeast	45,467	59,937	38,522	23,626	38,969	39,586	246,107
Mid-Atlantic	61,346	61,810	40,763	15,248	28,654	26,940	234,761
Southeast	53,911	65,015	44,887	13,473	27,677	33,948	238,911
Gulfsouth	27,435	32,177	24,727	10,084	17,582	19,914	131,918
Heartland	32,782	37,522	24,516	16,081	27,410	32,089	170,400
Southwest	57,508	60,169	41,603	14,838	25,137	24,674	223,930
Central	59,274	68,180	41,339	17,914	33,183	34,254	254,144
Southern California	42,403	38,721	21,433	6,985	12,720	13,492	135,754
Golden Gate	11,049	11,091	8,280	2,296	5,112	5,724	43,550
Northwest	21,341	24,847	14,581	4,813	8,877	9,585	84,044
Hawaii	14,578	13,794	6,071	1,146	2,023	2,029	39,641
Alaska	7,976	9,037	4,072	724	1,720	1,331	24,860
Total	435,070	482,300	310,792	127,228	229,064	243,567	1,828,020

TABLE B.6 Population Counts from Sample Frame
REGSMPL by ENLSMPL by AGESMPL

	Enrolled			Not Enrolled			
Region	0<=Age<=5	6<=Age<=12	13<=Age<=17	0<=Age<=5	6<=Age<=12	13<=Age<=17	Total
Northeast	47,431	59,632	39,965	23,420	39,466	40,834	250,748
Mid-Atlantic	61,949	63,277	37,796	14,500	26,294	27,800	231,616
Southeast	52,090	63,586	44,406	13,664	27,612	31,857	233,215
Gulfsouth	27,319	33,955	24,070	9,653	18,095	20,048	133,1440
Heartland	31,490	36,646	23,736	16,165	28,801	31,196	168,034
Southwest	57,319	63,136	41,666	13,955	23,261	24,531	223,868
Central	59,909	67,503	45,812	18,222	32,146	35,560	259,152
Southern California	41,326	38,587	22,127	6,855	12,317	13,446	134,658
Golden Gate	11,119	11,870	8,143	2,503	5,271	6,452	45,358
Northwest	21,533	23,899	16,281	4,475	8,671	10,094	84,953
Hawaii	14,656	13,578	6,295	1,270	2,162	2,176	40,137
Alaska	7,415	7,707	4,388	838	1,355	1,438	23,141
Total	433,556	483,376	314,685	125,520	225,451	245,432	1,828,020

TABLE B.7 Unweighted Values from Selected Sample Dataset
REGSMPLE by SEXSMPL

Region	Male	Female	Total
Northeast	2,294	2,200	4,494
Mid-Atlantic	2,058	1,976	4,034
Southeast	2,194	2,193	4,387
Gulfsouth	1,245	1,268	2,513
Heartland	1,585	1,538	3,123
Southwest	2,475	2,479	4,954
Central	2,343	2,414	4,757
Southern California	1,446	1,353	2,799
Golden Gate	488	480	968
Northwest	918	886	1,804
Hawaii	355	349	704
Alaska	230	233	463
Total	17,631	17,369	35,000

TABLE B.8 Weighted Values from Selected Sample Dataset
REGSMPLE by SEXSMPL

Region	Male	Female	Total
Northeast	125,014	121,093	246,107
Mid-Atlantic	118,927	115,835	234,761
Southeast	119,616	119,295	238,911
Gulfsouth	66,243	65,675	131,918
Heartland	86,205	84,195	170,400
Southwest	110,883	113,048	223,930
Central	125,690	128,454	254,144
Southern California	69,391	66,364	135,754
Golden Gate	21,185	22,366	43,550
Northwest	43,400	40,644	84,044
Hawaii	19,894	19,747	39,641
Alaska	12,471	12,388	24,860
Total	918,917	909,103	1,828,020

TABLE B.9 Population Counts from Sample Frame
REGSMPLE by SEXSMPL

Region	Male	Female	Total
Northeast	127,802	122,946	250,748
Mid-Atlantic	117,889	113,727	231,616
Southeast	118,141	115,074	233,215
Gulfsouth	67,699	65,441	133,140
Heartland	85,759	82,275	168,034
Southwest	113,567	110,301	223,868
Central	132,055	127,097	259,152
Southern California	68,771	65,887	134,658
Golden Gate	23,077	22,281	45,358
Northwest	43,663	41,290	84,953
Hawaii	20,528	19,609	40,137
Alaska	11,884	11,257	23,141
Total	930,835	897,185	1,828,020

APPENDIX C

VARIABLES DELIVERED TO NRC

LIST OF VARIABLES IN THE DATA SET DELIVERED TO NRC (FORM C)

#	Variable	Type	Length	Label	Values	Source
1	ACV	Char	1	Alternate Care Value	A = Active-duty Prime enrollee D = TRICARE Senior Prime enrollee E = TRICARE Prime enrollee U = Enrolled to Uniformed Services Family Health Plan (formerly USTFs) Blank = Not enrolled in TRICARE Prime or USFHP	DEERS
2	AGESMPL	Num	1	Age Sampling Variable	1 = 5 years or less 2 = 6 to 12 years 3 = 13 years or more	MPR
3	DAGEQY	Char	3	Beneficiary Age at time of Deers Extract	17 or younger, Blank as missing	DEERS
4	DBENCAT	Char	3	Beneficiary Category	ACT = Active Duty DA = Dependent of Active Duty GRD = Guard/Reserve DGR = Dependent of Guard/Reserve RET = Retiree DR = Dependent of Retiree DS = Survivor OTH = Other Z = Unknown	DEERS
5	DCATCH	Char	4	Catchment Area at Time of Extract		DEERS
6	DHSRGN	Char	2	Health Service Region	01 - Northeast 02 - Mid-Atlantic 03 - Southeast 04 - Gulf South 05 - Heartland 06 - Southwest 07 - Central 08 - Central 09 - Southern California 10 - Golden Gate 11 - Northwest 12 - Hawaii AK - Alaska 13 - Europe 14 - Pacific 15 - Latin America/Canada XX/ZZ - Unknown	DEERS
7	DMEDELG	Char	1	Medical Privilege Code	1 - Direct Care Only 2 - Direct Care and CHAMPUS 4 - Transitional Direct Care Only 5 - Transitional Direct Care and CHAMPUS 6 - Transitional Direct Care and Medicare 7 - Direct Care and Medicare	
8	DPRISM	Char	4	PRISM (20 mile) clinic service area		DEERS
9	DSPONSVC	Char	1	Derived Sponsor Branch of Service	A = Army C = Coast Guard F = Air Force M = Marine Corps N = Navy V = Navy Afloat X = Other Z = Unknown	DEERS
10	E1	Char	1	Eligibility Indicator - Period 1	Y = Yes, DEERS Eligible Period 1 N = No, Not DEERS Eligible Period 1	MPR
11	E2	Char	1	Eligibility Indicator - Period 2	Y = Yes, DEERS Eligible Period 2 N = No, Not DEERS Eligible Period 2	MPR

#	Variable	Type	Length	Label	Values	Source
12	E3	Char	1	Eligibility Indicator – Period 3	Y = Yes, DEERS Eligible Period 2 N = No, Not DEERS Eligible Period 2	MPR
13	ENBGSMPL	Char	2	Enrollment by beneficiary category	01 = "Active duty" 02 = "Active-duty fam,Prime,civ PCM" 03 = "Active-duty fam,Prime,mil PCM" 04 = "Active-duty fam,non-enrollee" 05 = "Retired,<65,civ PCM" 06 = "Retired,<65,mil PCM" 07 = "Retired,<65,non-enrollee"	MPR
14	ENLSMPL	Num	1	Enrollment Group	1 - Enrolled 2 - Not enrolled	MPR
15	ENRID	Char	4	Enrollment DMISID		DEERS
16	HADDFLG	Num	1	Residential Address - FLAG	0 = No address line1 1 = Address line1 present	DEERS
17	LEGDDSCD	Char	2	DEERS Dependent Suffix	01-16 = Dependent child 20 = Sponsor 30-36 = Spouse of sponsor 40-42 = Mother of sponsor 45-48 = Father of sponsor 50-54 = Mother-in-law of sponsor 55-56 = Father-in-law of sponsor 60 = Children where number greater than 19	DEERS
18	MACITYNM	Char	20	Residential Address - City		DEERS
19	MACTRYCD	Char	2	Residential Address, Country		DEERS
20	MALN1TX	Char	40	Residential Address - Line1		DEERS
21	MALN2TX	Char	40	Residential Address - Line2		DEERS
22	MAPRZIP	Char	5	Residential Address - ZIP		DEERS
23	MAPRZIPX	Char	4	Residential Address - ZIPX		DEERS
24	MASTCD	Char	2	Residential Address - State		DEERS
25	MBRRELCD	Char	1	Member Relationship Code	A = Self B = Spouse C = Child or stepchild D = Ward (not court ordered) E = Ward (court ordered) F = Dependent parent, stepparent, parent-in-law, or stepparent-in-law G = Surviving spouse H = Former spouse (20/20/20) I = Former spouse (20/20/15) J = Former spouse (10/20/10) K = Former spouse (transitional assistance (composite))	DEERS
26	MEDTYPE	Char	1	Medicare Eligibility	A - Medicare A B - Medicare B Blank - Neither Apply	
27	MPRID	Char	8	Unique MPR Identifier		MPR
28	MRTLSTAT	Char	1	Marital Status	A = Annulled D = Divorced I = Interlocutory decree L = Legally separated M = Married N = Never married S = Single / Not married [nonstandard] W = Widow or widower Z = Unknown	DEERS

#	Variable	Type	Length	Label	Values	Source
29	PATCAT	Char	7	Aggregated Beneficiary Category	ACTDTY = Active Duty and Guard/Reserve (no age cut). DEPACT = Dependent of Active Duty & Guard/Reserve (no age cut). NADD<65 = Retiree, Dependent of Retiree, Survivor, & Other under the age of 65. NADD65+ = Retiree, Dependent of Retiree, Survivor, & Other 65 years of age and older. UNKNOWN = Unknown (Derived Beneficiary Category equal to Z)	DEERS
30	PAYPLNCD	Char	5	Pay Plan Code		DEERS
31	PCM	Char	3	Enrolled to a Military or Civilian PCM	CIV = DMIS values of ‘8000’ to ‘8050’, or ‘6900’ to ‘6916’, or ‘7900’ to ‘7916’, or ‘0190’ to ‘0199’ (these last codes are USFHP enrollees). MIL = All other enrollment DMIS Codes. Blank = Not enrolled to TRICARE Prime or USFHP	DEERS
32	PGCD	Char	2	Pay Grade	00 = Unknown 00 – ZZ (not WW) = Used when pay plan is civil service 01 = Used when pay plan is cadet 01 – 05 = Used when pay plan is warrant officer 01 – 09 = Used when pay plan is enlisted 01 – 11 = Used when pay plan is officer	DEERS
33	PN1STNM	Char	20	Beneficiary First Name		DEERS
34	PNBRTHDT	Char	8	Beneficiary Date of Birth		DEERS
35	PNCDNCY	Char	4	Beneficiary Generation		DEERS
36	PNID	Char	9	Beneficiary/Dependent SSN		DEERS
37	PNLCATCD	Char	5	Personnel Category Code (Duty Status)	A = Active duty B = Presidential Appointee C = DoD civil service D = Disabled American veteran E = DoD contractor F = Former member H = Medal of Honor I = Other Government Agency Employee J = Academy student L = Lighthouse service M = Non-government Agency Personnel N = National Guard O = Other Government Agency Contractor Q = Reserve retiree R = Retired T = Foreign military U = Foreign national employee V = Reserve	DEERS
38	PNLSTNM	Char	26	Beneficiary Last Name		DEERS
39	PNMIDNM	Char	20	Beneficiary Middle Name		DEERS
40	PNSEXCD	Char	1	Beneficiary Sex	F = Female M = Male Z = Unknown	DEERS

#	Variable	Type	Length	Label	Values	Source
41	PNTYPCD	Char	1	Beneficiary Type Code	B = Both sponsor and dependent (i.e., the person has a joint marriage spouse) D = Dependent O = Other (e.g., someone who collapses in front of a military hospital and is treated at the hospital) S = Sponsor X = Prior sponsor (e.g., a sponsor who has been archived) Y = Prior dependent (e.g., a dependent who has been archived)	DEERS
42	PRN	Num	8	Permanent Random Number		MPR
43	RACEETHN	Char	1	Beneficiary Race/Ethnicity	A = American Indian or Alaskan Native B = Asian or Pacific islander C = Black (not Hispanic) D = White (not Hispanic) E = Hispanic X = Other Z = Unknown	DEERS
44	RANKCD	Char	6	Rank Code	See RANKCD.DOC for list of values	DEERS
45	REGSMPL	Num	2	Health Service Region Sampling Variable	01 - Northeast 02 - Mid-Atlantic 03 - Southeast 04 - Gulf South 05 - Heartland 06 - Southwest 07 - Central 08 - Central 09 - Southern California 10 - Golden Gate 11 - Northwest 12 - Hawaii 16 - Alaska 13 - Other	MPR
46	SADDFLG	Num	1	Sponsor Address - FLAG	0 = No address line1 1 = Address line1 present	DEERS
47	SPCITYNM	Char	20	Sponsor Address - City		DEERS
48	SPCTRYCD	Char	2	Sponsor Address, Country		DEERS
49	SPDUPID	Char	1	Family Sequence Number	1 = First occurrence of an SSN 2 = Second occurrence of an SSN 3 = Third occurrence of an SSN 4 = Fourth occurrence of an SSN	DEERS
50	SPLN1TX	Char	40	Sponsor Address - Line1		DEERS
51	SPLN2TX	Char	40	Sponsor Address - Line2		DEERS
52	SPONSSN	Char	9	Sponsor Social Security Number		DEERS
53	SPPRZIP	Char	5	Sponsor Residential Address - ZIP		DEERS
54	SPPRZIPX	Char	4	Sponsor Address - ZIPX		DEERS
55	SPSTCD	Char	2	Sponsor Residential Address - State		DEERS
56	SPTNUMCD	Char	14	Sponsor Phone Number		DEERS
57	SSNSMPL	Char	12	SPONSSN || SPDUPID || LEGDDSCD SSN Sampling Variable		MPR
58	STRATUM	Char	7	Stratum		MPR
59	SUPREG	Num	1	Super Region	1 = New regions (1,2,5) 2 = Mature regions (6,9-12,16) 3 = Other regions (3,4,7,8)	MPR

#	Variable	Type	Length	Label	Values	Source
60	SVCCD	Char	1	Branch of Service	A = Army N = Navy M = Marine Corps F = Air Force C = Coast Guard D = Office of the Secretary of Defense H = The Commissioned Corps of the PHS O = The Commissioned Corps of the NOAA 1 = Foreign Army 2 = Foreign Navy 3 = Foreign Marine Corps 4 = Foreign Air Force X = Not applicable	DEERS
61	TNUMCD	Char	14	Residence Telephone Number		DEERS
62	UADDFLG	Num	1	Unit Address - FLAG	0 = No address line1 1 = Address line1 present	DEERS
63	UICADD1	Char	30	Unit Address - Line1		DEERS
64	UICADD2	Char	30	Unit Address - Line2		DEERS
65	UICCITY	Char	30	Unit Address - City		DEERS
66	UICST	Char	2	Unit Address - State		DEERS
67	UICZIP	Char	5	Unit Address - ZIP		DEERS
68	ULOCDMIS	Char	4	Unit Address - DMIS Code		DEERS
69	ULOCGRN	Char	2	Unit Address - Region		DEERS

APPENDIX D

SAS CODE

1. Constructing the Extract and Crosswalk Files

```
****************************************************************************
*
* PROGRAM: STI.SAS
* TASK:    DOD Health Care Survey, Sampling (8860-210/220)
* PURPOSE: Split STI2003 raw datasets into smaller parts for CDs and
*          convert entire dataset into SAS/SD2 format.
*
* WRITTEN: 10/18/2000 BY KEITH RATHBUN
*
* MODIFIED: 1) 04/22/2002 BY KEITH RATHBUN, Removed TSPSITE from FREQs.
*
* INPUTS:
*
* 1) STI2003.001 - RAW 2003 Q3 DEERS Population Extract File (Tape Part 1)
* 2) STI2003.002 - RAW 2003 Q3 DEERS Population Extract File (Tape Part 2)
*
* OUTPUTS:
*
* 1) STI001.SD2 - 2003 Q3 DEERS Population Extract File (CD Part 1)
* 2) STI002.SD2 - 2003 Q3 DEERS Population Extract File (CD Part 2)
* 3) STI003.SD2 - 2003 Q3 DEERS Population Extract File (CD Part 3)
* 4) STI004.SD2 - 2003 Q3 DEERS Population Extract File (CD Part 4)
*
* INCLUDES:
*
* 1) LAYOUT.SAS - Input STEP For Raw Data From STI
*
* NOTES:
*
* 1) The tape file sent by STI exceeded 4 GB in size.  The tape software
*    crashed the computer at the 4 GB unload point.  In order to successfully
*    unload this file, I split the tape file into two parts (STI2003.001
*    and STI2003.002).
* 2) Under the new contract (8860), the suvey year was changed
*    to be based on the year the survey is administered (2002)
*    as opposed to the questioning reference frame (2001).  This program
*    references folders named according to the new convention [i.e.
*    the survey administration year (2002 for project 8860)].
*
****************************************************************************
*;
LIBNAME OUT V612 "D:\KEITH\DOD\Q3_2003\DATA";
OPTIONS PS=79 LS=132 COMPRESS=YES NOCENTER;

****************************************************************************
* PROCESS - MACRO PARAMETERS:
* 1) INUM  = Raw Input file extension
* 2) ONUM1 = SAS Output file 1 suffix
* 3) ONUM2 = SAS Output file 2 suffix
****************************************************************************;
%MACRO PROCESS(INUM=,ONUM1=,ONUM2=);

FILENAME  IN  "D:\KEITH\DOD\Q3_2003\DATA\STI2003.&INUM";

DATA OUT.STI&ONUM1 OUT.STI&ONUM2;
   INFILE IN LRECL=99999 RECFM=V MISSOVER;
```

```
   %INCLUDE "D:\KEITH\DOD\Q3_2003\SAMPLING\LAYOUT.SAS";
   IF _N_ LE 2500000 THEN OUTPUT OUT.STI&ONUM1;
   ELSE OUTPUT OUT.STI&ONUM2;
RUN;

%MEND PROCESS;
****************************************************************************
* END PROCESS MACRO
****************************************************************************;

%PROCESS(INUM=001,ONUM1=001,ONUM2=002);
%PROCESS(INUM=002,ONUM1=003,ONUM2=004);

****************************************************************************
* PRINTIT - MACRO PARAMETERS:
* 1) PNUM = SAS output file suffix
****************************************************************************;
%MACRO PRINTIT(PNUM=);

TITLE1 "DOD Health Care Survey, Sampling (8860-210/220)";
TITLE2 "PROGRAM: STI.SAS,WRITTEN BY: KEITH RATHBUN, April 2003";
TITLE3 "OUTPUT: STI&PNUM..SD2";

PROC CONTENTS DATA=OUT.STI&PNUM; RUN;

PROC FREQ DATA=OUT.STI&PNUM;
   TABLES
      PNTYPCD
      MRTLSTAT
      PNSEXCD
      PNARSNCD
      MDCABRSN
      LEGDDSCD
      PNLCATCD
      SVCCD
      PAYPLNCD
      PGCD
      MBRRELCD
      RANKCD
      ULOCGRN
      ULOCDMIS
      RACEETHN
      DCATCH
      DMEDELG
      DAGEQY
      DBENCAT
      DPRISM
      DHSRGN
      DSPONSVC
      MEDTYPE
      ENRID
      ACV
      PCM
      PATCAT
   /MISSING LIST;
RUN;
%MEND PRINTIT;
****************************************************************************
* END PRINTIT MACRO
```

```
***************************************************************************;

%PRINTIT(PNUM=001);
%PRINTIT(PNUM=002);
%PRINTIT(PNUM=003);
%PRINTIT(PNUM=004);
```

```
*************************************************************************
*
* PROGRAM:  EXTRACTC.SAS
* TASK:     DOD Health Care Survey, Sampling (8860-220)
* PURPOSE:  Build SAS extract file for the DOD sample
*
* WRITTEN:  10/19/2000 BY KEITH RATHBUN
*
* MODIFIED:
* 1) 01/18/2001 BY KEITH RATHBUN - Small changes for Q2 processing.
*    Removed sorting of XWALK and EXTRACT files by MPRID.
* 2) 02/08/2001 BY KEITH RATHBUN - Small changes for Q3 processing.
*    Added specific family exclusion criteria as include file.
* 3) 04/16/2001 BY KEITH RATHBUN - Small changes for child Q3 processing.
* 4) 04/23/2002 BY KEITH RATHBUN for Q3 2002 processing and removed TSPSITE.
* 5) 04/09/2003 BY KEITH RATHBUN for Q3 2003 processing.
*
* INPUTS:
* 1) STI001.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 1)
* 2) STI002.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 2)
* 3) STI003.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 3)
* 4) STI004.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 4)
* 5) XWALKC.SD2 - DEERS Child Population XWALK SAS data set (sorted by
SSNSMPL)
*
* OUTPUTS:
* 1) EXTRACTC.SD2 - DEERS Child Population EXTRACT SAS data set (complete -
sorted by SSNSMPL)
*
* INCLUDES:
* 1) EXCLUDE.SAS - Exclude specific family by SPONSSN.
*
* NOTES:
* 1) Under the new contract (8860), the suvey year was changed
*    to be based on the year the survey is administered (2002)
*    as opposed to the questioning reference frame (2001).  This program
*    references folders named according to the new convention [i.e.
*    the survey administration year (2002 for project 8860)].
*
*************************************************************************
*;
LIBNAME  IN  V612 "D:\KEITH\DOD\Q3_2003\DATA";
LIBNAME  OUT V612 "D:\KEITH\DOD\Q3_2003\DATA";
OPTIONS PS=79 LS=132 COMPRESS=YES NOCENTER;

*************************************************************************
* Set up MACRO to exclude specific families from survey.
*************************************************************************;
%INCLUDE "D:\KEITH\DOD\Q3_2003\SAMPLING\EXCLUDE.SAS";

*************************************************************************
* Extract key sampling variables.
*************************************************************************;
%MACRO SORTIT(NUM=);
   PROC SORT DATA=IN.STI&NUM
            (KEEP=SSNSMPL  PNTYPCD  MRTLSTAT  PNSEXCD
                  PNARSNCD MDCABRSN MDCAEFDT  MDCAEXDT
                  LEGDDSCD PNLCATCD SVCCD     PAYPLNCD
                  PGCD     MBRRELCD RANKCD    ULOCGRN
```

```
                  ULOCDMIS RACEETHN DCATCH    DMEDELG
                  DAGEQY   DBENCAT  DPRISM    DHSRGN
                  DSPONSVC MEDTYPE  ENRID     ACV
                  PCM      PATCAT)
              OUT=STI&NUM;
       BY SSNSMPL;
   RUN;
%MEND SORTIT;

%SORTIT(NUM=001);
%SORTIT(NUM=002);
%SORTIT(NUM=003);
%SORTIT(NUM=004);

****************************************************************************
* KEEP children (<18) and exclude specific families.
****************************************************************************;
DATA EXTRACT;
   SET STI001
       STI002
       STI003
       STI004
   ;
   BY SSNSMPL;
   IF NOT (DAGEQY GE "018" OR (DAGEQY = " " AND LEGDDSCD GE "20"));
   *************************************************************************
   * STI sent duplicates SSNSMPLs.  So, we let SAS remove them here.
   *************************************************************************;
   IF FIRST.SSNSMPL;
   *************************************************************************
   * Exclude specific families from survey.
   *************************************************************************;
   &EXCLUDE;
RUN;

DATA OUT.EXTRACTC;
   MERGE IN.XWALKC(IN=IN1) EXTRACT(IN=IN2);
   BY SSNSMPL;
   IF IN1 AND IN2;
   DROP SSNSMPL;
RUN;

TITLE1 "Build SAS EXTRACT file for the 2003 DOD Q3 child sample";
TITLE2 "Program Name: EXTRACTC.SAS, Written by Keith Rathbun, April 2003";

TITLE3 "CONTENTS of extract file";
PROC CONTENTS DATA=OUT.EXTRACTC; RUN;

TITLE3 "FREQS of key variables - 2003 Q3 DEERS child population extract:
EXTRACTC.SD2";
PROC FREQ DATA=OUT.EXTRACTC;
   TABLES
      PNTYPCD
      MRTLSTAT
      PNSEXCD
      PNARSNCD
      MDCABRSN
      LEGDDSCD
      PNLCATCD
```

```
      SVCCD
      PAYPLNCD
      PGCD
      MBRRELCD
      RANKCD
      ULOCGRN
      ULOCDMIS
      RACEETHN
      DCATCH
      DMEDELG
      DAGEQY
      DBENCAT
      DPRISM
      DHSRGN
      DSPONSVC
      MEDTYPE
      ENRID
      ACV
      PCM
      PATCAT
  /MISSING LIST;
RUN;
```

```
****************************************************************************
*
* PROGRAM:  XWALKC.SAS
* TASK:     DOD Health Care Survey, Adult Sampling (8860-220)
* PURPOSE:  Build SAS extract/cross-walk file for the DOD sample
*           and assign permanent random numbers (PRN).
*
* WRITTEN:  01/17/2001 BY KEITH RATHBUN
*
* MODIFIED:
* 1) 02/08/2001 BY KEITH RATHBUN for Q3 processing.  Also, added
*    specific family exclusion criteria as include file.
* 2) 07/09/2001 BY KEITH RATHBUN for Q4 processing.  Removed Q3-specific
*    processing.
* 3) 10/09/2001 BY KEITH RATHBUN for Q1 2002 processing.
* 4) 01/22/2002 BY KEITH RATHBUN for Q2 2002 processing.
* 5) 04/10/2002 BY KEITH RATHBUN for Q3 2002 processing.
* 6) 04/11/2002 BY KEITH RATHBUN, adapted from XWALK.SAS for Q3
*    2002 child processing.
* 7) 04/09/2003 BY KEITH RATHBUN, updated for Q3 2003 child processing.
*
* INPUTS:
* 1) STI001.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 1)
* 2) STI002.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 2)
* 3) STI003.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 3)
* 4) STI004.SD2 - 2003 Q3 DEERS Population SSN SAS data set (Part 4)
* 5) XWALKC.SD2 - 2002 Q3 DEERS Child Population XWALK SAS data set
*
* OUTPUTS:
* 1) XWALKC.SD2 - 2003 Q3 DEERS Child Population XWALK SAS data set
* 2) SEEDC.SD2  - 2003 Q3 DEERS Child Random SEED SAS data set
*
* INCLUDES:
* 1) EXCLUDE.SAS - Exclude specific family by SPONSSN.
*
* NOTES:
* 1) Under the new contract (8860), the suvey year was changed
*    to be based on the year the survey is administered (2002)
*    as opposed to the questioning reference frame (2001).  This program
*    references folders named according to the new convention [i.e.
*    the survey administration year (2002 for project 8860)].
*
****************************************************************************;
LIBNAME  IN1 V612 'D:\KEITH\DOD\Q3_2002\DATA'; * Previous XWALK;
LIBNAME  IN2 V612 'D:\KEITH\DOD\Q3_2003\DATA'; * Current STI Tape Files;
LIBNAME  OUT V612 'D:\KEITH\DOD\Q3_2003\DATA'; * Current Output;
OPTIONS PS=79 LS=132 COMPRESS=NO NOCENTER;

****************************************************************************
* Set period number as global variable.  Period 1,2,3 represent 2001, 2002,
* 2003 survey administration years.
****************************************************************************;
%LET PD = 3;

****************************************************************************
* Set up MACRO to exclude specific families from survey.
****************************************************************************;
%INCLUDE "D:\KEITH\DOD\Q3_2002\SAMPLING\EXCLUDE.SAS";
```

```
TITLE1 "Generate Child XWALK file from 2003 Q3 DOD DEERS Population Extract
File";
TITLE2 "Program Name: XWALKC.SAS, Written by Keith Rathbun, April 2003";

****************************************************************************
* Assign random SEED as global variable.  This will later be used as the
* starting point for random numbering.
****************************************************************************;
DATA OUT.SEEDC;
   SEED = INT(RANUNI(0)*1000000+1);
   CALL SYMPUT("SEED",SEED);
   PUT "Random SEED assigned for generating the permanent radom numbers: "
SEED;
RUN;

TITLE3 "Random SEED assigned for generating the permanent radom numbers:
SEEDC.SD2";
PROC PRINT; RUN;

****************************************************************************
* Assign LASTID from previous XWALKC file as global variable.  This will later
* be used as the starting point for assigning new MPRIDs.
****************************************************************************;
DATA _NULL_;
   SET IN1.XWALKC END=FINISHED;
   LENGTH MPRIDX 8; RETAIN MPRIDX;
   IF MPRID > MPRIDX THEN MPRIDX = MPRID;
   IF FINISHED THEN CALL SYMPUT("LASTID",MPRIDX);
RUN;

****************************************************************************
* Get SSNSMPLs from current quarter tape file.
****************************************************************************;
%MACRO SORTIT(NUM=);
   PROC SORT DATA=IN2.STI&NUM (KEEP=SSNSMPL LEGDDSCD DAGEQY) OUT=STI&NUM;
       BY SSNSMPL;
   RUN;
%MEND SORTIT;

%SORTIT(NUM=001);
%SORTIT(NUM=002);
%SORTIT(NUM=003);
%SORTIT(NUM=004);

****************************************************************************
* Remove children (<18) prior to assigning permanent random number (PRN).
****************************************************************************;
DATA SSN_Q(KEEP=SSNSMPL);
   SET STI001
       STI002
       STI003
       STI004
      ;
   BY SSNSMPL;
   IF NOT (DAGEQY GE "018" OR (DAGEQY = " " AND LEGDDSCD GE "20"));
   *************************************************************************
   * STI sent duplicates SSNSMPLs.  So, we let SAS remove them here.
   *************************************************************************;
   IF FIRST.SSNSMPL;
```

```
   *************************************************************************
   * Exclude specific families from survey.
   *************************************************************************;
   &EXCLUDE;
RUN;

****************************************************************************
* Combine Qn SSNSMPLs with previous XWALK (SSN_OLD) keeping only the
* new eligibles (SSN_NEW).
****************************************************************************;
DATA SSN_NEW OLDXWALK;
   MERGE SSN_Q(IN=IN1 KEEP=SSNSMPL) IN1.XWALKC(IN=IN2);
   BY SSNSMPL;

   *************************************************************************
   * Assign eligibility indicator for new eligibles.
   *************************************************************************;
   LENGTH E&PD $1;
   IF IN1 AND IN2 THEN E&PD = "Y";
   ELSE IF IN1    THEN E&PD = "Y";
   ELSE IF IN2    THEN E&PD = "N";
   LABEL E&PD = "Eligibility indicator for period = &PD";

   IF IN1 AND NOT IN2 THEN OUTPUT SSN_NEW;
   IF IN2 THEN OUTPUT OLDXWALK;
RUN;

****************************************************************************
* Assign PRN for all new eligibles.
****************************************************************************;
DATA NEWXWALK (KEEP=MPRID SSNSMPL PRN E&PD);
   SET SSN_NEW;
   LENGTH MPRID  $8;
   *************************************************************************
   * Assign eligibility indicator for new eligibles.
   *************************************************************************;
   LENGTH E&PD   $1;
   E&PD = "Y";
   LABEL E&PD = "Eligibility indicator for period = &PD";
   *************************************************************************
   * Assign PRN for new eligibles.
   *************************************************************************;
   PRN = RANUNI(&SEED);
   LABEL PRN = "Permanent Random Number";
   *************************************************************************
   * Assign MPRID starting with previous XWALKs LASTID+1.
   *************************************************************************;
   IF _N_ = 1 THEN MPRIDX = %EVAL(&LASTID+1);
   ELSE MPRIDX + 1; RETAIN MPRIDX;
   MPRID  = PUT(MPRIDX,Z8.);
RUN;

%MACRO XWALKC;
DATA OUT.XWALKC;
   SET NEWXWALK OLDXWALK;
   BY SSNSMPL;
   *************************************************************************
   * Recode missing values to Not eligible.
   *************************************************************************;
```

```
   %DO I = 1 %TO &PD;
       IF E&I = " " THEN E&I = "N";
   %END;
RUN;
%MEND XWALKC;
%XWALKC;

TITLE3 "XWALK file: XWALKC.SD2";
PROC CONTENTS; RUN;

PROC FREQ;
   TABLES E1-E&PD E1*E2*E3 /MISSING LIST;
RUN;
```

2. Construct Sample Frame C

```
****************************************************************************
***
*** Project:        2003 Health Care Survey of DoD Beneficiaries - Child
*** Project Number: 8860
*** Task Number:    220
***
*** Purpose: Create the frame for the child survey.
***
*** Date:       April 23, 2003
*** Programmer: Nancy A. Clusen
***
*** Program: D:\projects\8860-210\child_2003\framec.sas,
***          Creates the child sampling frame.
***
*** Inputs:  \\DOD\Q3_2003\Data\Afinal\extractc.sd2
***          Extracted DoD data set used to create the child sampling frame.
***
***          E:\xwalkc.sd2
***          Provides the family identifier (on CD).
***
*** Outputs: D:\projects\8860-210\child_2003\Data\framec.sd2
***          Child sampling frame created from the extracted DoD data set.
***
*** Notes: None
***
****************************************************************************;

*** Setup the titles. ***;
title1 '2003 Health Care Survey of DoD Beneficiaries - Child';
title2 'Program: D:\projects\8860-210\child_2003\framec.sas by Nancy A.
Clusen';
title3 'Create the Child Sampling Frame';

*** Setup the options. ***;
options ls=132 ps=79 nocenter compress=yes;

*** Setup the paths where the files are located. ***;
libname inx  'F:\DOD\Q3_2003\Data\Afinal';
*libname ine v6 'E:\';
*libname in  v6 'D:\projects\8860-210\child_2003\Data';
libname in   'D:\projects\8860-210\child_2003\Data';
libname out  'D:\projects\8860-210\child_2003\Data';

title5 'Check the Contents of the Extracted DoD Data Set';
proc contents data=inx.extractc;
run;

title5 'Check Some Important Variables';
proc freq data=inx.extractc;
table dageqy patcat pcm dhsrgn pnsexcd svccd / list missing;
run;

title5 'Check the Contents for the Family Identifier';
proc contents data=in.xwalkc;
run;
```

```
*** Create the formats. ***;
proc format;
   value agesmpl 1 = 'Younger than 6'
                 2 = '6 to 12'
                 3 = '13 to 17'
             4 = 'Other'
             other = 'Error';
   value bgcsmpl 1 = 'Active Duty'
                 2 = 'Active Duty Family Member'
             3 = 'Retirees and Family Member'
             4 = 'Other'
             other = 'Error';
   value enlsmpl 1 = 'Enrolled'
                 2 = 'Not Enrolled'
                 4 = 'Other'
                other = 'Error';
   value regsmpl 1 = 'Northeast'
                 2 = 'Mid-Atlantic'
             3 = 'Southeast'
             4 = 'Gulf South'
             5 = 'Heartland'
             6 = 'Southwest'
             7 = 'Central'
                 8 = 'Central'
             9 = 'Southern California'
            10 = 'Golden Gate'
            11 = 'Northwest'
            12 = 'Hawaii'
            16 = 'Alaska'
            13 = 'Other'
            other = 'Error';
   value supreg 1 = 'New Region'
                2 = 'Mature Region'
                3 = 'Other Region'
                4 = 'Not a Region'
                other = 'Error';
   value sexsmpl 1 = 'Male'
                 2 = 'Female'
             3 = 'Other'
             other = 'Error';
   value svcsmpl 1 = 'Army'
                 2 = 'Navy'
             3 = 'Air Force'
             4 = 'Marine Corps'
             5 = 'Coast Guard'
             6 = 'Other'
             other = 'Error';
run;

*** Sort the data sets. ***;
proc sort data=inx.extractc;
by mprid prn;
run;

proc sort data=in.xwalkc;
by mprid prn;
run;

*** Merge the data sets to create the frame. ***;
```

```
data out.framec;
merge inx.extractc in.xwalkc;
by mprid prn;
*** Create the age group stratification variable, agesmpl. ***;
if dageqy = ' ' then agesmpl = 1;
else if '000' <= dageqy < '006' then agesmpl = 1;
else if '006' <= dageqy < '013' then agesmpl = 2;
else if '013' <= dageqy <= '017' then agesmpl = 3;
else agesmpl = 4;
*** Create a numberic age variable, age_n. ***;
age_n = input(dageqy,3.0);
if age_n = . then age_n = 0;
*** Create the beneficiary group variable, bgcsmpl. ***;
if patcat = 'DEPACT' then bgcsmpl = 2;
else if patcat = 'NADD<65' then bgcsmpl = 3;
else if patcat = 'ACTDTY' then bgcsmpl = 1;
else bgcsmpl = 4;
*** Create the enrollment status of beneficiary variable, enlsmpl. ***;
*** Changed form 3 levels to 2 levels Dod q3 2002. ***;
if pcm in( 'MTF', 'CIV') then enlsmpl = 1;
else if pcm = ' ' then enlsmpl = 2;
else enlsmpl = 4;
*** Create the geographic region variable, regsmpl. ***;
select (dhsrgn);
when ('01') regsmpl =  1;
when ('02') regsmpl =  2;
when ('03') regsmpl =  3;
when ('04') regsmpl =  4;
when ('05') regsmpl =  5;
when ('06') regsmpl =  6;
when ('07') regsmpl =  7;
when ('08') regsmpl =  8;
when ('09') regsmpl =  9;
when ('10') regsmpl = 10;
when ('11') regsmpl = 11;
when ('12') regsmpl = 12;
when ('AK') regsmpl = 16;
otherwise regsmpl = 13;
end;
*** Create the geographic area variable, supreg. ***;
if regsmpl in (1,2,5) then supreg = 1;
else if regsmpl in (6,9,10,11,12,16) then supreg = 2;
else if regsmpl in (3,4,7,8) then supreg = 3;
else supreg = 4;
*** Create the beneficiary gender variable, sexsmpl. ***;
*** Missing, Z or ' ', is considered male. ***;
if pnsexcd in ('M','Z') then sexsmpl = 1;
else if pnsexcd = ' ' then sexsmpl = 1;
else if pnsexcd = 'F' then sexsmpl = 2;
else sexsmpl = 3;
*** Create the branch of service variable, svcsmpl. ***;
select (svccd);
when ('A') svcsmpl = 1;
when ('N') svcsmpl = 2;
when ('F') svcsmpl = 3;
when ('M') svcsmpl = 4;
when ('C') svcsmpl = 5;
otherwise svcsmpl = 6;
end;
```

```
*** Create the sampling stratum. ***;
length stratum $3.;
stratum = put(supreg,1.) || put(enlsmpl,1.) || put(agesmpl,1.);
*** Create the family variable, family. ***;
family = input(substr(ssnsmpl,1,9),9.0);
*** Label the variables. ***;
LABEL SVCSMPL = 'SVCSMPL - Branch of Service'
      AGESMPL = 'AGESMPL - Age'
      SEXSMPL = 'SEXSMPL - Sex'
      REGSMPL = 'REGSMPL - Region'
      STRATUM = 'prelim STRATUM: supreg+enlsmpl+agesmpl'
      BGCSMPL = 'BGCSMPL - Beneficiary Group'
      ENLSMPL = 'ENLSMPL - Enrollment Sampling Group'
      SUPREG  = 'SUPREG - Super Region'
      FAMILY  = 'FAMILY - Family';
*** Exclude people not in the United States. ***;
if regsmpl ne 13 then output out.framec;
run;

title5 'Check the Constructed Variables';
proc freq data=in.framec;
table agesmpl*dageqy
      age_n*dageqy
      bgcsmpl*patcat
      enlsmpl*pcm
      regsmpl*dhsrgn
      supreg*regsmpl
      sexsmpl*pnsexcd
      svcsmpl*svccd
      / list missing;
format agesmpl agesmpl.
       bgcsmpl bgcsmpl.
       enlsmpl enlsmpl.
       regsmpl regsmpl.
       supreg supreg.
       sexsmpl sexsmpl.
       svcsmpl svcsmpl.;
run;

proc freq data=in.framec;
table stratum*supreg*enlsmpl*agesmpl / list missing;
run;

*** Create the family code variable, famcode. ***;
proc sort data=in.framec;
by family;
run;

data out.framec;
set in.framec;
by family;
retain famcode 0;
if first.family then famcode = famcode + 1;
label famcode = 'FAMCODE - Family Code';
run;

proc print data=out.framec (obs = 500);
var famcode family ssnsmpl;
run;
```

```
*** Create the sampling strata variable, sampstr. ***;
data out.framec;
set in.framec;
rannum = ranuni(8892704);
run;

proc sort data=in.framec;
by famcode rannum;
run;

data out.framec;
set in.framec;
by famcode;
retain sampstr '000';
if first.famcode = 1 then sampstr = stratum;
label sampstr = 'SAMPSTR-final sampling stratum';
;
run;

proc print data=in.framec (obs = 500);
var famcode sampstr;
run;

proc freq data=in.framec;
table sampstr*stratum / list missing;
run;

title5 'Population Counts by Stratification Variables';
proc freq data=in.framec;
table stratum sampstr / list missing;
run;
```

```
***************************************************************************
***
*** Project:        2003 Health Care Survey of DoD Beneficiaries - Child
*** Project Number: 8860
*** Task Number:    220
***
*** Purpose: Create the count data set for the child survey.  This consists
***          of the population counts by various cell definitions:
***
***              PSUM0 = Stratifcation Variable Count
***              PSUM1 = SUPREG Count
***              PSUM2 = ENLSMPL Count
***              PSUM3 = AGESMPL Count
***              TOTAL = Total Population
***
*** Date:       April  2003
*** Programmer: Nancy A. Clusen Initial program by Keith Rathbun.
***
*** Program: D:\projects\8860-210\child_2003\countc.sas,
***          Creates the child sampling frame.
***
*** Inputs:  D:\projects\8860-210\child_2003\Data\framec.sd2
***          Child sampling frame created from the extracted DoD data set.
***
*** Outputs: D:\projects\8860-210\child_2003\Data\countc.sd2
***          Population counts by various cell definitions.
***
*** Notes: None
***
***************************************************************************;

*** Setup the titles. ***;
title1 '2003 Health Care Survey of DoD Beneficiaries - Child';
title2 'Program:  D:\projects\8860-210\child_2003\countc.sas by Nancy A.
Clusen';
title3 'Create population counts by various cell definitions.';

*** Setup the options. ***;
options ls=132 ps=79 nocenter compress=yes mlogic mprint symbolgen;

*** Setup the paths where the files are located. ***;
libname in   'D:\projects\8860-210\child_2003\Data';
libname out  'D:\projects\8860-210\child_2003\Data';

*** Set the stratification variable. ***;
%let strata = sampstr;

data framec ;
set in.framec; /*(keep = sampstr prn)*/
supreg = input(substr(sampstr,1,1),1.);
enlsmpl = input(substr(sampstr,2,1),1.);
agesmpl = input(substr(sampstr,3,1),1.);
run;

TITLE5 "FREQS of sample FRAMEC.SD2";
PROC FREQ DATA=framec;
   TABLES &strata. SUPREG ENLSMPL AGESMPL
  /MISSING LIST;
```

```
RUN;

PROC SORT DATA=framec OUT=FRAMEC;
   BY &strata. SUPREG ENLSMPL AGESMPL;
RUN;

PROC MEANS DATA=FRAMEC NOPRINT;
   BY  &strata. SUPREG ENLSMPL AGESMPL;
   VAR ENLSMPL;
   OUTPUT
   OUT=T0(KEEP=&strata. SUPREG ENLSMPL AGESMPL)
   N=DUMMY;
RUN;

PROC FREQ DATA=FRAMEC NOPRINT;
   TABLES &strata.
  /MISSING LIST OUT=T1(RENAME=(COUNT=PSUM0)
                KEEP=COUNT &strata.) NOPERCENT NOCUM NOPRINT;
RUN;

PROC FREQ DATA=FRAMEC NOPRINT;
   TABLES SUPREG
  /MISSING LIST OUT=T2(RENAME=(COUNT=PSUM1)
                KEEP=COUNT SUPREG) NOPERCENT NOCUM NOPRINT;
RUN;

PROC FREQ DATA=FRAMEC NOPRINT;
   TABLES ENLSMPL
  /MISSING LIST OUT=T3(RENAME=(COUNT=PSUM2)
                KEEP=COUNT ENLSMPL) NOPERCENT NOCUM NOPRINT;
RUN;

PROC FREQ DATA=FRAMEC NOPRINT;
   TABLES AGESMPL
  /MISSING LIST OUT=T4(RENAME=(COUNT=PSUM3)
                KEEP=COUNT AGESMPL) NOPERCENT NOCUM NOPRINT;
RUN;

PROC SORT DATA=T0; BY &strata.; RUN;
DATA T0;
   MERGE T0 T1;
   BY  &strata.;
RUN;

PROC SORT DATA=T0; BY SUPREG; RUN;
DATA T0;
   MERGE T0 T2;
   BY  SUPREG;
RUN;

PROC SORT DATA=T0; BY ENLSMPL; RUN;
DATA T0;
   MERGE T0 T3;
   BY  ENLSMPL;
RUN;

PROC SORT DATA=T0; BY AGESMPL; RUN;

proc means data=framec noprint;
```

```
var prn;
output out=total n=total;
run;

DATA OUT.COUNTC;
if _n_=1 then set total(drop = _type_ _freq_);
   MERGE T0 T4;
   BY  AGESMPL;
   LABEL PSUM0 = 'PSUM0 - &strata. Count'
         PSUM1 = 'PSUM1 - SUPREG Count'
         PSUM2 = 'PSUM2 - ENLSMPL Count'
         PSUM3 = 'PSUM3 - AGESMPL Count'
         TOTAL = 'TOTAL Population'
        ;
RUN;

TITLE5 "Information for COUNTC.SD2";

PROC CONTENTS data=in.countc;
RUN;

PROC PRINT data=in.countc;
var &strata. supreg enlsmpl agesmpl psum0-psum3 total;
RUN;
```

3. Construct Sample C

```
***************************************************************************
*** Project:        2003 Health Care Survey of DoD Beneficiaries - Child
***
*** Purpose: Select the child sample from the child sampling frame.
***
*** Date:       April 30, 2002
*** Programmer: Nancy A. Clusen
***
*** Program:  D:\projects\8860-210\child_2003\samplc.sas,
***          Selects the sample from the child sampling frame.
***
*** Inputs:   D:\projects\8860-210\child_2003\data\framec.sd2
***          Child sampling frame created from the extracted DoD data set.
***
***           D:\projects\8860-210\child_2003\data\csize.sd2
***          Sample size information for each stratum.
***
*** Outputs: D:\projects\8860-210\child_2003\data\samplc.sd2
***          The child sample created from the child sampling frame.
***
*** Notes: None
***
***************************************************************************;

*** Setup the titles. ***;
title1 '2003 Health Care Survey of DoD Beneficiaries - Child';
title2 'Program:  D:\projects\8860-210\child_2003\data\samplc01.sas by Nancy';
title3 'Select the child sample from the child sampling frame.';

*** Setup the options. ***;
options ls=132 ps=79 nocenter compress=yes;

*** Setup the paths where the files are located. ***;
libname in  v6 'D:\projects\8860-210\child_2003\data';
libname out v8 'D:\projects\8860-210\child_2003\data';

*** Change the information in the csize data set to conform to proc
surveyselect. ***;
title5 'Information for the CSIZE.SD2 Data Set';
proc contents data=in.csize;
run;

data sam_size (keep = sampstr _nsize_);
set in.csize (rename = (nhff = _nsize_));
run;

proc contents data=sam_size;
run;

title5 'Information for the FRAMEC.SD2 Data Set';
data framec;
set in.framec;
run;

proc contents data=framec;
run;
```

```
proc sort data=framec;
by sampstr;
run;

proc sort data=sam_size;
by sampstr;
run;

title5 'Information for the Child Sample';
proc surveyselect
   data = framec
   out = samplc
   method = sys
   sampsize = sam_size
   sort = serp
   seed = 1465233
   stats;
strata sampstr;
control famcode regsmpl age_n sexsmpl;
run;

data out.samplc01;
   set samplc (rename = (samplingweight = BWT selectionprob = SEL_PROB));
run;

title5 'Check for Multiple Children with the Same Sponsor';
proc freq data=out.samplc01 noprint;
table famcode / list missing out=m_fam;
run;

data m_fam;
set m_fam (keep = famcode count);
if count > 1 then output m_fam;
run;

proc print data=m_fam;
run;

proc sort data=out.samplc01;
by famcode;
run;

proc sort data=m_fam;
by famcode;
run;

data out.multifam mf_only s_only;
merge out.samplc01 (in = in_s) m_fam (in = in_mf);
by famcode;
if in_s = 1 and in_mf = 1 then output out.multifam;
else if in_s = 0 and in_mf = 1 then output mf_only;
else if in_s = 1 and in_mf = 0 then output s_only;
run;

proc freq data=out.multifam;
table famcode*sampstr / list missing;
run;

title5 'Check the Actual Stratum Sizes';
```

```
proc freq data=out.samplc01;
table stratum / list missing;
run;

title5 'Region, Age, and Enrollee Group for the Frame';
proc freq data=in.framec;
table regsmpl age_n enlsmpl sexsmpl / list missing;
run;

title5 'Region, Age, and Enrollee Group for the Sample';
proc freq data=out.samplc01;
table regsmpl age_n enlsmpl sexsmpl / list missing;
run;

title5 'Cross of Region and Age for the Frame and Sample';
proc freq data=in.framec noprint;
table regsmpl*age_n / list missing out=frmpct;
run;

data frmpct (keep = regsmpl age_n frame);
set frmpct (rename = (percent = frame));
run;

proc freq data=out.samplc01 noprint;
table regsmpl*age_n / list missing out=smplpct;
run;

data smplpct (keep = regsmpl age_n sample);
set smplpct (rename = (percent = sample));
run;

proc sort data=frmpct;
by regsmpl age_n;
run;

proc sort data=smplpct;
by regsmpl age_n;
run;

data percents;
merge frmpct smplpct;
by regsmpl age_n;
run;

proc print data=percents;
var regsmpl age_n frame sample;
sum frame sample;
run;

title5 'Sample Counts';
proc freq data=out.samplc01;
table sampstr / list missing;
run;

title5 'Sampling Weight by Stratum';
proc freq data=out.samplc01;
table sampstr*BWT / list missing;
run;
```

```
title5 'Weighted Sample Counts';
proc freq data=out.samplc01;
table sampstr / list missing;
weight BWT;
run;

title5 'Frame Counts';
proc freq data=in.framec;
table sampstr / list missing;
run;

****************************************************************************
* Create the ENBGSMPL variable.
* Note: This var was created in samplc03.sas in q3 2000.
****************************************************************************;

DATA out.samplc01;
   SET out.samplc01;

select (patcat);
   when ('ACTDTY') ENBGSMPL='01';
   when ('DEPACT')
      do;
         select (pcm);
            when ('CIV') ENBGSMPL='02';
            when ('MTF') ENBGSMPL='03';
            when (' ')   ENBGSMPL='04';
            otherwise ENBGSMPL='c';
         end;
      end;
   when('NADD<65')
      do;
         select (pcm);
            when ('CIV') ENBGSMPL='05';
            when ('MTF') ENBGSMPL='06';
            when (' ')   ENBGSMPL='07';
            otherwise ENBGSMPL='d';
         end;
      end;
   when('NADD65+') ENBGSMPL='10';
   when('UNKNOWN')
      do;
         if pntypcd='S' then
            do;
              if pnlcatcd in ('A','J','N','V') then ENBGSMPL='01';
              else if dageqy = ' ' then ENBGSMPL='f';
              else if dageqy <= '064' then
                 do;
                    select (pcm);
                       when ('CIV') ENBGSMPL='05';
                       when ('MTF') ENBGSMPL='06';
                       when (' ')   ENBGSMPL='07';
                       otherwise    ENBGSMPL='g';
                    end;
                 end;
              else if dageqy > '064' then ENBGSMPL='10';
            end;
         else if pntypcd='D' then
            do;
```

```
               if pnlcatcd in ('A','J','N','V') then
                  do;
                     select (pcm);
                        when ('CIV') ENBGSMPL='02';
                        when ('MTF') ENBGSMPL='03';
                        when (' ')   ENBGSMPL='04';
                        otherwise    ENBGSMPL='h';
                     end;
                  end;
               else if dageqy = ' ' then ENBGSMPL='i';
               else if dageqy <= '064' then
                     do;
                        select (pcm);
                           when ('CIV') ENBGSMPL='05';
                           when ('MTF') ENBGSMPL='06';
                           when (' ')   ENBGSMPL='07';
                           otherwise    ENBGSMPL='j';
                        end;
                     end;
               else if dageqy > '064' then ENBGSMPL='10';
            end;
         else ENBGSMPL='e';
      end;
   otherwise ENBGSMPL='b';
end;
DROP PATCAT PNTYPCD PNLCATCD;
LABEL ENBGSMPL = 'ENBGSMPL-Enrollment Beneficiary group';
RUN;

PROC FREQ;
   TABLES ENBGSMPL dageqy ENBGSMPL*dageqy /MISSING LIST;
RUN;


***************************************;
*CREATE IN_HOUSE AND CLIENT DATASETS***;
***************************************;

data out.samplc;
set out.samplc01;
run;

proc contents;
run;

data out.samplc01 (keep= AGESMPL AGE_N BWT DAGEQY DHSRGN ENLSMPL FAMCODE
                         MPRID PCM REGSMPL SAMPSTR SEL_PROB SEXSMPL STRATUM
SUPREG ENBGSMPL);
set out.samplc01;
run;

proc contents;
run;
```

```
****************************************************************************
*
* PROGRAM: SAMPLC02.SAS
* TASK:    DOD Health Care Survey, Sampling (8860-220)
* PURPOSE: Attach DEERS variables to FORM C Sample, Step 2
*
* WRITTEN: 05/07/2001 BY KEITH RATHBUN
*
* INPUTS:
* 1) SAMPLC01.SD2 - 2003 Q3 DOD FORM C Sample
* 2) XWALKC.SD2 - 2003 Q3 DEERS FORM C XWALK File
* 3) DOD 2003 Q3 DEERS Extract File
*    a) STI001.SD2 - 2003 Q3 DEERS Population Extract File (Part 1)
*    b) STI002.SD2 - 2003 Q3 DEERS Population Extract File (Part 2)
*    c) STI003.SD2 - 2003 Q3 DEERS Population Extract File (Part 3)
*    d) STI004.SD2 - 2003 Q3 DEERS Population Extract File (Part 4)
*
* OUTPUTS:
* 1) SAMPLC02.SD2 - 2003 Q3 DOD FORM C Sample combined with DEERS extract
*
* INCLUDES: None
*
****************************************************************************;
LIBNAME  OUT V612 "D:\KEITH\DOD\Q3_2003\DATA";
LIBNAME  IN  "D:\KEITH\DOD\Q3_2003\DATA";
OPTIONS LS=132 PS=79 NOCENTER COMPRESS=YES;

****************************************************************************
* Attach SSNSMPL variable and keep only the sampled records.
****************************************************************************;
PROC SORT DATA=OUT.XWALKC   OUT=XWALKC;   BY MPRID; RUN;
PROC SORT DATA=IN.SAMPLC01 OUT=SAMPLC01; BY MPRID; RUN;

DATA SAMPLC02;
   MERGE XWALKC(IN=IN1) SAMPLC01(IN=IN2);
   BY MPRID;
   IF IN1 AND IN2;
RUN;

PROC SORT DATA=SAMPLC02; BY SSNSMPL; RUN;
LIBNAME  IN V612 "D:\KEITH\DOD\Q3_2003\DATA"; *Switch back to V612;

%MACRO PROCESS(DSN=);
****************************************************************************
* COMBINE each part (1-4) of the address/extract information file with
* sample file information.  DROP sampling variables (already on the file).
****************************************************************************;
PROC SORT DATA=OUT.&DSN
    (DROP=PNARSNCD MDCABRSN MDCAEFDT  MDCAEXDT)
    OUT=TEMP;
    BY SSNSMPL;
RUN;

****************************************************************************
* MERGE the DEERS extract file information with the Form C Sample by SSNSMPL.
****************************************************************************;
DATA &DSN;
   MERGE TEMP(IN=IN1) SAMPLC02(IN=IN2);
   BY SSNSMPL;
```

```
   IF IN1 AND IN2;
RUN;

****************************************************************************
* DELETE temporary dataset to conserve disk space.
****************************************************************************;
PROC DATASETS; DELETE TEMP; RUN;

%MEND PROCESS;

%PROCESS(DSN=STI001);
%PROCESS(DSN=STI002);
%PROCESS(DSN=STI003);
%PROCESS(DSN=STI004);

****************************************************************************
* STACK the combined DEERS extract/sample file information into one dataset.
****************************************************************************;
DATA SAMPLC02;
   SET STI001 STI002 STI003 STI004;
   BY SSNSMPL;
   DROP AGE_N BWT FAMCODE SAMPSTR SEL_PROB SEXSMPL;
   *************************************************************************
   * STI sent duplicates SSNSMPLs.  So, we let SAS remove them here.
   *************************************************************************;
   IF FIRST.SSNSMPL;
RUN;

****************************************************************************
* SORT the combined DEERS extract/sample file information by MPRID.
****************************************************************************;
PROC SORT DATA=SAMPLC02 OUT=OUT.SAMPLC02; BY MPRID; RUN;

TITLE1 "2003 Q3 DOD Health Care Survey Sampling (8860-220)";
TITLE2 "PROGRAM: SAMPLC02.SAS, OUTPUT: SAMPLC02.SD2";
TITLE3 "WRITTEN BY: KEITH RATHBUN, May 2003";

PROC CONTENTS; RUN;

PROC FREQ DATA=IN.SAMPLC02
    (DROP=
       MALN1TX
       MALN2TX
       MACITYNM
       MAPRZIP
       MAPRZIPX
       MPRID
       PN1STNM
       PNBRTHDT
       PNID
       PNLSTNM
       PNMIDNM
       PRN
       SPCITYNM
       SPLN1TX
       SPLN2TX
       SPONSSN
       SPPRZIP
       SPPRZIPX
```

```
       SPTNUMCD
       SSNSMPL
       TNUMCD
       UICADD1
       UICADD2
       UICCITY
       UICZIP
      );
   TABLES _ALL_ /MISSING LIST;
RUN;
```

```
***************************************************************************
***
*** Project:        2003 Health Care Survey of DoD Beneficiaries - Child
*** Project Number: 8860
*** Task Number:    220
***
*** Purpose: Sample size determination for the 2003 DoD Child sample design
***
***
*** Date:       April  2003
*** Programmer: Nancy A. Clusen Initial program by Keith Rathbun.
***
*** Program: D:\projects\8860-210\child_2003\samsizec.sas,
***          Determine sample sizes for all child samples
***
*** Inputs:  D:\projects\8860-210\child_2003\Data\countc.sd2
***          Population counts by various cell definitions.
***
*** Outputs: D:\projects\8860-210\child_2003\Data\csize.sd2
***          Sample sizes by various cell definitions.
***
*** Notes: None
***
***************************************************************************;
*** Setup the titles. ***;
title1 '2003 Health Care Survey of DoD Beneficiaries - Child';
title2 'Program:  D:\projects\8860-210\child_2003\samsizec.sas by Nancy A.
Clusen';
title3 'Determine sample sizes for all child samples.';

LIBNAME IN   'D:\projects\8860-210\child_2003\Data';
OPTIONS PS=79 LS=132 ERRORS=2 NOCENTER mlogic mprint symbolgen;

%LET P = .5;            ***PRODUCE THE MOST CONSERVATIVE SAMPLE SIZES****;
%LET Z = 1.96;          ***97.5TH PERCENTILE FOR Z-DIST******************;
%LET HLA0 = .040501;        ***HALF LENGTH FOR EACH STRATUM*****;
%LET SSQUARE = &P*(1-&P); ***FORMULA FOR VARIANCE OF P****************;

/*------------------------------------------------------------------------
        MACRO:  CALCULATE NUMERICAL PORTIONS OF VARIANCES GIVEN SAMPLE SIZES
-------------------------------------------------------------------------*/
%MACRO VAR(DAT,DOMAIN,POPSIZE,NH,ODAT);
DATA VARA;
        SET &DAT;BY &DOMAIN;
        VH=&POPSIZE**2*((&POPSIZE-&NH)/(&POPSIZE-1))*DE*&SSQUARE/&NH;
RUN;

PROC MEANS DATA=VARA NOPRINT;
        VAR VH;BY &DOMAIN;
        OUTPUT OUT=&ODAT SUM=VSUM;
RUN;
%MEND VAR;

************************************************************************
*       TO DETERMINE OPTIMAL STRATUM SIZES GIVEN PREDETERMINED VARIANCE
***********************************************************************;
%MACRO OPTALLO(DAT,DOMAIN,POPSIZE,V0,ODAT);
/*----------------------------------------------------------
        TO CALCULATE PARTIAL SUMS OF REMAINING DOMAIN SIZES
```

```
        NOTE:  THIS SUM can be DIFFERENT FROM THE DOMAIN TOTAL !!!
--------------------------------------------------------------*/
DATA &DAT;SET &DAT;
        DEN = (&POPSIZE/DSUM&ITE)**2*DE/(&POPSIZE-1);
        COM = &POPSIZE*SQRT(DE*&POPSIZE/(&POPSIZE-1));
        NUM = COM/DSUM&ITE;
RUN;
PROC MEANS DATA=&DAT NOPRINT;
        VAR NUM DEN COM;BY &DOMAIN;
        OUTPUT OUT=DSIZEA SUM=NUMS DENS COMS;
RUN;

DATA &ODAT;
        MERGE &DAT DSIZEA;BY &DOMAIN;
        ND=(&SSQUARE*NUMS**2)/(&V0+&SSQUARE*DENS);
        NHO=ND*COM/COMS;
        DROP ND NUM DEN COM NUMS DENS COMS;
RUN;
%MEND OPTALLO;
/*----------------------------------------------------------------
        TO RETRIVE THE NUMBER OF OBSERVATIONS IN A SAS DATA SET
------------------------------------------------------------------*/
%MACRO NUMOBS(DSN);
        %GLOBAL NUM;/* THIS MACRO CONTAINS THE NUMBER OF OBS IN THE DATA*/
        DATA _NULL_;
                IF 0 THEN SET &DSN NOBS=COUNT;
                CALL SYMPUT('NUM',LEFT(PUT(COUNT,8.)));
                STOP;
        RUN;
%MEND NUMOBS;

/*----------------------------------------------------------------
        ITERATE UNTIL THE REMAINING DOMAINS HAVE NHO GREATER THAN
        THE PREVIOUS SAMPLE SIZES
-----------------------------------------------------------------*/
%MACRO ITERATE;
%OPTALLO(STE,DOM&ITE,POPSIZE,VSTAR,OSTAT);

DATA FIN&I STE;
        SET OSTAT;
        IF NHF < NHO THEN FIN = FIN +1;
IF FIN=&I then output FIN&I;
IF FIN = &I + 1 then output STE;
RUN;

%VAR(FIN&I,DOM&ITE,POPSIZE,NHF,SUMMARY);

DATA STE;
        MERGE STE (IN=A) SUMMARY ;BY DOM&ITE;
        IF A;
        IF VSUM=. THEN VSUM=0;****SHULD EXIST!!!;
        VSTAR= VSTAR - VSUM/DSUM&ITE**2;
        DROP VSUM;
RUN;
%MEND ITERATE;

/*--------------------------------------------------------------------
        MAIN PART OF THE PROGRAM:  'ITE' INDICATES THE LEVEL OF DOMAINS
----------------------------------------------------------------------*/
```

```
%MACRO MPART(ITE);
PROC SORT data=indata;BY DOM&ITE;RUN;

%VAR(INDATA,DOM&ITE,POPSIZE,NHF,SUMMARY);

DATA CHKVAR;***TO COMPARE THE VARIANCE TO THE PRECISION REQUIREMENT;
        MERGE SUMMARY INDATA;BY DOM&ITE;
        FIN=1;
        MARGIN=SQRT((VSUM/DSUM&ITE**2)*1.96**2)/HL&ITE;
        IF MARGIN > 1 THEN FIN=FIN+1;
        DROP VSUM MARGIN; /* SHOULD DROP 'VSUM'VARIABLE HERE !!! */
RUN;

***DATA SET INCLUDING STRATA HAVING FINAL SAMPLE SIZE AT THIS STEP***;

DATA FIN1 STE;
        SET CHKVAR;BY DOM&ITE;
        VSTAR=(HL&ITE/1.96)**2;
IF FIN=1 then output FIN1;
IF FIN=2 then output STE;
RUN;

%NUMOBS(STE);

%LET I = 1;
%IF &NUM=0 %THEN %GOTO FDSN;
/*------------------------------------------------------------------------------
-
        ITERATE MACRO TO UPDATE SAMPLE SIZES TO MEET THE PRECISION
REQUIREMENTS
        THIS PART NEEDS TO BE REFINED TO ALLOW TO STOP THE PROGRAM WHENEVER
NEEDED
--------------------------------------------------------------------------------
--*/
%DO %UNTIL(&NUM = 0);
        %LET I = %EVAL(&I +1);
        %ITERATE;
        %NUMOBS(FIN&I);
%END;
/*----------------------------------------------------------------------------
        GIVE THE REMAINING DOMAINS OPTIMAL SAMPLE SIZES
-----------------------------------------------------------------------------*/
%LET I = %EVAL(&I +1);
DATA FIN&I;SET STE;
        NHF = NHO;
RUN;
/*-----------------------------------------------------------------------------
        COMBINE THE DATASETS INTO ONE
-------------------------------------------------------------------------------
*/
%FDSN:
DATA STEP9;
        SET FIN1;

%DO J=2 %TO &I;
        DATA STEP9;
                SET STEP9 FIN&J;
        RUN;
%END;
```

```
%MEND MPART;

***************************************************************************
*       START THE MAIN PROGRAM:
---------------------------------------------------------------------------;
DATA INDATA;
        SET IN.countc;
        DOM0 = sampstr/*STRATUM*/;
        dom1 = SUPREG;
        DOM2 = enlsmpl;
        DOM3 = agesmpl;
        DOM4 = 1;
        popsize = psum0;
        dsum1 = psum1;
        dsum2 = psum2;
        dsum3 = psum3;
        dsum4 = total;
        de = 1;
**************************************************
*       SET INITIAL SAMPLE SIZES
**************************************************;
        NUM=&Z**2*DE*&SSQUARE/&HLA0**2;
        NHZERO=NUM/(1+(NUM-1)/POPSIZE);
        NHF = NHZERO;
**************************************************
*       PRECISION REQUIREMENTS
**----------------------**************************;
        HL1 = 0.02;     ***FOR SUPER REGIONS******;
        HL4 = 0.01;     ***FOR THE WHOLE*********;
        DROP NUM COUNT;
RUN;
*--------------------------------------------------------------------------***
*       ADJUST INITIAL SAMPLE SIZE TO SATISFY THE DOM&ITE PRECISION
REQUIREMENT
----------------------------------------------------------------------------*;
%MPART(1);
**--------------------------------------------------------------------------
*       CTEATE STATUS&ITE SO THAT FIN VALUES CAN REFLECT ITE TOO
----------------------------------------------------------------------------*;
DATA INDATA;SET STEP9;
        STATUS1=10+FIN;
DROP FIN;
RUN;

************************************************************
*       ACCOUNT FOR OVERALL PRECISION REQUIREMENT
***********************************************************;
%mpart(4)
DATA FINAL;SET STEP9;
        STATUS4=40+FIN;
        NHF4=NHF;
        VH=POPSIZE**2*((POPSIZE-NHF)/(POPSIZE-1))*DE*&SSQUARE/NHF;
RUN;
*---------------------------------------------------------------------------
        CHECK IF THE FINAL SAMPLE SIZES MEET ALL PRECISION REQUIREMENTS
----------------------------------------------------------------------------;
PROC SORT DATA=FINAL;BY DOM1;RUN;
PROC MEANS NOPRINT DATA=FINAL;VAR VH;BY DOM1;
        OUTPUT OUT=FDATA1 SUM=V1;
```

```
RUN;
DATA FINAL;MERGE FINAL FDATA1;BY DOM1;

PROC SORT DATA=FINAL;BY DOM2;RUN;
PROC MEANS DATA=FINAL NOPRINT;VAR VH;BY DOM2;
        OUTPUT OUT=FDATA2 SUM=V2;
RUN;
DATA FINAL;MERGE FINAL FDATA2;BY DOM2;

PROC SORT data=final;BY DOM3;RUN;
PROC MEANS DATA=FINAL NOPRINT;VAR VH;BY DOM3;
        OUTPUT OUT=FDATA3 SUM=V3;
RUN;
DATA FINAL;MERGE FINAL FDATA3;BY DOM3;
PROC MEANS DATA=FINAL NOPRINT;VAR VH;
        OUTPUT OUT=FDATA4 SUM=V4;
RUN;
DATA FINAL;IF _N_ = 1 THEN SET FDATA4;
        SET FINAL;
        P0=SQRT(((POPSIZE-NHF)/(POPSIZE-1))*DE*&SSQUARE/NHF)*1.96;
        P1=SQRT((V1/DSUM1**2)*1.96**2);
        P2=SQRT((V2/DSUM2**2)*1.96**2);
        P3=SQRT((V3/DSUM3**2)*1.96**2);
        P4=SQRT((V4/DSUM4**2)*1.96**2);
RUN;
*******************************************************
*       ACCOUNT FOR EXPECTED RESPONSE RATES
*******************************************************;
DATA RESP;
        SET FINAL;
        NHFF=INT(NHF/0.31)+1;
RUN;

DATA LAST;SET RESP;
        nhf = int(nhf)+1;
        nhzero = int(nhzero)+1;
        BWT = POPSIZE/NHFF;
PROC SORT data=LAST;BY DOM0;run;
PROC MEANS DATA=LAST;VAR NHZERO nhf NHFF BWT;RUN;
PROC PRINT DATA=LAST;VAR DOM0 DOM1 P0 P1 DOM2 P2 DOM3 P3 p4 POPSIZE NHFF;sum
nhff; RUN;
*************************************************************
*       CREATE THE DATA SET CONTAINING THE FINAL SAMPLE SIZES
*************************************************************;
DATA IN.CSIZE;
        SET LAST;
        KEEP sampstr/*STRATUM*/ POPSIZE NHFF;
```

4. Include Files

```
***************************************************************************
*
* PROGRAM:  LAYOUT.SAS
* TASK:     DOD Health Care Survey, Sampling (8860-210/220)
* PURPOSE:  INPUT step for the 2000 DEERS Extract file from STI
*
* WRITTEN:  10/18/2000 BY KEITH RATHBUN
*
* MODIFIED: 1) 04/22/2002 BY KEITH RATHBUN, Removed TSPSITE from layout.
*
***************************************************************************
***************************************************************************
* Input RAW data (ignore delimiters!)
***************************************************************************;
INPUT
   @1      SPONSSN    $CHAR9.
   @11     SPDUPID    $CHAR1.
   @13     PNTYPCD    $CHAR1.
   @15     PNID       $CHAR9.
   @25     PNBRTHDT   $CHAR8.
   @34     MRTLSTAT   $CHAR1.
   @36     PNSEXCD    $CHAR1.
   @38     PNARSNCD   $CHAR2.
   @41     MDCABRSN   $CHAR1.
   @43     MDCAEFDT   $CHAR8.
   @52     MDCAEXDT   $CHAR8.
   @61     LEGDDSCD   $CHAR2.
   @64     PNLCATCD   $CHAR1.
   @66     SVCCD      $CHAR1.
   @68     PAYPLNCD   $CHAR5.
   @74     PGCD       $CHAR2.
   @77     MBRRELCD   $CHAR1.
   @79     MALN1TX    $CHAR40.
   @120    MALN2TX    $CHAR40.
   @161    MACITYNM   $CHAR20.
   @182    MASTCD     $CHAR2.
   @185    MACTRYCD   $CHAR2.
   @188    MAPRZIP    $CHAR5.
   @194    MAPRZIPX   $CHAR4.
   @199    HADDFLG    $CHAR1.
   @201    TNUMCD     $CHAR14.
   @216    PNLSTNM    $CHAR26.
   @243    PN1STNM    $CHAR20.
   @264    PNMIDNM    $CHAR20.
   @285    PNCDNCY    $CHAR4.
   @290    RANKCD     $CHAR6.
   @297    ULOCGRN    $CHAR2.
   @300    ULOCDMIS   $CHAR4.
   @305    RACEETHN   $CHAR1.
   @307    DCATCH     $CHAR4.
   @312    DMEDELG    $CHAR1.
   @314    DAGEQY     $CHAR3.
   @318    DBENCAT    $CHAR3.
   @322    DPRISM     $CHAR4.
   @327    DHSRGN     $CHAR2.
   @330    DSPONSVC   $CHAR1.
   @332    MEDTYPE    $CHAR1.
```

```
   @334    UICADD1    $CHAR30.
   @365    UICADD2    $CHAR30.
   @396    UICCITY    $CHAR30.
   @427    UICST      $CHAR2.
   @430    UICZIP     $CHAR5.
   @436    UADDFLG    $CHAR1.
   @438    SPLN1TX    $CHAR40.
   @479    SPLN2TX    $CHAR40.
   @520    SPCITYNM   $CHAR20.
   @541    SPSTCD     $CHAR2.
   @544    SPCTRYCD   $CHAR2.
   @547    SPPRZIP    $CHAR5.
   @553    SPPRZIPX   $CHAR4.
   @558    SADDFLG    $CHAR1.
   @560    SPTNUMCD   $CHAR14.
   @575    ENRID      $CHAR4.
   @580    ACV        $CHAR1.
   @582    PCM        $CHAR3.
   @586    PATCAT     $CHAR7.
 ;
************************************************************************
* Construct SSNSMPL as SPONSSN & SPDUPID & LEGDDSCD
************************************************************************;
LENGTH SSNSMPL $12;
SSNSMPL = SPONSSN || SPDUPID || LEGDDSCD ;

************************************************************************
* LABEL variables
************************************************************************;
LABEL
    SSNSMPL  =   "SSNSMPL - SPONSSN & SPDUPID & LEGDDSCD"
    SPONSSN  =   "Sponsor SSN"
    SPDUPID  =   "Family Sequence Number"
    PNTYPCD  =   "Person Type Code"
    PNID     =   "Person SSN"
    PNBRTHDT =   "Person Birth Date"
    MRTLSTAT =   "Marital Status"
    PNSEXCD  =   "Person Gender"
    PNARSNCD =   "Person Association Reason Code"
    MDCABRSN =   "Medicare A Begin Reason Code"
    MDCAEFDT =   "Medicare A Effective Date"
    MDCAEXDT =   "Medicare A Expiration Date"
    LEGDDSCD =   "DDS Code"
    PNLCATCD =   "Personnel Category Code (Duty Status)"
    SVCCD    =   "Branch of Service"
    PAYPLNCD =   "Pay Plan Code"
    PGCD     =   "Pay Grade"
    MBRRELCD =   "Member Relationship Code"
    MALN1TX  =   "Residential Address, Line 1"
    MALN2TX  =   "Residential Address, Line 2"
    MACITYNM =   "Residential Address, City"
    MASTCD   =   "Residential Address, State"
    MACTRYCD =   "Residential Address, Country"
    MAPRZIP  =   "Residential Address, ZIP Code"
    MAPRZIPX =   "Residential Address, ZIP Code Extension"
    HADDFLG  =   "Residential Address Flag"
    TNUMCD   =   "Residence Telephone Number"
    PNLSTNM  =   "Person Last Name"
    PN1STNM  =   "Person First Name"
```

```
   PNMIDNM  =   "Person Middle Name"
   PNCDNCY  =   "Person Generation (Cadency)"
   RANKCD   =   "Rank Code"
   ULOCGRN  =   "Unit Region"
   ULOCDMIS =   "Unit DMISID"
   RACEETHN =   "Race/Ethnic Code"
   DCATCH   =   "Catchment Area"
   DMEDELG  =   "Medical Privlege Code"
   DAGEQY   =   "Age (As of 28 February 2003)"
   DBENCAT  =   "Beneficiary Category"
   DPRISM   =   "PRISM (20 mile) clinic service area"
   DHSRGN   =   "Health Service Region"
   DSPONSVC =   "Derived Sponsor Branch of Service"
   MEDTYPE  =   "Medicare Type"
   UICADD1  =   "Unit Address, Line 1"
   UICADD2  =   "Unit Address, Line 2"
   UICCITY  =   "Unit Address, City"
   UICST    =   "Unit Address, State"
   UICZIP   =   "Unit Address, ZIP Code"
   UADDFLG  =   "Unit Address Flag"
   SPLN1TX  =   "Sponsor Address, Line 1"
   SPLN2TX  =   "Sponsor Address, Line 2"
   SPCITYNM =   "Sponsor Address, City"
   SPSTCD   =   "Sponsor Address, State"
   SPCTRYCD =   "Sponsor Address, Country"
   SPPRZIP  =   "Sponsor Address, ZIP Code"
   SPPRZIPX =   "Sponsor Address, ZIP Code Extension"
   SADDFLG  =   "Sponsor Address Flag"
   SPTNUMCD =   "Sponsor Telephone Number"
   ENRID    =   "Enrollment DMISID"
   ACV      =   "Alternate Care Value"
   PCM      =   "Primary Manager Code (CIV or MIL)"
   PATCAT   =   "Aggregated Beneficiary Category"
;
```

APPENDIX E

TECHNICAL BACKGROUND IN DETERMINING THE SAMPLE SIZES

Technical Background for the Algorithm

To attain the required half-length *HL* for confidence intervals, the required sample size *n* was obtained while incorporating finite population correction factors.

For simple random samples (SRS) of size *n* from finite populations of size *N*, the variance of *p* is:

(E.1)
$$V_{SRS}(p) = \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$$

Because the expected sample sizes for all strata for the 2003 HCSDB Child survey are sufficiently large, the standard formula (4.1) in Chapter IV can be used in constructing the confidence interval of *P*. Let *B* denote the required half-length interval for *P*. With the variance of *P*, we can determine the sample size to attain the precision requirement *B* by solving the following equation with respect to *n*:

(E.2)
$$B = z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)}$$

implies

(E.3)
$$n = \frac{\dfrac{z^2_{1-\alpha/2}\,[P(1-P)]}{B^2}}{1+\dfrac{1}{N}\left(\dfrac{z^2_{1-\alpha/2}\,[P(1-P)]}{B^2}\right)}$$

This formula was used as the first step in determining initial sample sizes for all strata in the 2003 HCSDB.

Note from formula (E.3), sample sizes vary according to values of the proportion *P*. As *P* becomes closer to 0.5, *n* becomes larger. Because characteristics of interest of this survey could have values ranging from zero to one, the resulting sample sizes lie within a wide range of values with the largest value associated with *P*=0.5. For sample size determination, we used a *P* value of 0.5, which ensures that the sample size will be large enough to meet or exceed the predetermined precision requirement for all proportions to be estimated.

Since the sample size is being defined to construct a 95 percent interval for P = 0.5 with a half-length interval less than or equal to B, $z_{1-\alpha/2}$ can be replaced with $z_{.975}$ which is 1.96. Formula (E.3) can then be specified as the following:

(E.4)
$$n = \frac{\dfrac{.9604}{B^2}}{1 + \dfrac{1}{N}\left(\dfrac{.9604}{B^2}\right)}$$

where .9604 was obtained from $z_{.975}\, P(1-P)$ with P = 0.5. The formula (E.4) can then be applied to determine the sample size to achieve B in estimating stratum-level estimates.

Recall that the 2003 HCSDB employs a stratified sample design. Since we wish to estimate the proportion of beneficiaries from domain *d* having a certain characteristic, an estimate of the proportion P_d can be obtained as the weighted sum of stratum-level proportion estimates:

(E.5)
$$p_d = \sum_{h \in d} \sum_{i \in d} \frac{N_d(h)}{N_d(+)} \; p_h ,$$

where N_h is the population size for stratum *h*, $N_{\overset{+}{d}}$ is the sum of N_h over domain *d*, and p_h is the estimated proportion for the *h*-th stratum. Since the sampling is independent across strata, the variance of estimated proportion p_d is the sum of stratum-level variances:

(E.6)
$$V_d = \sum_{h \in d} \left(\frac{N_h}{N_d}\right)^2 \left(\frac{N_h - n_h}{N_h - 1}\right) \frac{P_h(1-P_h)}{n_h}$$

where n_h is the sample size in stratum *h* and P_h is the stratum-level proportion for stratum *h*. Like the single stratum case, all stratum-level proportions are assumed to be 0.5, and thus the formula (E.6) can be reduced to the following:

(E.7)
$$V_d = \sum_{h \in d} \left(\frac{N_h}{N_d}\right)^2 \left(\frac{N_h - n_h}{N_h - 1}\right) \frac{.25}{n_h}$$

The minimum sample size satisfying the requirements for a predetermined half-length interval B_d is:

(E.8)
$$n_d = \frac{\left(\sum_{h \in d} \dfrac{N_h}{N_d} \sqrt{\dfrac{N_h}{N_h - 1}} \sqrt{P_h(1-P_h)}\right)^2}{\dfrac{B_d^2}{z_{1-\alpha/2}^2} + \sum_{h \in d} \dfrac{N_h^2}{N_d^2} \left(\dfrac{1}{N_h - 1}\right) P_h(1-P_h)}$$

With the same specifications above, formula (E.8) can be specified as:

(E.9)

$$n = \frac{.25\left(\sum_{h\in d}\frac{N_h}{N_d}\sqrt{\frac{N_h}{N_h - 1}}\right)^2}{\frac{B_d^2}{3.8416} + .25\sum_{h\in d}\frac{N_h^2}{N_d^2}\frac{1}{N_h - 1}},$$

where $P_h(1 - P_h)$ = (.5)(.5) = 0.25 for all *h* and $z_{.975}^2$ = 3.8416.

The domain sample size n_d in (E.9) is based on the following optimal stratum sample sizes:

(E.10)

$$n_h = n_d \frac{N_h\sqrt{\frac{N_h}{N_h - 1}}\sqrt{P_h(1 - P_h)}}{\sum_{h\in d} N_h\sqrt{\frac{N_h}{N_h - 1}}\sqrt{P_h(1 - P_h)}}$$

Likewise, this formula becomes

(E.11)

$$n_h = n_d \frac{N_h\sqrt{\frac{N_h}{N_h - 1}}}{\sum_{h\in d} N_h\sqrt{\frac{N_h}{N_h - 1}}}$$

After the stratum size for eligible respondents was finally determined, an anticipated response rate *R* was incorporated to get the final stratum sample size:

(E.12)

$$n_{h,F} = \frac{n_h}{R}$$