DESIGN EFFECTS OF SURVEY ESTIMATES DERIVED FROM THE 1996 MEDICAL EXPENDITURE PANEL SURVEY (MEPS)

William Yu, Agency for Health Care Policy and Research
2101 E. Jefferson St., Suite 500, Rockville, MD 20852

Key Words: design effect, health care utilization, Medical Expenditure Panel Survey (MEPS)

Abstract:

The sample design of the 1996 Medical Expenditure Panel Survey(MEPS) is characterized by a multistage, complex area probability design that includes disproportionate sampling of specified policy relevant population groups. 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. The design effect is defined as the ratio of the true variance of a statistic to the variance derived under simple random sampling assumptions. It is expected that the variances of survey estimates derived from the 1996 MEPS will generally exhibit design effects that are greater than unity. This paper will evaluate the design effects achieved for national estimates of health care utilization and insurance coverage; the level of design effect variation in related survey estimates; and design effect variation by alternative population subgroups and by different geographic regions of the nation. The results may be used to help improve the sample design specification for the selection of future new sample panels of households for the annual MEPS.

Introduction

Complex survey design components often include unequal selection probabilities of elements in the population, with several stages of clustering. 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. The design effect is defined as the ratio of the true variance of a statistic to the variance derived under simple random sampling assumptions. Based on data from the Household Component of the 1996 Medical Expenditure Panel Survey (MEPS), this paper will evaluate the design effects achieved for national estimates of health care utilization and insurance coverage; the level of design effect variation in related survey estimates; and design effect variation by alternative population subgroups and by different geographic regions of the nation.

Design of the MEPS Household Component

The MEPS Household Component (MEPS HC), a nationally representative survey of the U.S. civilian noninstitutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The survey is sponsored by the Agency for Health Care Policy and Research with co-sponsorship by the National Center for Health Statistics.

The 1996 MEPS Household Component sample was selected from households that responded to the 1995 National Health Interview Survey (NHIS). This selection consists of 195 Primary Sampling Units (PSUs), 1,675 sample segments (second-stage sampling units) and 10,597 responding households. It is designed to produce unbiased estimates for the four Census regions, with over-sampling of households with Hispanics and blacks at a ratio of approximately 2.0:1 for Hispanics and 1.5:1 for blacks. The average design effect target for survey estimates of health care use and expenditure estimates for the 1996 MEPS is 1.6 (Cohen S. 1997).

The 1995 NHIS response rate achieved for MEPS-eligible households was 93.9 percent. Of 10,639 responding NHIS dwelling units eligible for MEPS, 99.6 percent were identified with enough information to allow MEPS data collection. Of the11,424 eligible reporting units targeted for interviews in Round 1, 9,488 (83.1 percent) responded. Overall, the joint NHIS-Round 1 response rate for the 1996 MEPS household survey was 77.7 percent (.939 x .996 x .831).

The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of six rounds of interviews over a 2 ½-year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for 2 calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when

The views expressed in this paper are those of the author and no official endorsement by the Department of Health and Human Service or the Agency for health Care Policy and Research is intended or should be inferred. The author wishes to thank Dr. Steven Cohen and Dr. Alan Monheit for their helpful reviews of the paper.

combined with other ongoing panels, will provide continuous and current estimates of health care expenditures (Cohen J. 1997).

Design Effect in the 1996 MEPS HC

Given the complex nature of the 1996 MEPS HC survey design, the assumptions of independence and equal selection probabilities are not satisfied. Its impact on variance estimation is best described as follows:

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

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 sample random sampling assumptions,
ρ is the intra cluster correlation coefficient, and
$\tilde{n}$ is the average cluster size.

The design effect is consequently 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 the Taylor series linearization method to account for survey design complexities (shah, 1996).

Evaluation of Design Effect Variation in the 1996 MEPS HC

Based on the 1996 MEPS HC data, design effects are determined for a representative set of survey statistics which estimate medical care utilization and health insurance coverage of the U.S. population. For the nation, the design effects ranged from 1.01 for "number of zero-night hospitals stays" to 6.55 for "PID covered by Medicaid" with an overall average of 2.07. Figure 1 is a bar chart comparing the level of design effects achieved for national estimates of health care utilization and insurance coverage.

Demographic variables used to form population subgroups in this analysis include gender (male, female), age (<19, 20-44, 45-64, 65+), race/ethnicity (Hispanic, black/non-Hispanic, others), Census region (Northeast, Midwest, South, West).

Figure 2 presents a comparison across all the alternative population groups and by different geographic regions of the nation. Overall, Hispanics and black/non-Hispanics have the two lowest average design effects at 1.23 and 1.38 respectively while the Census region - South has the highest average design effect of 2.27. The average design effects for males and females appear to be identical at 1.63. There is a notable difference between age groups on the value of average design effect. The average design effect is approximately 1.6 for age groups 0-19 and 20-44, compared to 1.3 for age groups 45-64 and 65+. For the census regions, Northeast has the lowest average design effect at 1.62 and South has the highest at 2.27.

For a selected set of representative statistics, domain estimates were generated in terms of population means and population proportions when appropriate. The domain estimates are defined by marginal or cross-classified distributional categories of the selected demographic variables. For example, for the mean number of physician visits within specific age-race/ethnicity-sex-census region classes of the U.S. population, the domain estimate, $\bar{Y}_g$, is derived as:

$$\bar{Y}_g = (\Sigma_i W_i X_{gi} Y_i) / (\Sigma_i W_i X_{gi})$$

where

Y_i is the i[th] individual's number of physician visits,
W_i is the i[th] individual's sampling weight,
X_{gi} = 1 if the individual is a member of the g[th] age-race/ethnicity-sex-census region domain,
= 0 otherwise.

Tables 1-4 present the design effect variation for domain estimates of the selected variables expressed in terms of population means or proportions. Three hundred (300) domain estimates were grouped into sixteen strata defined by the cross-classification of quartile boundaries on sample size and mean (or proportion) estimates of the respective health care measures. Within each of these strata and their marginal classes, the average design effect, its standard error and sample range of design effects were derived.

The most notable pattern in design effect variability was the positive incremental effect of sample size on the value of average design effect. The pattern was most obvious for domain estimates of the proportion of population with private insurance. As shown in Figure 3, the average design effect ranged from 1.224 (SE = .064) on sample size less than or equal to 155, to 3.173 (SE = .142) for sample size greater than 1,208. Similar, but more moderate, patterns were observed for the other selected health care utilization measures.

No distinct relationship was observed in Figure 4 between the average design effect and the respective quartile boundaries which characterized the distribution of criterion variable domain estimates. However, a positive incremental effect on the average design effect was observed in relation to the quartile distribution of domain estimates for private insurance coverage. As

Figure 1 - Design Effect for Mean Estimate of Medical Utilization and Insurance Coverage Measures from the 1996 MEPS - National Average

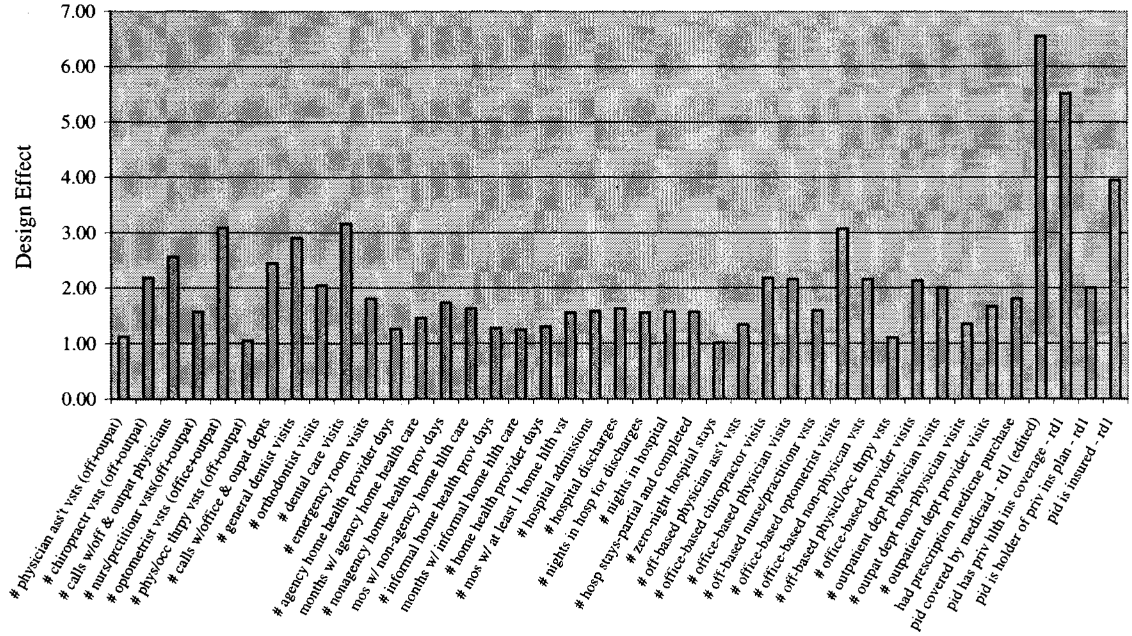
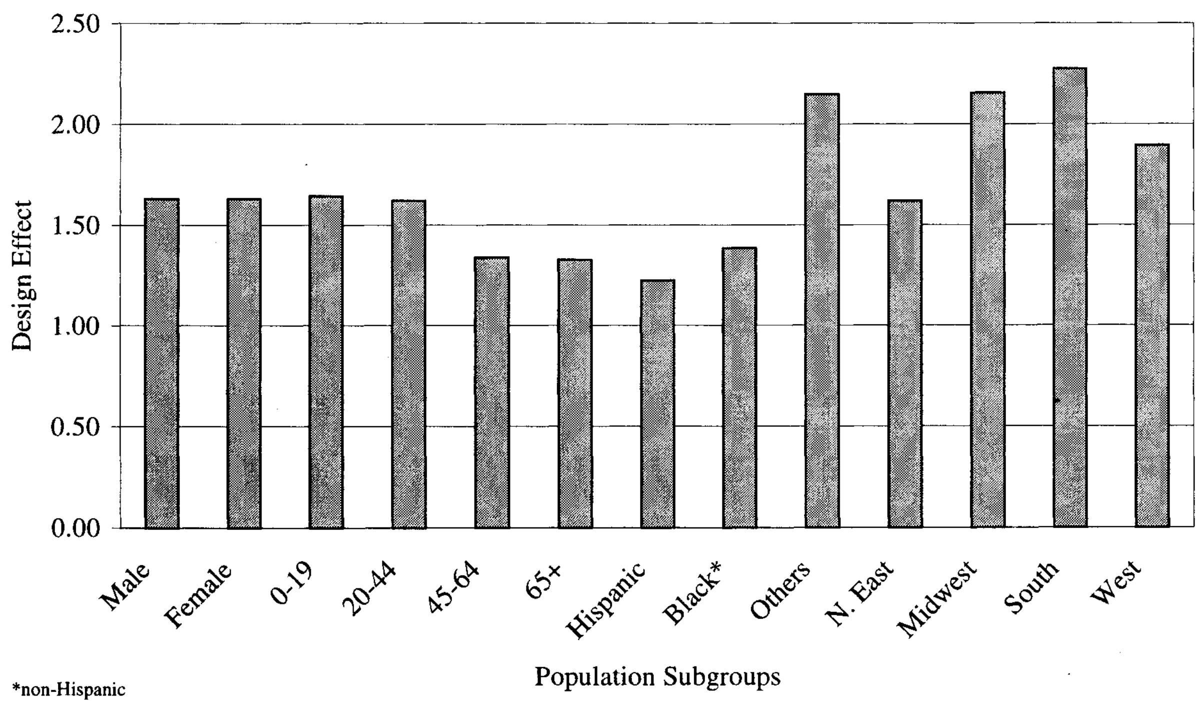


Figure 2 - Average Design Effect for Alternate Population Subgroups



*non-Hispanic

Data Source: 1999 MEPS HC-003

Table 1 - Design Effect Variation for Domain Estimates of the Mean Number of Hospital Admissions

Sample Size	Design	Mean Number of Hospital Admissions (Quartile Boundaries)				
Quartile Boundaries	Effect	< 0.059	0.059 - 0.090	0.091 - 0.122	> 0.122	Total
<= 155	Mean	0.633	0.614	0.723	0.834	0.715
	SE	0.057	0.069	0.099	0.083	0.041
	Range	< 0.193 - 1.087 >	< 0.255 - 1.316 >	< 0.259 - 1.325 >	<0.260 - 2.203 >	< 0.193 - 2.203 >
156 - 495	Mean	1.073	1.011	0.991	1.006	1.020
	SE	0.131	0.158	0.102	0.042	0.049
	Range	< 0.354 - 2.168 >	< 0.420 - 2.881 >	< 0.402 - 1.930 >	< 0.505 - 1.456 >	< 0.354 - 2.881 >
496 - 1,229	Mean	1.076	1.284	1.032	1.078	1.124
	SE	0.055	0.131	0.057	0.080	0.046
	Range	< 0.598 - 1.671 >	< 0.493 - 2.779 >	< 0.535 - 1.535 >	< 0.518 - 1.753 >	< 0.493 - 2.779 >
1,230 - 21,571	Mean	1.230	1.912	1.318	1.058	1.242
	SE	0.082	0.067	0.059	0.069	0.037
	Range	< 0.862 - 1.780 >	< 0.572 - 1.805 >	< 0.701 - 2.331 >	< 0.790 - 1.303 >	< 0.572 - 2.331 >
Total	Mean	0.994	1.058	1.113	0.970	1.035
	SE	0.049	0.061	0.044	0.037	0.024
	Range	< 0.193 - 2.168 >	< 0.255 - 2.881 >	< 0.259 - 2.331 >	< 0.260 - 2.203 >	< 0.193 - 2.881 >

Table 2 - Design Effect Variation for Domain Estimates of the Mean Number of Physician Visits

Sample Size	Design	Mean Number of Office-based and Outpatient Physician Visits (Quartile Boundaries)				
Quartile Boundaries	Effect	< 2.266	2.266 - 3.327	3.328 - 4.748	> 4.748	Total
<= 155	Mean	1.141	0.942	0.839	0.780	0.915
	SE	0.117	0.125	0.085	0.044	0.047
	Range	< 0.578 - 2.954 >	< 0.436 - 1.632 >	< 0.328 - 1.523 >	<0.381 - 1.328 >	< 0.328 - 2.954 >
156 - 495	Mean	1.493	1.271	1.150	1.312	1.238
	SE	0.106	0.161	0.094	0.081	0.054
	Range	< 0.531 - 1.912 >	< 0.581 - 3.080 >	< 0.291 - 1.698 >	< 0.646 - 2.375 >	< 0.291 - 3.080 >
496 - 1,229	Mean	1.401	1.314	1.394	1.459	1.389
	SE	0.082	0.080	0.138	0.129	0.051
	Range	< 0.865 - 3.054 >	< 0.744 - 1.921 >	< 0.733 - 2.881 >	< 0.932 - 2.569 >	< 0.733 - 3.054 >
1,230 - 21,571	Mean	1.382	1.551	1.760	1.518	1.601
	SE	0.104	0.064	0.074	0.090	0.043
	Range	< 0.981 - 1.921 >	< 0.858 - 2.821 >	< 1.239 - 2.618 >	< 1.189 - 1.839 >	< 0.858 - 2.821 >
Total	Mean	1.272	1.361	1.347	1.163	1.286
	SE	0.053	0.054	0.063	0.054	0.028
	Range	< 0.531 - 3.054 >	< 0.436 - 3.080 >	< 0.291 - 2.881 >	< 0.381 - 2.569 >	< 0.291 - 3.080 >

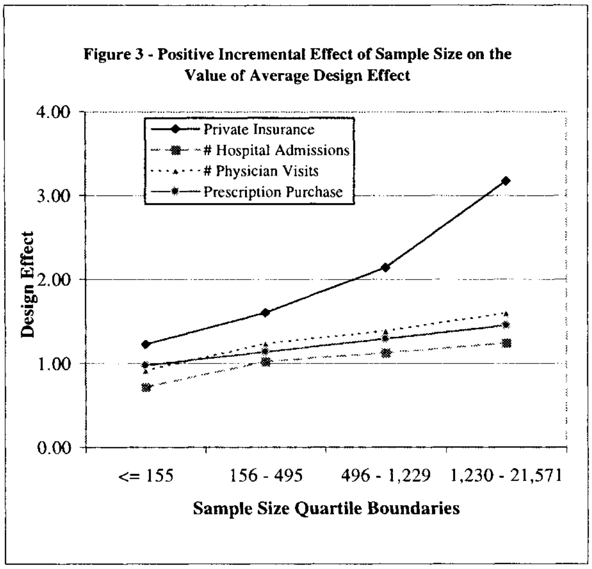
Table 3 - Design Effect Variation for Domain Estimates of the Proportion of Population with Prescription Medicine Purchase

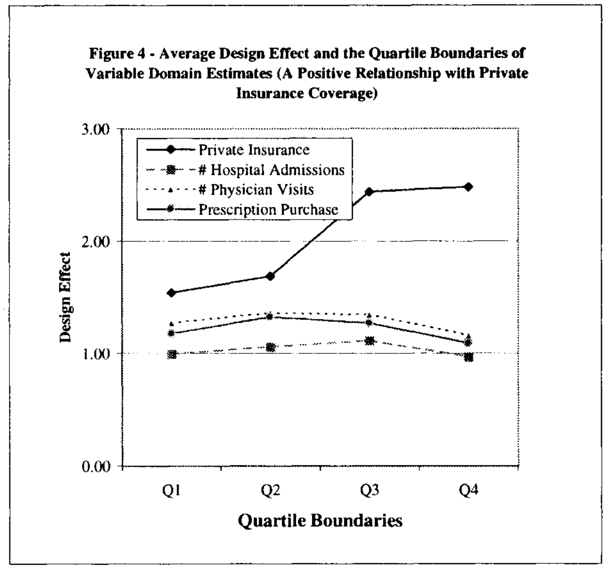
Sample Size	Design	Proportion of Population with Prescription Medicine Purchase (Quartile Boundaries)				
Quartile Boundaries	Effect	< 0.535	0.535 - 0.629	0.629 - 0.755	> 0.755	Total
<= 155	Mean	0.990	1.088	1.088	0.846	0.975
	SE	0.079	0.098	0.095	0.077	0.045
	Range	< 0.635 - 1.529 >	< 0.617 - 1.685 >	< 0.585 - 2.048 >	<0.386 - 2.130 >	< 0.386 - 2.130 >
156 - 495	Mean	1.048	1.211	1.061	1.214	1.138
	SE	0.072	0.103	0.059	0.061	0.037
	Range	< 0.442 - 1.839 >	< 0.684 - 1.974 >	< 0.632 - 1.350 >	< 0.588 - 1.812 >	< 0.442 - 1.974 >
496 - 1,229	Mean	1.270	1.383	1.184	1.323	1.290
	SE	0.047	0.060	0.057	0.045	0.028
	Range	< 0.788 - 1.762 >	< 0.784 - 1.874 >	< 0.875 - 1.583 >	< 1.089 - 1.607 >	< 0.784 - 1.874 >
1,230 - 21,571	Mean	1.382	1.438	1.529	1.271	1.451
	SE	0.054	0.027	0.060	0.075	0.028
	Range	< 1.099 - 1.703 >	< 1.052 - 1.667 >	< 0.975 - 2.140 >	< 1.095 - 1.505 >	< 0.975 - 2.140 >
Total	Mean	1.175	1.322	1.268	1.091	1.214
	SE	0.035	0.034	0.042	0.045	0.020
	Range	< 0.442 - 1.839 >	< 0.617 - 1.974 >	< 0.585 - 2.140 >	< 0.386 - 2.130 >	< 0.386 - 2.140 >

Table 4 - Design Effect Variation for Domain Estimates of the Proportion of Population with Private Insurance

Sample Size	Design	Proportion of Population with Private Insurance (Quartile Boundaries)				
Quartile Boundaries	Effect	< 0.451	0.451 - 0.621	0.621 - 0.725	> 0.725	Total
<= 155	Mean	1.093	1.343	1.476	1.202	1.224
	SE	0.076	0.116	0.202	0.300	0.064
	Range	< 0.581 - 2.314 >	< 0.634 - 3.121 >	< 0.782 - 1.915 >	<0.262 - 2.254 >	< 0.262 - 3.121 >
156 - 494	Mean	1.520	1.591	1.659	1.667	1.604
	SE	0.132	0.133	0.303	0.153	0.083
	Range	< 0.805 - 3.072 >	< 0.739 - 3.503 >	< 0.943 - 4.728 >	< 1.069 - 3.363 >	< 0.739 - 4.728 >
495 - 1,207	Mean	2.088	1.991	2.115	2.313	2.139
	SE	0.158	0.152	0.215	0.214	0.102
	Range	< 1.218 - 3.587 >	< 1.037 - 2.894 >	< 1.117 - 6.853 >	< 1.347 - 4.977 >	< 1.037 - 6.853 >
1,208 - 21,260	Mean	2.812	2.683	3.188	3.359	3.173
	SE	0.364	0.330	0.208	0.260	0.142
	Range	< 1.943 - 4.098 >	< 1.267 - 3.871 >	< 1.767 - 6.101 >	< 1.408 - 6.730 >	< 1.267 - 6.730 >
Total	Mean	1.538	1.686	2.437	2.479	2.035
	SE	0.088	0.089	0.146	0.155	0.066
	Range	< 0.581 - 4.098 >	< 0.634 - 3.871 >	< 0.782 - 6.853 >	< 0.262 - 6.730 >	< 0.262 - 6.853 >

Data Source: 1996 MEPS HC-003





shown in table 4, the mean design effect was 1.538 for proportional estimates less than 0.451, increasing to 1.686 for the proportional range 0.451 - 0.621, 2.437 for the proportional range 0.621-0.725, and measured at 2.479 for proportional range in excess of 0.725.

Summary

The study findings revealed that the original average design effect target for the 1996 MEPS for mean estimates of hospital stays, emergency room visits, prescribed medicines, home health provider days, and provider visits generally were satisfied.

Overall, for the selected health care utilization and insurance coverage measures, the average design effects are the same between gender but varied significantly between race/ethnicity groups (Hispanics/blacks vs. others), age categories (<45 years old vs. >= 45 years old), and Census regions.

Positive incremental effects on the average design effect were observed in relation to sample size for all the selected variables and to the domain estimates for private insurance coverage in particular. One possible explanation for the relationship between average design effect and proportion of population with private insurance coverage is that the ultimate cluster units in the 1996 MEPS HC sample design are the household or family. It is expected that a strong positive correlation exists between individuals in the same household with respect to their insurance coverage.

References

Cohen, S. B. (1982) "Comparison of Design Effect and Relative Variance Curve Strategy for Variance Estimation from Complex Survey Data" presented to the annual meetings of the American Public Health Association, November 1982.

Cohen J. W. (1997). "Design and methods of the 1996 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 S. B. (1997) "Sample design of the 1996 Medical Expenditure Panel Survey Household Component". Rockville (MD): Agency for Health Care Policy and Research; 1997. MEPS methodology Report No. 2. AHCPR Pub. No. 97-0027.

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