COMBINED ESTIMATES FROM FOUR QUARTERLY SURVEY DATA SETS

Esther M. Friedman, Don Jang, Mathematica Policy Research
Thomas V. Williams, Department of Defense
Esther M. Friedman, 600 Maryland Ave SW Washington, DC 20024 (efriedman@mathematica-mpr.com)

Key Words: Combined estimate; quarterly survey; temporal aggregation; cross-sectional aggregation; rolling sample; Health Care Survey of Department of Defense Beneficiaries (HCSDB)

I. INTRODUCTION

The concept of cumulating surveys over time is hardly new. One of its early advocates was Leslie Kish, who first proposed the concept as early as 1965 (Kish, 1965, p.475), and discussed it in more detail in several papers in the 1970's and, more recently, in the 1990's (see, for example, Kish, 1979; Kish, 1990; Kish 1999). Kish advocates using a "rolling sample" design with non-overlapping monthly panels, cumulated over different lengths of time for domains of different sizes. He feels that researchers focus almost exclusively on variations in spatial domains, such as counties and provinces, while ignoring variations in temporal dimensions (Kish, 1998). He argued that cumulating data from periodic surveys into "rolling samples" addresses the need for both spatial and temporal detail.

The recent proposal by the U.S. Bureau of the Census to replace the long form of the Census with the American Community Survey (ACS) has spurred a renewed interest in "rolling samples." The ACS, a monthly survey of different geographic areas, is rolled into an annual sample and is often cumulated over several years, as well, depending on the size of the analytic domain (Alexander, 1998; Alexander and Wetrogan, 2000). This design ensures that a full survey dataset is available every year and that the individual monthly domains are more current than they would be if all areas were collected in one time period. As with the ACS, the quarterly Adult Health Care Survey of Department of Defense Beneficiaries (HCSDB) is cumulated into an annual dataset, to produce frequent and accurate data without sacrificing the sample sizes necessary to make reliable small area estimates.

As rolling samples become more prevalent, a major problem that arises is determining which weighting technique to employ in generating combined estimates. This paper briefly discusses when and why one might opt to survey respondents with more frequency, how to ensure that sample sizes are large enough for small area estimates, how the HCSDB quarterly samples were combined into an annual dataset, and how that technique compares to alternative methods for generating combined estimates.

II. WHY SPLIT A LARGE SURVEY INTO SMALLER REPEATED SAMPLES?

There is a need for frequent and accurate data for small areas and other small-sized domains, yet it is costly to obtain both the frequency desired and the sample size required to make small area estimations. Until recently, timeliness has been sacrificed for sample size, with the best example being that of the decennial census. The Census is rich in detail and allows for unbiased estimates of small domains, yet current information is made available only once every ten years and must, therefore, be supplemented with more recent information from other sources (e.g. the Current Population Survey).

Rolling samples is one way to obtain frequent and accurate data for small areas. Conducting monthly surveys, instead of one large, decennial survey, or conducting quarterly surveys in place of one annual survey, allows researchers to estimate trends with more frequency, and thereby detect seasonal trends or the effect of an irregular or sudden change. Furthermore, the average estimate obtained from repeated surveys is often more reliable than one point-in-time estimate (Kish, 1998). This is especially true when, as is often the case, the fielding period for the survey is simply the most convenient, if not necessarily the most representative, date.

III. METHODS FOR COMBINING SURVEYS

When fielding several small surveys, in place of one annual survey, one must combine the data in order to obtain accurate small area estimates. There are many different ways to combine data sets over time, and the optimal technique depends, in large part, on the data in question. Kish (1998, 1999) suggests a number of techniques for weighting data combined over a series of years. One approach is to give the full weight to the final year and assign a weight of zero to all other years. This approach lends itself to national and large domain estimates, where timeliness dominates sampling precision because the sample sizes are large enough. Equally weighted years, on the other hand, may be appropriate for estimates that are more stable over time. This approach is suggested for the ACS multi-year cumulations when there is little variation within the time period (Alexander, 2002). Sometimes one may want to weight the years differently, depending on their recency or some other criteria. For these cases, Kish recommends monotonically nondecreasing weights, and suggests determining the curve by way of a model, empirical data, or some combination of the two. Others employ moving average models to weight across years (Chand and Alexander, 2000). Certainly, for more complex samples that must be combined across both years and quarters, across different domains, or when additional auxiliary variables are available for the years, modeling-using either a linear or non-linear model-is probably the optimal technique (see, for example, Binder and Hidiroglou, 1989; Singh and Merkouris, 1995).

IV. THE ADULT HCSDB

The Adult Health Care Survey of Department of Defense Beneficiaries (HCSDB) is the primary tool with which the TRICARE Management Activity (TMA) of the Assistant Secretary of Defense (Health Affairs) monitors the opinions and experiences of military health system (MHS) beneficiaries. Specifically, the HCSDB is designed to monitor experience with, access to, and care provided by the MHS. The HCSBD has been conducted annually since 1995, and was first fielded quarterly in 2001.

Quarterly surveys were implemented primarily to obtain current information about MHS beneficiaries with more frequency. As funding makes it impossible to survey 180,000 respondents per quarter, the sample was divided by four for each quarter, with the intention that the data sets would be combined after four quarters of data are acquired. This larger dataset is necessary in order to ensure appropriate sample sizes for analysis of small geographic domains.

The samples for the 2001 HCSDB quarterly surveys were selected using a stratified sample design, stratified by 126 military *catchment* areas and six enrollment-beneficiary groups. In order to account for the survey design and to control for the biasing effects of nonresponse, weights were calculated for each quarter of data. The final weight includes a sampling weight that reflects the differential selection probabilities used to sample beneficiaries across strata and weighting class adjustments for non-response. The quarterly surveys were mailed to representative, independently selected samples of 45,000 MHS beneficiaries per quarter. Permanent random numbers were assigned to beneficiaries to ensure that no one is sampled more than once in a two-year period. Response rates range from 33.3% in quarter three to 36.1% in quarter two. After four quarters of data were acquired, it became necessary to combine the data sets into one large cumulative dataset, in order to provide reliable estimates of small domains.

V. METHODS FOR COMBINING THE ADULT HCSDB

The data from the adult HCSDB are primarily used to obtain basic descriptions and comparisons of geographic areas and domains and to measure changes in health care experiences over time, within specific geographic areas. In order to obtain the appropriate estimates from the combined data sets, it is necessary to determine the best technique for combining the quarterly estimates to weight the annual file. The appropriate weight for the quarterly data sets was determined by WCOM = q_i X Wq_i where i ranges from 1 to 4 for the four quarters and the sum of q1-q4 must be equal to one. The value of q*i* must be determined and varies depending on the weighting technique used.

The quarterly samples were combined into an annual dataset simply by assigning equal weights to each of the four quarterly surveys. This ensures that each sample contributes equally to the overall annual dataset. Several alternative techniques are discussed below and are compared to the equal weighting technique, including assigning weights proportional to the reference period, with the most recent reference period (quarter four) assigned the largest weights, and weighting proportional to each quarter's domain size for domain specific estimates.

Equal Weights

The technique employed in combining the HCSDB samples involved assigning equal weights to each quarter. This method assumes that the variation in the estimates from one quarter to the next is due merely to sampling variation. Thus, combined estimates can be calculated from the four independent samples by averaging the estimates for the four quarters. These combined estimates will, in fact, be more precise than the quarterly estimates because they average out the variation across the quarters.

Without loss of generality, let the current quarter be denoted as q4. Then, the combined dataset includes the four quarterly data sets: q1, q2, q3, q4. Let the final quarterly survey weights be denoted by Wq1, Wq2, Wq3, and Wq4. To retain the sum of the weights from the combined data as the population count, each quarterly survey weight will need to be rescaled to meet the following condition:

$WCOM_{EQUAL}$ = q_i X Wq_i where q_i is between 0 and 1 with the constraint: $q_1 + q_2 + q_3 + q_4 = 1$. Since equal relationships were assumed among the quarters, $q_1 = 0.25$, $q_2 = 0.25$, $q_3 = 0.25$, and $q_4 = 0.25$.

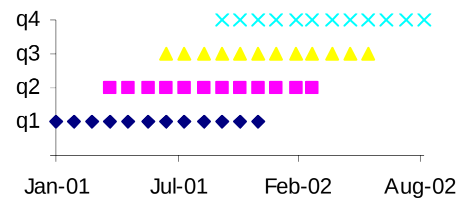
The abovementioned technique is an optimal choice for it is simple to implement and easy to interpret.

Weight Proportional to Reference Period

One problem, however, with assigning equal weights to the data, is that results obtained for the more recent quarters may be more relevant to data users or policy makers than those obtained from previous quarters. For example, if MHS beneficiaries of the last three quarters were satisfied, on the average, with a given service, but beneficiaries in the fourth (most recent) quarter showed a marked difference in level of satisfaction, it may be more useful to policy-makers if the overall estimate gives more weight to more recent respondents. Conversely, if first quarter respondents differ from respondents in quarters two through four, this may reveal a new trend in beneficiarys' attitudes and the information from the first quarter is not as relevant as the more recent data. Weidman et al. (1995) discuss the advantages and disadvantages of weighting based on reference period rather than using equal weights. Equal weighting gives minimum variance and ensures that the difference between successive yearly (or quarterly) estimate models is, simply, the difference between the two estimates. Increasing the proportion of the weight for more recent quarters, on the other-hand, gives averages that are more current, yet this benefit is complicated by the fact that a succeeding quarter's difference represents a linear combination of estimates, rather than the simple difference between the quarters, and is therefore more difficult to interpret. Interestingly, this is the reason the equal weighting technique is suggested by Alexander (2002) for the ACS's multi-year cumulations, rather than the reference period approach recommended by Kish for these data.

In the case of the adult HCSDB, there is an additional motivation for combining estimates proportional to reference period. Although the HCSDB surveys are now fielded quarterly, they have a reference period of a year, and, therefore, they question respondents about their experiences over the past year, so more recent surveys cover a greater percentage of the prior calendar year. For example, quarter one was fielded in January 2002, and asked respondents about their experiences and use of care over the past year, that is, from January 2001 to January 2002. The quarter two survey was fielded in April of 2002 and asked respondents about their experience and use of health care from April 2001 to April 2002. This nine-month overlap is apparent between each survey and its succeeding survey. Figure 1 below reveals this overlap.

Figure 1: HCSDB Reference Period by Quarter



In assigning weights relative to sampling reference period, q4 was determined to be the most recent reference, and q1 the least recent. The sum of the weights from the combined data must sum to the population count; therefore, each quarterly survey weight had to meet the following condition: $WCOM_{REF} = q_i \times Wq_i$, with q_i between 0 and 1 and with the constraints $q_1 + q_2 + q_3 + q_4 = 1$ and the additional constraint $q_1 < q_2 < q_3 < q_4$. Since the time lag between each quarter is identical, and each quarterly survey is deemed as relevant as the others, with the exception of the recency of the fielding period, the more recent reference periods are given a proportionally greater weight, and q_i is set to $i/10$. That is, $q_1 = 1/10$, $q_2 = 2/10$, $q_3 = 3/10$, $q_4 = 4/10$.

Weight Proportional To Each Quarter's Domain Size For Domain-Specific Estimates

Although results from the HCSDB are sometimes reported as overall estimates, they are often reported by one of several domains. Population sizes for these domains may be dynamic and may vary between quarters, especially for smaller sized domains. This is especially true of the active duty population, which is typically quite mobile. By assigning weights proportional to each quarter's domain size, one can account for potential variations in population sizes.

Three domains that are commonly used in analyses (and included in this study) are geographic region, respondent's enrollment type, and military *catchment* area of the health care facility. Geographic region refers to eleven geographic regions in the US, including the Northeast, Mid-Atlantic, Southeast, Gulfsouth, Heartland, Southwest, Two Central America regions, Southern California, Golden Gate, and the Northwest, as well as five regions outside of the US, such as Hawaii, Europe, Western Pacific, Latin America, and Alaska. Enrollment type includes active duty respondents, respondents enrolled in TRICARE Prime who are 65 and older; those enrolled who are under 65, non-enrollees 65+ and enrollees 65+. Military *catchment* areas refer to 126 small geographic areas surrounding military health care facilities or service areas both in the US and abroad.

Estimates were weighted using the following formula:

$$\mathrm{WCOM}_{i,d} = \frac{(\sum \mathrm{WQ}_{i,d})^{-1}}{(\sum \mathrm{WQ}_{1,d})^{-1} + (\sum \mathrm{WQ}_{2,d})^{-1} + (\sum \mathrm{WQ}_{3,d})^{-1} + (\sum \mathrm{WQ}_{4,d})^{-1}} \times \mathrm{WQ}_{i,d}$$

where i = quarters 1, 2, 3, 4, and *d* indicates the domain of interest.

VI. COMPARISON OF THE WEIGHTING TECHNIQUES

In order to compare the weighting techniques above to the equal weighting technique implemented for these data, the relative errors of the estimates were obtained for 23 key response variables in the HCSDB survey relating to use and rating of health services.

The point estimate, $\bar{Y}_t$, was computed for all weighting techniques, *t*, such that $\bar{Y}_t = \frac{\sum \mathrm{WCOM}_{ti} Y_i}{\sum \mathrm{WCOM}_{ti}}$.

The relative error of the estimate was then computed as: $RE\_P = \frac{\bar{Y}_t - \bar{Y}_{equal}}{\bar{Y}_{equal}}$, where $\bar{Y}_{equal}$ refers to the point estimate when the quarterly datasets were combined using the equal weighting technique.

VII. FINDINGS

Relative errors were obtained for all techniques for the 23 key variables. Table 1 below lists the number of estimates for which relative errors were obtained, by technique and domain.

Relative errors were, for the most part, negligible. However, when estimates were computed by military *catchment* area and weighted using the reference period technique, several estimates had large absolute relative errors (>20%). A complete list of all estimates with an absolute relative error greater than 20 percent appears in table 2 below, with their estimates for quarters one through four, the four weighting techniques, and absolute relative errors for each technique.

Table1: Number of Estimates By Technique and By Domain

Techniques	Domains			
	Overall	Region	Military area	Enrollee-beneficiary group
$RE_{Quarter}$	23 estimates	23x16 estimates	23x126 estimates	23x5 estimates
$RE_{Qrt+Domain}$	n/a	23x16	23x126	23x5
$RE_{Reference}$	23	23x16	23x126	23x5

Table 2: Estimates and Relative Errors for RE>20%

VARIABLE*	Military/Catchment Area	Est Q1	Est Q2	Est Q3	Est Q4	Est Equal	Est Qrt	Est Domain	Est Ref	RE Qrt	RE Domain	RE Ref
7+ civilian prescriptions filled by military pharmacy	Seoul	0.07	0.02	0.02	0.01	0.03	0.03	0.03	0.02	0.1%	2.4%	31.8%
	Out/Area-Europe	0.01	0.02	0.04	0.04	0.02	0.02	0.03	0.03	0.2%	24.2%	25.1%
	Wuerzburg	0.07	0.00	0.03	0.01	0.03	0.03	0.03	0.02	0.3%	0.3%	24.3%
	NMCL Annapolis	0.05	0.02	0.02	0.01	0.02	0.02	0.03	0.02	0.1%	1.2%	22.3%
	Out/Area-Latin America	0.03	0.01	0.03	0.08	0.04	0.04	0.04	0.04	0.2%	1.5%	21.4%
Beneficiary covered by civilian insurance	NH Corpus Christi	0.02	0.01	0.03	0.07	0.04	0.04	0.03	0.04	0.2%	2.5%	22.6%
	Patrick AFB	0.02	0.04	0.04	0.08	0.05	0.05	0.05	0.06	0.0%	1.0%	21.4%
Big problem getting necessary care	Scott AFB	0.15	0.06	0.03	0.03	0.06	0.06	0.06	0.05	0.0%	0.3%	28.9%
	Pearl Harbor	0.25	0.14	0.08	0.05	0.14	0.14	0.14	0.10	0.0%	0.5%	24.6%
	Ft. Eustis	0.03	0.01	0.05	0.08	0.04	0.04	0.04	0.05	0.2%	0.0%	24.2%
	NH Cherry Point	0.04	0.06	0.02	0.01	0.03	0.03	0.03	0.02	0.4%	1.2%	22.5%
	Pearl Harbor	0.10	0.04	0.02	0.03	0.05	0.05	0.05	0.04	0.0%	0.6%	22.0%
	Ft. Jackson	0.08	0.03	0.03	0.02	0.04	0.04	0.04	0.03	0.1%	0.7%	21.4%
	NH 29-Palms	0.16	0.13	0.06	0.05	0.10	0.10	0.10	0.08	0.1%	1.5%	21.2%
	Dover AFB	0.04	0.15	0.12	0.21	0.13	0.13	0.13	0.15	0.1%	0.0%	20.4%
Intention to disenroll	NH Pensacola	0.16	0.08	0.04	0.01	0.07	0.07	0.07	0.05	0.0%	0.4%	33.9%
	Travis AFB	0.10	0.05	0.04	0.01	0.05	0.05	0.05	0.03	0.0%	0.8%	26.7%
	McGuire AFB/Ft. Dix	0.20	0.06	0.10	0.02	0.10	0.10	0.10	0.07	0.2%	0.2%	25.7%
	Ft. Leonard Wood	0.12	0.13	0.05	0.02	0.08	0.08	0.08	0.06	0.2%	1.1%	22.2%
	Lackland AFB	0.01	0.04	0.04	0.05	0.03	0.03	0.03	0.04	0.0%	0.4%	22.1%
	Out/Area-GulfSouth	0.08	0.03	0.03	0.02	0.04	0.04	0.04	0.03	0.1%	0.6%	20.7%
	Ft Wainwright	0.10	0.15	0.04	0.03	0.08	0.08	0.08	0.06	0.3%	0.1%	20.5%
	NH Cherry Point	0.12	0.11	0.04	0.04	0.07	0.07	0.07	0.06	0.1%	1.2%	20.5%

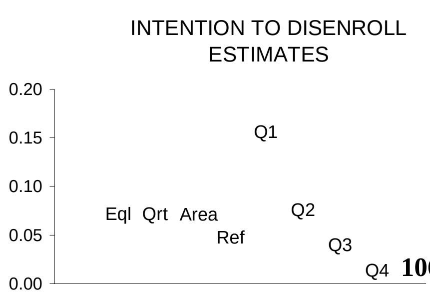
*Variables are indicators with values of 0-1

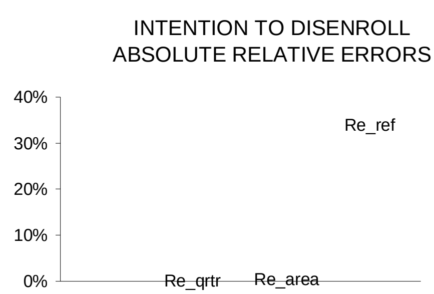
As table 2 indicates, estimates that were found to have larger relative errors for the reference period technique also showed a distinct trend over the four quarters. This trend is not apparent for the same variables when estimates are viewed at the larger national or regional levels, but is only apparent by military *catchment* area. This is better illustrated by figures 2-4 below. Figures 2-4 are plots of the estimates for all four quarters and all weighting techniques and relative errors for all techniques for the *intention to disenroll* indicator variable. Figures 2a and 2b reveal these estimates and relative errors by the military *catchment* area level for Pensacola, Florida. Figures 3a and 3b show the estimates and errors for the entire Gulf South region (which encompasses the Pensacola military *catchment* area), and Figures 4a and 4b are plots of the estimates and relative errors on the overall level.

Clearly, there is a noticeable downward trend from quarters one to four, which indicates that respondents' *intention to disenroll* has been steadily decreasing across the four quarters of 2001. It is this trend that causes the relative error for the reference period to reach nearly 34%, while the relative errors for the other techniques are negligible. On the regional level, the trend is still apparent, though somewhat less pronounced, and the relative error for the reference period drops to a little under 17%. On the overall level, however, a trend is harder to distinguish, and the relative error for all techniques approximate zero. Similar trends can be seen for the other estimates with large relative errors.

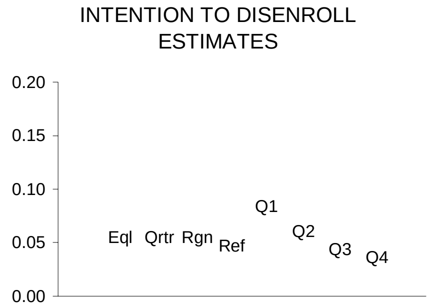
This indicates that combining HCSDB quarterly estimates by assigning equal weights to each quarter is indeed an appropriate technique for these data, as little difference was found between the equal weighting technique and the other techniques demonstrated above. Caution should be used, however, when combining estimates by military *catchment* area, as estimates obtained using the reference period technique may differ from those obtained using the equal weighting technique due to trends over time in the quarterly data.

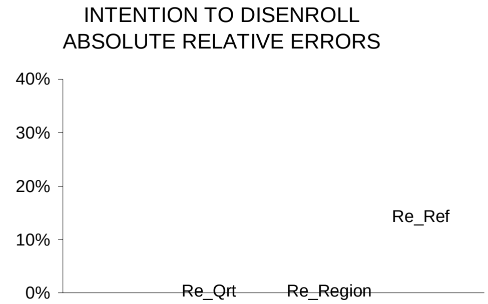
Figures 2a-2b: Pensacola Area



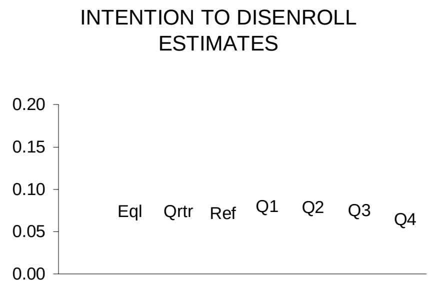


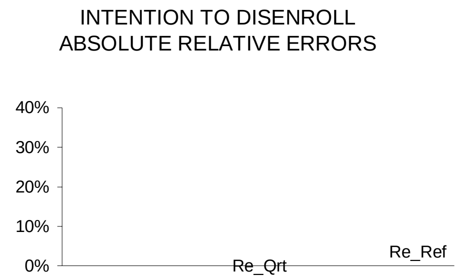
Figures 3a-3b:
GulfSouth Region





Figures 4a-4b:
Overall





REFERENCES

Alexander, C. H. (1998), "Recent Developments in the American Community Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, Alexandria, VA: American Statistical Association, 92-100.

Alexander, C. H. and Wetrogan S. (2000), "Integrating The American Community Survey and the Intercensal Demographic Estimates Programs," Proceedings of the American Statistical Association, Survey Research Methods Section, Alexandria, VA: American Statistical Association., 295-300.

Alexander, C. H. (2002) "Still Rolling: Leslie Kish's 'Rolling Samples' and The American Community Survey," *Survey Methodology*, 28:1, 35-41.

Binder, D.A. and Hidiroglou, M.A. (1989). "Modeling and Estimation of Repeated Surveys," *Survey Methodology*, 15, 29-45.

Chand, N. and Alexander C.H. (2000), "Multi Year Averages From a Rolling Sample Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, Alexandria, VA: American Statistical Association., 301-306.

Kish, L. (1965), *Survey Sampling*. New York: John Wiley.

Kish, L. (1979). "Samples and Censuses," *International Statistical Review,* 47, 99-109.

Kish, L. (1990), "Rolling Samples and Censuses," *Survey Methodology*, 16: 1, 63-7 1.

Kish, L. (1998), "Space/Time Variations and Rolling Samples," *Journal of Official Statistics*, Vol. 14, No. 1, 31-46.

Kish, L. (1999), "Cumulating/Combining Population Surveys," *Survey Methodology*, 25-2, 129-138.

Singh, A. and Merkouris, P. (1995). "Composite Estimation by Modified Regression for Repeated Surveys," Proceedings of the American Statistical Association, Survey Research Methods Section, Alexandria, VA: American Statistical Association, 420-425.

Weidman, L., C.H. Alexander, G. Diffendal, S. Love (1995), "Estimation Issues for the Continuous Measurement Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, Alexandria, VA: American Statistical Association, 596-601.