Design Effect Variation for Estimates Derived from the MEPS (1996-1998)

William Yu, Agency for Health Care Research and Quality
2101 E. Jefferson St., Suite 5W29, Rockville, MD 20852

KEY WORDS: Design Effect, Health Care Expenditure, MEPS

Introduction

The sample design of the Medical Expenditure Panel Survey (MEPS) is an overlapping panel design characterized by a multistage, complex area probability design that includes disproportionate sampling of specified policy relevant population groups. Standard methods of variance estimation which assume simple random sampling generally result in an under-estimation of variance, when used with data from a complex survey design (Cohen S., 1982). The extent of this departure from simple random sampling assumptions and its impact on the variances of survey estimates may be measured by the design effect. A summary of design effect variation for 1997 MEPS estimates has been previously published (Yu, W., 2001). Based on data from the 1996, 1997, and 1998 MEPS, this paper evaluates and contrasts: (1) the design effects achieved over time for national estimates of health care utilization and expenditures; (2) design effect variation by alternative population subgroups and by different geographic regions of the nation; and (3) the level of design effect variation in related survey estimates.

MEPS Household Component

MEPS is conducted annually to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. It is co-sponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS).

The core survey for MEPS is the Household Component (HC). The MEPS-HC collects data through an overlapping panel design. In this design, data are collected through a series of five rounds of interviews over a period of 2 ½ years. Interviews are conducted with one member of each family, who reports on the health care experiences of the entire family. Two calendar years of medical expenditure and utilization data are collected in each household and captured using computer-assisted personal interviews. This series of data collection rounds is launched again each subsequent year on a new sample of households to provide overlapping samples of survey data that provide continuous and current estimates of health care expenditures (Cohen J., 1997).

The sampling frame for the MEPS-HC is drawn from respondents to the previous year's National Health Interview Survey (NHIS), conducted by NCHS. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized population, with over sampling of Hispanics and blacks.

Source and Definition of Data

This study is based on three years of MEPS-HC data: HC-012 (1996), HC-020 (1997), and HC-028 (1998). Expenditures in MEPS are defined as the sum of direct payments for health care provided during the year, including out-of-pocket payments and payments by private insurance, Medicare, Medicaid, and other sources. Payments for over the counter drugs, alternative care services, and phone contacts with medical providers are not included in MEPS total expenditure estimates. Indirect payments unrelated to specific medical events such as Medicaid Disproportionate Share and Medicare Direct Medical Education subsidies also are not included (Cohen JW, Machlin SR, Zuvekas SH, *et al.*, 2000).

The use and expenditure data included in this paper were derived from the MEPS-HC and Medical Provider Components (MPC). MPC data were collected for some office-based visits to physicians (or medical providers supervised by physicians), hospital-based events (e.g. inpatient stays, emergency room visits, and outpatient department visits), and prescribed medicines. HC data were collected for physician visits, dental and vision services, other medical equipment and services, and home health care not provided by an agency. Data on expenditures for care provided by home health agencies were collected only in the MPC. MPC data were used if complete; otherwise HC data were used if complete. Missing data for events where HC data were not complete and MPC data were not collected or not complete were derived through an imputation process (Cohen S. and Carlson B., 1994).

In this study, the evaluation of design effect variation is based on cross-sectional estimates for all in scope persons living in the United States on December 31 of each target year. The sample sizes for the target

The views expressed in this paper are those of the author and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred. The author wishes to thank Trena Ezzati-Rice and Doris C. Lefkowitz for their helpful reviews of the paper.

populations were 21,326, 32,226, and 22,630 for calendar years 1996, 1997, and 1998, respectively. Table 1 contains 25 utilization and expenditure variables selected from the 1996-1998 MEPS consolidated data files for this study.

Table 1 - Utilization and Expenditure Variables by Health Services Category

Health Service Category	Utilization Variables	Expenditure Variables
All Health Services	–	TOTEXPyy
Office Based Visits		
Total Office Based Visits (Physician + Non-physician + Unknown)	OBTOTVyy	OBVEXPyy
Visits to Physicians	OBDRVyy	OBDEXPyy
Visits to Non-physicians	OBOTHVyy	OBOEXPyy
Hospital Outpatient Visits		
Total Outpatient Visits (Physician + Non-physician + Unknown)	OPTOTVyy	OPEXPyy(1)
Visits to Physicians	OPDRVyy	OPYEXPyy(1)
Visits to Non-physicians	OPOTHVyy	OPNEXPyy(1)
Total Emergency Room Visits	ERTOTyy	EREXPyy(1)
Total Inpatient Stays (Including Zero Night Stays)	IPDISyy	IPEXPyy(1)
Total Dental Visits	DVTOTyy	DVTEXPyy
Total Home Health Care	HHTOTDyy	HHEXPyy(1)
Other		
Prescription Medicines	RXTOTyy	RXEXPyy
Vision Aids	–	VISEXPyy
Other Medical Supplies and Equipment	–	OTHEXPyy

yy-applicable year of data (e.g., 96, 97, or 98)

(1) Sum of facility and "separately billing doctor"(SBD) expense

Variables used to form population, geographic, and economic subgroups in this analysis are gender (male, female), age (<=17, 18-44, 45-64, 65+), race/ethnicity (Hispanic, black/non-Hispanic, others), Census region (Northeast, Midwest, South, West), and family income as percent of poverty line (0%-199%, 200%+).

Design Effect in the MEPS-HC

Given the complex nature of the MEPS-HC survey design, the assumptions of independence and equal selection probabilities are not satisfied. The resulting impact on variance estimation may be described as follows (Cohen, S., 1982):

$$\sigma^2_{complex} = \sigma^2_{SRS} [1 + \rho (\tilde{n} - 1)]$$

where

$\sigma^2_{complex}$ is the true variance of a statistic given the complex survey design,

σ^2_{SRS} is the variance estimate obtained for the statistic under simple random sampling assumptions,

ρ is the intra cluster correlation coefficient, and

$\tilde{n}$ is the average cluster size.

The design effect consequently is expressed as:

$$\text{Design Effect} = (\sigma^2_{complex} / \sigma^2_{SRS}) = [1 + \rho (\tilde{n} - 1)]$$

The design effect deviates from unity when the effects of clustering are dominant in a survey design and the average cluster size is moderate to large. Variances of all estimated parameters presented in this paper were derived using SUDAAN (Shah, Barnwell, Bieler, *et al.*, 1996) with the Taylor series linearization method to account for the complex survey design.

Evaluation of Design Effect Variation

For each of the three years covered in this study, design effects were determined for each of the 25 selected survey statistics for the nation overall as well as for alternative population subgroups and different geographic regions of the nation. Figure 1 compares the level of design effects achieved for the selected set of national mean health care use and expenditure estimates from 1996 to 1998.

The overall design effects are summarized further in Table 2 below where the variables with the lowest and highest design effects are noted in {}. For example, in 1996, the design effects ranged from 1.29 for the estimated number of home health provider days to 3.17 for the estimated number of dental care visits with an overall average of 1.76.

Table 2 - Summary of Design Effect Variation: MEPS 1996-1998

Year	Lowest Estimate {variable}	Highest Estimate {variable}	Mean (Std)
1996	1.29 {# home health provider days}	3.17{# dental care visits}	1.76 (0.40)
1997	1.14 {#outpatient Dr visits}	3.59 {# dental care visits}	2.06 (0.58)
1998	0.66 {Total other equip/ sply (excl. Diab.) exp}	3.61 {Total glass/contact lens exp}	1.94 (0.73)

Figure 2 and Table 3 present comparisons of average design effects from the selected health care use and expenditure measures across all of the alternative population groups and by different geographic regions of the nation from 1996 to 1998. Comparing the three years, 1996 has the lowest average design effect for almost all population subgroups. The only exception is in the male

Figue 1-Design Effects for Mean Use and Expenditure Estimates (1996-1998 MEPS)

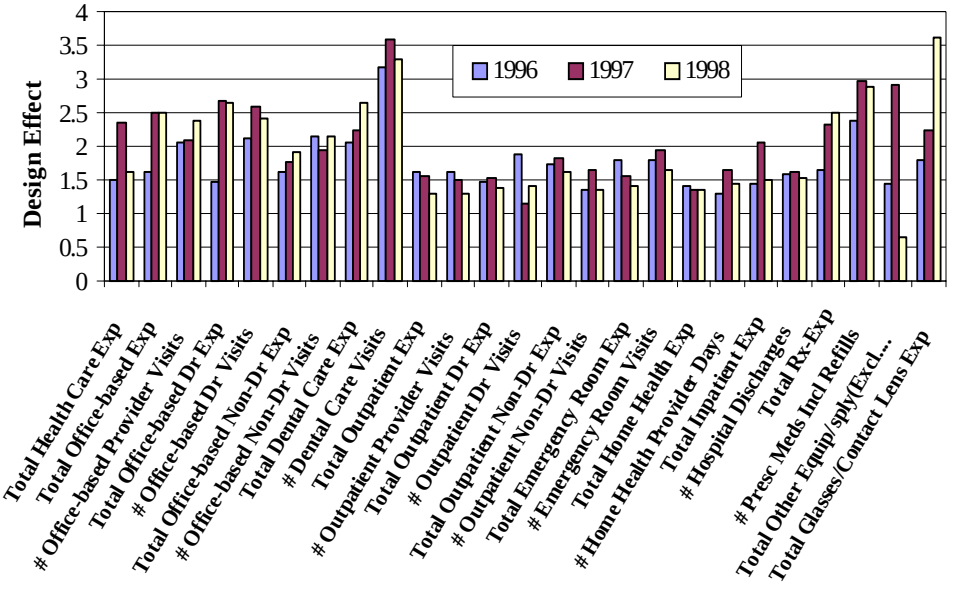
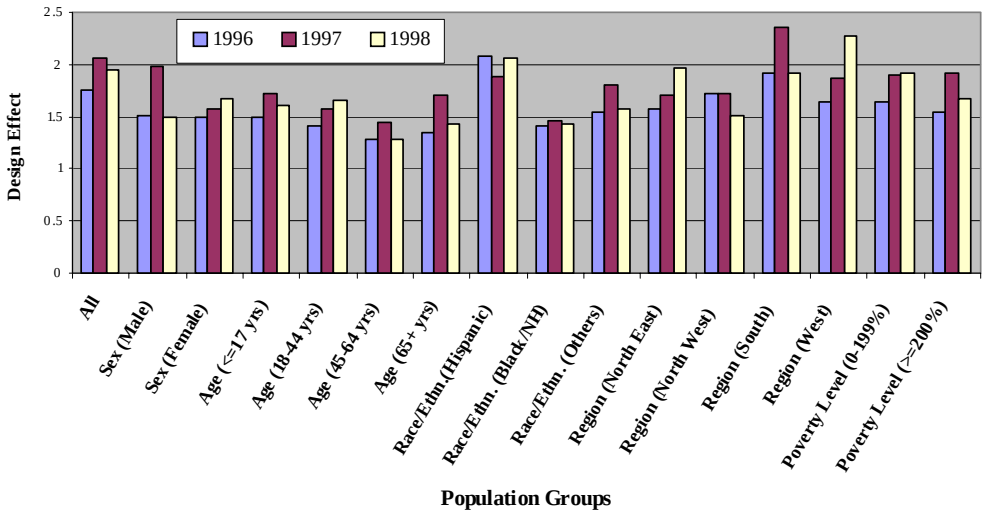


Figure 2 - Average Design Effects for Population Groups (1996-1998 MEPS)



group where the average design effect is 1.51 for 1996, but 1.50 for 1998. Overall, age group 45-64 has the lowest average design effects at 1.29, 1.45, and 1.29 for 1996, 1997, and 1998 respectively. The highest average design effects are 2.07 for the Hispanic group in 1996, 2.35 for the South census region in 1997, and 2.27 for the West census region in 1998.

Table 3 - Summary of Design Effects by Population Groups: MEPS 1996-1998

	Mean(Std)**		
Population Subgroups	1996	1997	1998
All	1.76 (0.40)	2.06 (0.58)	1.94 (0.73)
Gender			
Male	1.51 (0.30)	1.98 (0.52)	1.50 (0.33)
Female	1.50 (0.30)	1.58 (0.36)	1.67 (0.51)
Age Group			
<=17 years	1.50 (0.46)	1.72 (0.48)	1.61 (0.41)
18-44 years	1.41 (0.24)	1.57 (0.25)	1.66 (0.29)
45-64 years	1.29 (0.23)	1.45 (0.33)	1.29 (0.39)
65+ years	1.34 (0.24)	1.71 (0.63)	1.43 (0.44)
Race/Ethnicity			
Hispanics	2.07 (0.50)	1.88 (0.63)	2.06 (0.56)
Black/non-Hispanics	1.41 (0.35)	1.46 (0.50)	1.43 (0.44)
Others	1.54 (0.32)	1.80 (0.43)	1.58 (0.49)
Census Region			
North East	1.58 (0.56)	1.71 (0.72)	1.97 (0.90)
Midwest	1.72 (0.56)	1.72 (0.62)	1.51 (0.64)
South	1.91 (0.48)	2.35 (0.91)	1.91 (0.61)
West	1.64 (0.52)	1.87 (0.54)	2.27 (1.53)
Poverty Level			
0 - 199%	1.64 (0.29)	1.90 (0.51)	1.92 (0.38)
200% +	1.55 (0.31)	1.92 (0.43)	1.68 (0.60)

**Mean and standard deviation of design effects for the 25 mean use and expenditure estimates listed in Table 1.

There is a notable downward trend for the values of average design effect by ascending age groups 0-17, 18-44, and 45-64 in 1996 and 1997. For race/ethnicity, the black/non-Hispanic group has the lowest average design effects for all three years. There is no notable pattern of difference in the average design effects for people classified in each of the census regions or below or above 200 percent of poverty line.

The following subset of representative use and expenditure measures were selected for a more detailed study of design effect variation:

- Total health care expenditures
- Total office-based expenditures
- Total Rx-expenditures
- Total inpatient expenditures
- Total # of office-based provider visits
- Total # of Rx medicine including refills
- Total # of hospital discharges

For each of the selected variables, domain estimates of design effects were generated in terms of population means. The domain estimates are defined by marginal or cross-classified distributional categories of the selected variables.

The quartile boundaries on sample size for the set of domain estimates under investigation were cross-classified by the quartile boundaries on the resultant mean estimates of the respective health care utilization and expenditure measures, yielding 16 strata. Within each of these strata and their marginal classes, the average design effect and the standard error of the design effects were derived.

The most notable pattern in design effect variability was the positive incremental impact of sample size on the average design effect. This pattern was most obvious for the estimates of design effect of the mean on total number of prescription medicines including refills as shown in Figures 3-5. Similar patterns were observed for the other selected health care utilization and expenditure measures. No distinct relationship was observed between the average design effect and the respective quartile boundaries which characterized the distribution of criterion-variable domain estimates.

Summary

A part of the overall precision requirements for the MEPS survey is the achievement of an average design effect of approximately 1.7 (Cohen S.B. 2000) for the survey estimates of policy relevant population subgroups (e.g., households with Hispanics and blacks, persons with family incomes less than 200% of poverty line, persons 65 years or older, adults and/or children with functional impairments). The study findings confirmed that this requirement was generally satisfied (mean ± std, see Table 3) with respect to selected measures of health care utilization and expenditures for the three data years studied.

Positive incremental effects on the average design effect were observed in relation to sample size for all the selected variables (Figures 3-5). In addition, for the selected measures of health care utilization and expenditures, most of the average design effects by population subgroups (Table 3) appeared to be larger in 1997 compared to 1996 and 1998. This finding was

Figure 3 - Effect of Sample Size on the Values of Average Design Effect (1996 MEPS)

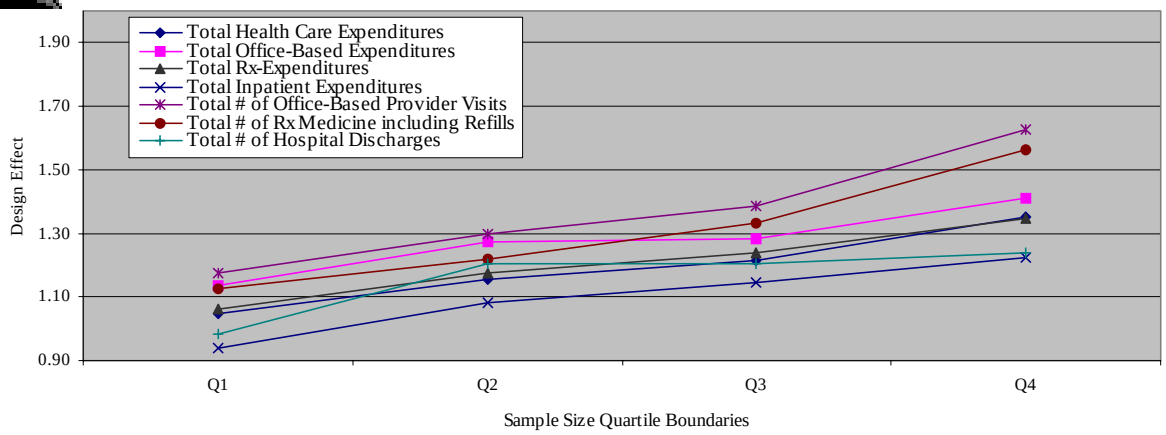


Figure 4 - Effect of Sample Size on the Values of Average Design Effect (1997 MEPS)

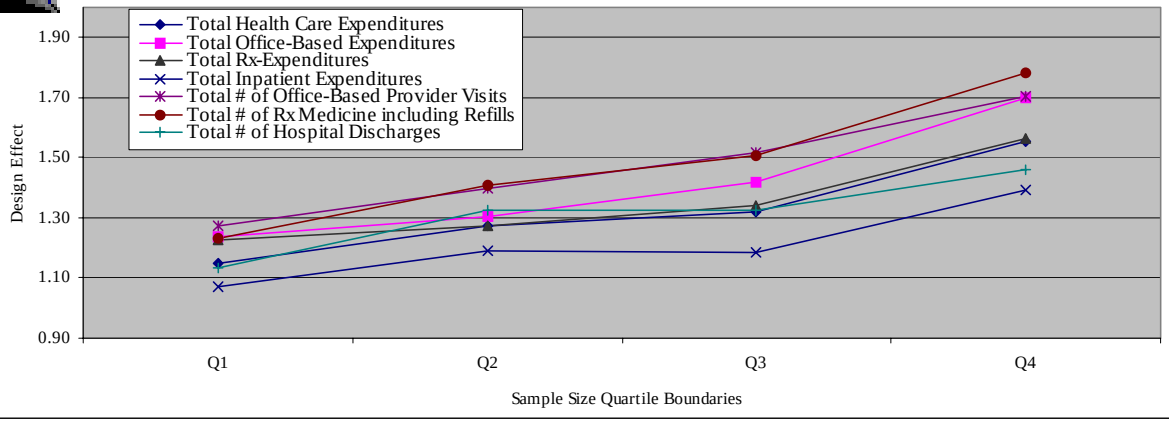
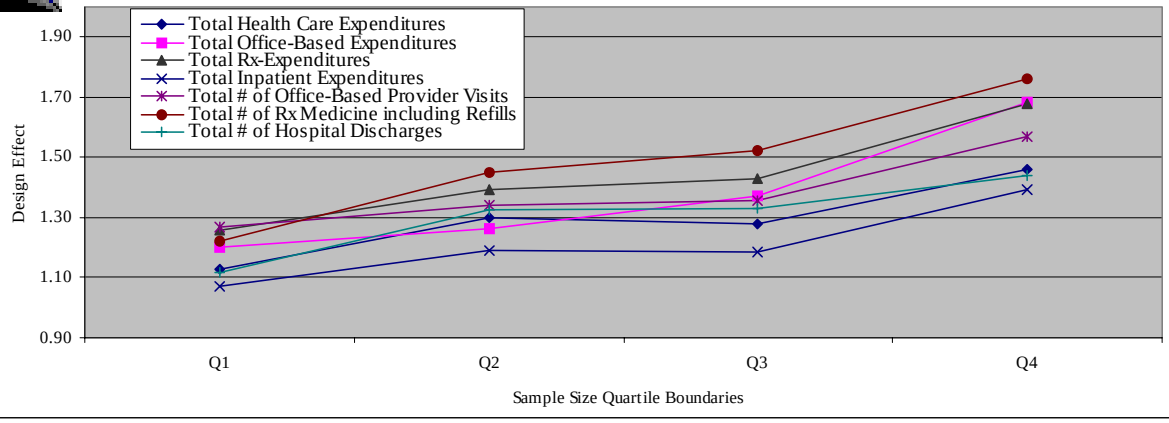


Figure 5 - Effect of Sample Size on the Values of Average Design Effect (1998 MEPS)



based on sample sizes of 21,326, 32,226 and 22,630 for the 1996, 1997 and 1998 MEPS, respectively.

It also was observed that the design effects for the mean estimate of uncommon events (e.g., 1.59, 1.62, 1.53 for number of hospital discharges in 1996, 1997, and 1998, respectively) were closer to unity than other planned events (e.g., 3.17, 3.59, and 3.30 for number of dental care visits in 1996, 1997, and 1998 respectively). One possible explanation is that the ultimate cluster units in the MEPS-HC sample design are the household or family. A strong positive correlation would be expected between individuals in the same household with respect to the total number of dental care visits compared to the number of times discharged from the hospital.

References

Cohen SB., "Comparison of Design Effect and Relative Variance Curve Strategy for Variance Estimation from Complex Survey Data," presented at the annual meetings of the American Public Health Association, 1982.

Yu WW. (2001), "Precision of Survey Estimates Derived from the Medical Expenditure Panel Survey(MEPS)." 2001 Proceedings of the American Statistical Association, Section on Survey Research Methods [CD-ROM], Alexandria VA: American Statistical Association.

Cohen JW., "Design and Methods of the Medical Expenditure Panel Survey Household Component." Rockville (MD): Agency for Health Care Policy and Research; 1997. MEPS Methodology Report No.1. AHCPR Pub. No. 97-0026.

Cohen JW, Machlin SR, Zuvekas SH, *et al*., "Health care expenses in the United States, 1996." Rockville (MD): Agency for Healthcare Research and Quality; 2000. MEPS Research Findings 12. AHRQ Pub. No. 01-0009.

Cohen S. and Carlson B., "A comparison of household and medical provider reported expenditures in the 1987 NMES." *Journal of Official Statistics* 1994; 10(1):3-29.

Shah, B.V., B.G. Barnwell, G.S. Bieler, K.E. Boyle, R.E. Folsom, L. Lavange, S.C. Wheeless, and R.Williams, Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0, Research Triangle Park, NC: Research Triangle Institute, 1996.

Cohen SB, "Sample design of the 1997 Medical Expenditure Panel Survey Househpld Component." Rockville (MD): Agency for Healthcare Research and Quality; 2000. MEPS Methodology Report No. 11. AHRQ Pub. No. 01-0001.