

Accounting for Clustering in the Analysis of MEPS Data: Frequently Asked Questions (FAQ)

1: When analyzing MEPS annual data, what needs to be done to account for intra-family clustering that exists because all family members living in sampled households are included in the survey?

In general, software developed for analyzing complex survey data take care of clustering/correlations at different levels such as within family, within household, within segments, and so on. All that needs to be done is to properly specify the MEPS stratum and PSU design variables (VARSTR and VARPSU respectively) or to use the BRR replicates provided which are developed based on VARSTR and VARPSU. See *MEPS Software Examples of Programming Codes* ([PDF 288KB](#)) or [HTML](#). These procedures use the robust variance estimator (see *R. L. Williams "A Note on Robust Variance Estimation for Cluster-Correlated Data" (June 2000), Biometrics*) which correctly accounts for all correlations that are nested below the first-stage sampling level (i.e., PSU), including intra-family correlation. In summary, clustering at all stages of sample selection is accounted for once the primary sampling unit (VARPSU) is specified.

2: When analyzing pooled MEPS data from two consecutive annual files, what needs to be done to account for the fact there are multiple observations for some sample persons due to the MEPS overlapping panel design?

While it is true that having repeated observations for the same person introduces another level of clustering, the additional clustering is also accounted for by specifying the MEPS stratum and PSU design variables (VARSTR and VARPSU, respectively) as discussed in FAQ 1 above. As for analyses of individual MEPS annual files, all that needs to be done is to properly specify the MEPS stratum and PSU design variables (VARSTR and VARPSU respectively) in a complex survey analysis statistical software procedure. See *MEPS Software Examples of Programming Codes* ([PDF 288KB](#)) or [HTML](#). These procedures use the robust variance estimator (see *R. L. Williams "A Note on Robust Variance Estimation for Cluster-Correlated Data" (June 2000), Biometrics*) which correctly accounts for all correlations that are nested below the first-stage sampling level (i.e., PSU), including both intra-family and intra-person correlation. In summary, clustering at all stages of sample selection is accounted for once the primary sampling unit (VARPSU) is specified.

3: Is it necessary to analyze MEPS data using hierarchical (multilevel) models because the complex sample design involves clustering at multiple stages of selection?

Given that special procedures and software exist for complex survey data analysis, it is not necessary to analyze MEPS data using hierarchical (multilevel) models. Hierarchical or multilevel models are often used in analyzing clustered-correlated data to account for correlation and, in some cases, to estimate effects at different levels of clustering. However, the theoretical development of hierarchical modeling is mostly limited to a classical framework where the data are assumed to be generated by the underlying multilevel model and not based on a complex survey design. Applying hierarchical modeling to complex survey data involves additional complexities due to stratification, clustering, unequal selection

Accounting for Clustering in the Analysis of MEPS Data: Frequently Asked Questions (FAQ)

probability, sampling weight, etc. Complex survey data are commonly analyzed using either the Taylor series approach or some form of replication method. These procedures, as implemented in different software packages, can be used to produce common types of estimates from complex survey data (e.g., variances of means, proportions, ratios and totals) or to fit linear or logistic models. Producing effects at different levels of clustering is usually not of interest when analyzing complex survey data. Expert statistical consultation is recommended for hierarchical modeling of complex survey data.