FAMILY UNIT ANALYSIS IN THE NATIONAL MEDICAL CARE EXPENDITURE SURVEY

Steven B. Cohen, National Center for Health Services Research

Introduction

The National Medical Care Expenditure Survey (NMCES) has been established to provide an assessment of the utilization, costs, and sources of payment associated with medical care in the United States. The complex survey design, which is a stratified multistage national area probability design, allowed for the analysis of family health characteristics by defining the household as the data collection unit. These households, which are subsets of the housing units,[1] consist of individuals related by blood, marriage or adoption. Unrelated individuals residing in the same housing unit were treated as distinct single member households. Data collection was applied to the same panel of sample households in six rounds of interviewing with 1977 as the reference period.

In the data collection process, the households were more formally defined as reporting units. The purpose of this mode of data collection was to obtain health expenditure, utilization and insurance coverage information on households with related individuals. These reporting units were further classified as either primary or secondary reporting units. Unrelated individual units include single lodgers or roomers unrelated to the primary family occupants of a housing unit, single person households, and members of group quarters (a noninstitutional civilan structure in which five or more unrelated persons reside). Each housing unit was assigned a primary reporting unit which, in the case of a housing unit with multiple reporting units, was generally the reporting unit with the largest number of related individuals. Any non-primary reporting unit in a multiple reporting unit setting was designated as a secondary reporting unit.

The reporting unit data collection scheme, entwined in a panel design, provided an initial framework from which to formally develop a family unit analytical file. More detailed rules were provided to characterize family continuation, dissolution, or formation, which evolved from the initial reporting unit specifications. A weighting scheme was then devised to yield unbiased national estimates of the health experiences for a dynamic population of families. In this paper, the framework for family unit analysis that was adopted for NMCES is presented.

The Reporting Unit Model

Consideration of the reporting unit model to characterize families would allow families to continue their existence when movement in or out of the household occurred, provided at least one key participant remained in the family. Key participants were defined as persons who were initially selected in the household study, in addition to babies born into sample households during the study and household members who were initially ineligible (i.e. who entered the military or an institution) but later became eligible by returning to their household. In this setting, the potential exists for undercoverage of the set of individuals initially ineligible for NMCES as a consequence of their membership in the military or an institution and who do not belong to any existing family when they rejoin the civilian household population. With respect to the non-key individuals, their health care data were collected for the period of time they were in reporting units which contained other key members.

To monitor the formation and dissolution of families in NMCES, the survey instruments provided for the retrieval of information at the start of each round concerning change in the family composition. One question elicited information on individuals no longer in the original reporting unit, the reasons for the change and the date. The new address was obtained for those members who moved, a new reporting was assigned to all members of the original household that moved to the new address and an attempt was made to establish contact. Another question probed to ascertain any additions to the original reporting units.

The process of reporting unit "splits" is a function of movement out of an intact household, where the spawned family is treated as new. More specifically, consider an original round one reporting unit consisting of a head of household, a spouse and three children (all key participants). By the subsequent round, the oldest child has moved out to marry. In the process, the original reporting unit continues its existence and a new reporting unit will be spawned.

In the NMCES, there were 13,955 initial reporting units identified in the first round of data collection with a questionnaire response. Over five rounds of data collection, 14,789 unique reporting units were identified. This constitutes a 5.97 percent increase in the reporting unit population. There were 11,653 reporting units with the same composition of participants for all rounds of data collection, which accounted for 79 percent of all reporting units or 86 percent of the reporting units existing for all five rounds. There were 13,514 reporting units that existed for all five rounds of data collection in which 14 percent, or 1,861 had a change in composition. Another 1,275 reporting units were in existence for less than five rounds, due to creations or dissolutions during the study. The results of this tabulation is presented in Table I.

The distribution of reporting units by patterns of attrition or additions of family members can be viewed in Table II. There were 1,114 reporting units with attrition across rounds, 1,774 reporting units with additions across rounds and 248 reporting units which manifested both attrition and

additions over the year. As indicated, 79 percent of all reporting units experienced no change in composition. Of the 40,323 NMCES participants, 39,196 belonged to only one reporting unit over five rounds of interviewing. There were another 1,061 participants who belonged to two distinct reporting units, and 66 participants with links to three reporting units over the year.

The reporting unit structure allows for the derivation of national estimates of household characteristics, where households are consistent with the definitions that identify reporting units. In the estimation component, reporting unit weights were initially defined in terms of the reciprocal of their selection probabilities. Since the primary and secondary reporting units owe their creation to their linkage with the respective housing units selected for inclusion in the NMCES, they retain the housing unit's probability of selection. Reporting units that are formed by splits of an existing reporting unit retain the sampling weight of the original unit. In those rare circumstances when two reporting units merge, the average of their sampling weights can be used to characterize the weight of the newly formed unit.

Family Unit Analysis

The analytical considerations of the proposed family unit analysis in the NMCES could not conform to the restrictive definitions of households or families that were implicit in the reporting unit model. Since many of the analyses focused on the health utilization, expenditure, and insurance coverage data at the family level, which were assumed to be quite sensitive to changes in family composition, an alternative family unit analysis strategy was required. A family was defined as specified in the reporting unit definition, consisting of one or more individual related by blood, marriage or adoption. The difference in family unit and reporting unit definitions was manifest in the treatment of the formation and dissolutions of families. Two new families were to be formed in the family unit framework when the head or spouse moved out of the existing family unit. In this process, the original spawning family ceased to exist. Whenever there was a change in the head or spouse due to death, movement out of the original family followed by nonresponse, institutionalization, or movement into military, only one new family was assigned. As before, the original family ceased to exist. For changes in family composition concerning family members other than head or spouse, such as birth, death, movement out, nonresponse of some family unit member, or a member institutionalized or in military, the family did not change. The composition of most family units were unaffected by change for the survey year, as one expected from trends in the general population. Those families that exhibited transformations allowed for the generation of dynamic estimates of family unit health experiences over the survey year.

Similar to the reporting units, the family units were affected by the undercoverage of the set of individuals initially ineligible for NMCES as a consequence of their membership in the military or an institution who do not belong to any existing family when they rejoined the civilian household population. Another source of potential bias was present as a consequence of a failure on the part of the NMCES questionnaire to collect information on family deletions that occured between January 1, 1977 and the round one interview date. To adjust for the impact of this limitation, health care data were collected for individuals who died in 1978 prior to the round five interview date. These deaths were assigned to round one reporting units as a function of the characteristics of the surviving families from which they departed.

The Family Unit Weighting Strategy

The derivation of national estimates of health care parameters for families in a panel design requires the appropriate formulation of sampling weights. The population of family units existing on any one day during the calendar year 1977 is potentially different from that existing on any other day due to the definition of family formation and dissolution. The families are defined with a beginning date, an ending date and a set of participants in the household survey that qualify as eligible members of the family. Consequently, there is a linkage between potential round one NMCES survey participants and the universe of all family units ever containing one or more of the potential participants during 1977. To characterize this linkage between unique families and potential round one participants (Bentley and Folsom, 1981), g(ij) terms are defined as zero-one indicators such that

g (ij) = 1 if the participant ever belonged to Family Unit j in 1977,
= 0 otherwise.

The total number of potential round one NMCES participants belonging to family unit j is defined as the family unit's multiplicity and specified in terms of membership indicators as g(+j), where

$$g(+j) = \sum_{i=1}^{I} g(ij)$$

To determine the total number of family units existing on day t, the existence indicators E_t (j) were defined, where,

$E_t(j)$ = 1 if the family unit j exists on day t
= 0 otherwise

Consequently, the sum of these existence indicators,

$$F(t) = \sum_{j=1}^{J} E_t(j)$$

over all family units J, provides a count of the total number of family units existing on day t. This can be re-expressed using a multiplicity framework such that

$$F(t) = \sum_{i=1}^{I} \sum_{j=1}^{J} g(ij)\, E_t(j)/g(+j)$$

This expression is now defined in terms of participant level quantities. Since the NMCES family units are distinguished by their composition of sample participants, the sample weight for a family can be formulated to incorporate the weights of its members. The individual's unadjusted sampling weight is the reciprocal of its selection probability. This selection is equivalent to the probability of selection for the sample housing unit within which the sample participant resides. These sampling weights were further adjusted to reflect the nonresponse of entire reporting units and smoothed to reduce the effect of unequal weighting (Cohen and Kalsbeek, 1977). Within this framework, an unbiased estimate of the number of family units existing on day t, F(t) can be expressed as

$$F(t) = \sum_{i \epsilon S} W_2(i) \sum_{j=1}^{S'} \frac{g(ij)\, E_t(j)}{g(+j)}$$

where $W_2(i)$ denotes the nonresponse adjusted, smoothed and contractor adjusted weight for the round one sampling unit that contained participant i, i S considers all eligible participants listed in the rosters of round one reporting units, and S' represents all sampled NMCES families.

In this setting, the family unit weights are defined as

$$W_f(j) = \sum_{i \epsilon S} W_2(i) g(ij)/g(+j)$$

$$= \sum_{i \epsilon S(j)} W_2(i)/g(+j)$$

where S(j) includes all eligible household participants found in responding round one reporting units, who at one time during 1977 belonged to family unit j.

If we defined R(j) as the set of responding round one reporting units linked to participants in S(j) and $n_j(K)$ as the number of eligible participants from reporting unit K in R(j), then the appropriate weight for family unit j can be re-expressed as

$$W_f(j) = \sum_{K \epsilon R(j)} n_j(k) W_2(k)/g(+j)$$

where $W_2(K) = W_2(i)$ for all i in K, $n_j(K)$ is the number of Key participants belonging to family unit j from reporting unit K, excluding births and previously ineligible military or institutionalized participants, and g(+j) is the sum of all Key and non-Key participants belonging to family unit j, excluding the same individuals excluded in $n_j(K)$.

The unbiased estimate of the number of family units existing on day t can be re-expressed as

$$F(t) = \sum_{j \epsilon S'} W_f(j) E_t(j)$$

where S' represents the set of all sampled NMCES families.

To arrive at annual and quarterly estimates of the number of families existing during the time period, the daily average number of families existing during the time interval is calculated. With days numbered from t=1 to t=365, the annual daily average can be expressed as

$$\sum_{t=1}^{365} F(t)/365 = \sum_{j \epsilon S'} W_f(j) \sum_{t=1}^{365} E_t(j)/365$$

The sum of existence indicators can be computed as

$$\sum_{t=1}^{365} E_t(j) = \text{Min}\left\{\left[90, ED(j)\right] - \text{Max}\left[1, BD(j)\right]\right\} + 1$$

where BD(j) is the beginning date and ED(j) is the ending date for family j.

Using q=1,2,3, or 4 to denote the quarterly time periods and q=5 to represent the annual period, and $PE_q(j)$ as the fraction of days during time interval q that family unit j exists, then the appropriate weight for estimating the average number of families existing per day during time q takes the form:

$$W_{f2q}(j) = W_f(j) \cdot PE_q(j)$$

On occasion, the number of days that data were provided by a family was not equivalent to the number of days for which it was eligible to report data. A weighting class adjustment was considered to adjust for the nonresponse induced by this difference between eligibility and responding days. Weighting classes were formed by crossing the following family level variables :

Size and type of family
- 1-4: Husband and wife with size=2,3,4, and 5+,
- 5-8: Female head - no husband present- size= 1,2,3, and 4+,
- 9-10: Male head - no wife present - size= 1, and 2+

Race of Head
- 1: White
- 2: Black or Other

Age of Head
1. Less than 35
2. 35-44
3. 45-64
4. 65+

It was assumed that the specified weighting classes would adequately capture most of the significant variation in response rates.

To derive the nonresponse adjustment factor, the variable $PR_q(j)$ was specified to indicate the fraction of days during time period q that family j had at least one responding participant. Weights of the form

$$W_{f3q}(j) = W_f(j) \cdot PR_q(j)$$

were then produced and incorporated into the nonresponse adjustment factor $AF_q(d)$ for time periods q=1,2,...5 where

$$AF_q(d) = \frac{\sum_{j \epsilon d} W_{f2q}(j)}{\sum_{j \epsilon d} W_{f3q}(j)}$$

and d=1,2...80 represents the eighty respective weighting classes. When fewer than 20 responding families were identified in a weighting class or $AF_q(d) \geq 2$, the classes were collapsed to satisfy these constraints. Family unit weights were specified as a function of this nonresponse adjusted factor, and took the form

$$W_q(j) = W_{f3q}(j) \cdot AF_q(d).$$

These weights were designed to be used directly in the derivation of family estimates for the nation that are expressed in terms of population totals on percents.

The evaluation of quarterly and annual family health care utilization and expenditure rates required application of both the $W_f(j)$ and $W_q(j)$ family weights. Consequently, the mean family utilization or expenditure estimate, Rq, took form

$$R_q = \frac{\sum_{j \epsilon S'} W_f(j)\, Y(j)}{\sum_{j \epsilon S'} W_q(j)}$$

where q = 1,2,3,4, or 5 for the respective quarter of 1977 or the year(q=5), and Y(j) is the j^{th} family's utilization or expenditure data for the time period of interest.

A Comparison of Demographic Measures Between Stable and Dynamic Families

A critical analytical concern of the family level data analysis was the determination of those characteristics that distinguished families experiencing a change in composition over the survey year and those remaining intact. Consequently, a comparison of demographic measures which characterize these distinct populations was considered. From preliminary NMCES family level data, it was possible to classify the sample of family units into two mutually exclusive groups which identified the stable and dynamic types. More specifically, the stable group consisted of families with no change in membership over the year, but allowing for the nonresponse of individual members. The dynamic group consisted of families with at least one member belonging to more than one family. In the NMCES survey, 86 percent of the families were classified as stable.

The demographic measurements on families that were considered included the age of the family head, the education level of the family head, the employment status of the head, the health status of the head, the martial status of the head, poverty level status of the head, region, size of city, sex, and race of the family head.

The respective classes for each of these demographic measures are:

Age	Years of Education
1. ≤18	1. 0-8 years
2. 19-24	2. 9-11 years
3. 25-54	3. 12 years
4. 55-64	4. 13-15 years
5. 65 or older	5. 16-18 years

Employment Status	Health Status
1. Worked	1. Excellent
2. Unemployed	2. Good
3. Not in labor force	3. Fair
	4. Poor

Martial Status	Poverty Level Status
1. Never married	1. Poor, below poverty level
2. Married	2. Near poverty level
3. Widowed	3. Low income
4. Separated	4. Middle income
5. Divorced	5. High Income

Region	Size of City
1. Northeast	1. 16 largest SMSA
2. North Central	2. SMSA ≥ 500,000 but not in 1
3. South	3. SMSA less than 500,000
4. West	4. Not SMSA, less than 60% rural
	5. Not SMSA, 60% or more rural

Sex	Race
1. Male	1. White
2. Female	2. Black or other

Parameter estimates of these demographic distributions which characterize the alternative family types can be observed in Table 3. To determine whether the estimated parameters of the distributions were equivalent across family types, a test of homogeneity appropriate for data from a complex survey design was considered. Once the respective

vectors of estimated proportions characterizing each demographic distribution and their associated variance-covariance matrices were specified for the stable and dynamic families, the weighted least squares methodology described by Grizzle, Starmer, and Koch (1969) and implemented in the analysis of data from complex surveys by Koch, Freeman, and Freeman (1975) and Freeman et al. (1976), was used to test for the equivalence in parameter estimates across family types. This method is analogous to multiple regression analysis using weighted least squares to estimate model parameters and test relevant hypotheses. Because sample sizes are relatively large for the respective estimated parameters, it can be assumed they have an approximate multivariate normal distribution. Consequently, statistical inferences can be made for model parameters by computing Wald statistics which have approximate chi-square distributions. Variances of all the estimated population parameters considered in this paper were derived using the Taylor series linearization method (Woodruff, 1971).

The age of head distributions for the stable and dynamic populations differed significantly at the .05 level (Table 3). From NMCES preliminary data, the difference was most notable across family types within age of head categories 19-24 and 65 or over. The stable families exhibited a significantly greater proportion of family heads 65 or older than in the dynamic group. This supports the notion that the stable families have a greater concentration of older families than their dynamic counterparts. Contrarily, the dynamic families had a greater representation of family heads in the younger age categories. These families are composed of younger individuals more likely to experience the effects of marriage, divorce, graduation from college or assumption of new employment.

The marital status distribution which characterized family heads also differed significantly across family types. This differential was primarily explained by the greater proportionate representation of never married family heads characterizing the dynamic population, as contrasted by the larger relative frequency of families with widowed heads in the stable group. In addition, a significantly greater representation of families with female heads was observed in the dynamic population.

It was also noted that the employment status distributions determined by the family head differed significantly between the stable and dynamic families. The stable families exhibited a greater proportional representation of family heads not in the labor force. Again, these families are more likely to consist of older heads. Comparisons of the population distributions characterized by region and size of city measures revealed no significant differences across family types. Consequently, there was no observed disproportionate concentration of either family group in a particular geographic setting, whether distinguished by region or metropolitan and non-metropolitan breakdowns. Similarly, no significant differences between the stable and dynamic families were noted when examining distributions characterized by the family head's years of education, health status, poverty level status, and race.

Summary

In this paper, a framework for family unit analysis is presented which allows for the derivation of national estimates of family health characteristics. Particular attention is given to the adopted weighting strategy and resultant estimates for a set of relevant demographic measures which characterize the families. The weighting strategy considered a multiplicity framework to yield annual and quarterly estimates of the average number of families existing during the period, in addition to national estimates of relevant NHCES health care measures.

An analysis to determine demographic characteristics that distinguished between stable and dynamic families was also presented. Findings revealed significant differences across families by age, marital status, sex and employment status of the head. Further analyses are planned to examine the differentials in health care utilization, expenditure and insurance coverage between the stable and dynamic families.

FOOTNOTE

1. A housing unit is defined as a house, apartment, group of rooms, or a single room which is occupied (or vacant but intended for occupancy) as a separate living quarters.

REFERENCES

1. Cohen, S.B. and Gridley G. (1981). "Effects of Alternative Weighting Strategies On Households As Sampling Units in the National Medical Care Expenditure Survey." Presented to the annual meetings of the American Public Health Association and available from the National Center for Health Services Research.
2. Cohen, S.B. and Kalsbeek, W.D. (1981). National Medical Care Expenditure Survey; Estimation and Sampling Variances in the Household Survey. National Center for Health Services Research, Instruments and Procedures Series, No. 2, Department of Health and Human Services Publication No. (PHS) 81-3281. Washington: Government Printing Office.
3. Bentley, B.S., and Folsom, R.E. (1981). "Family Unit Analysis Weighting Methodology for the National Medical Care Expenditure Survey." Research Triangle Institute, Research Triangle Park, NC.
4. Freeman, D.H., Jr., Freeman, J.L., Brock, D.B., and Koch, G.G. (1976). "Strategies in the Multivariate Analysis of Data from Complex Surveys II: An Application to the United States

National Health Interview Survey." International Statistical Review. 44, 317-330.

5. Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). "Analysis of Categorical Data by Linear Models." Biometrics. 25, 489-504.
6. Koch, G.G., Freeman, D.H., Jr., and Freeman, J.L. (1975). "Strategies in the Multivariate Analysis of Data from Complex Surveys." International Statistical Review. 43, 59-78.
7. Woodruff, L.S. (1971). "A Simple Method for Approximating the Variance of a Complicated Estimate." Journal of the American Statistical Association. 66, 411-414.

ACKNOWLEDGEMENTS

The author wishes to thank Barbara S. Bentley and Ralph E. Folsom for their valuable assistance in the determination of the family unit weighting strategy, and John Carrick for his skillful typing of this manuscript.

The views expressed in this paper are those of the author, and no official endorsement by the National Center for Health Services Research is intended or should be inferred.

Special thanks to Vicki Burt for her invaluable assistance in the statistical analyses.

TABLE I: Reporting Unit Distribution

	Number	% All RUID's
Reporting units with the same composition of participants for all 5 Rounds	11,653	78.8
Reporting units with a changing composition of participants which had at least one participants the same in all 5 Rounds	1,861	12.6
Reporting units existing all 5 Rounds	13,514	
Reporting units existing less than 5 Rounds (either created or destroyed during study)	1,275	8.6
Total Reporting units in existence at least during some period of the study	14,789	100%

TABLE II: Distribution of Reporting Units by Patterns of Attrition and/or Addition

	Number	%
Number of Reporting units with the same composition of PID's for all 5 Rounds	11,653	78.8
Reporting units with the attrition of one or more participants over 5 Rounds	1,114	7.5
Reporting units with the addition of one or more participants over the 5 Rounds	1,774	12.0
Reporting units with both the attrition and addition of one or more participants over the 5 Rounds	248	1.7
Total	14,789	100%

Table III: A comparison of Demographic Measures Between the Stable and Dynamic NMCES Families

Demographic measure	Stable families: Proportion	Stable families: Standard error	Dynamic families: Proportion	Dynamic families: Standard error
Age of Head				
1. < 18	.001	.000	.012	.002
2. 19-24	.061	.003	.150	.009
3. 25-54	.538	.007	.537	.020
4. 55-64	.170	.004	.167	.017
5. 65 or older	.230	.007	.133	.012
		Q=112.243 (5 degrees of freedom, P<.001)*		
Years of Education				
1. 0-8	.224	.008	.226	.017
2. 9-11	.160	.004	.155	.014
3. 12	.315	.006	.317	.018
4. 13-15	.137	.004	.157	.016
5. 16-18	.164	.007	.145	.015
		Q=1.611 (4 degrees of freedom, P=.807)*		
Employment Status				
1. Worked	.660	.007	.695	.017
2. Unemployed	.022	.002	.029	.004
3. Not in labor force	.317	.007	.275	.016
		Q=6.074 (2 degrees of freedom, P=.048)*		
Health Status				
1. Excellent	.429	.007	.406	.018
2. Good	.378	.006	.382	.016
3. Fair	.143	.005	.155	.011
4. Poor	.051	.003	.057	.006
		Q=2.459 (3 degrees of freedom, P=.483)*		
Martial Status				
1. Never married	.101	.005	.162	.013
2. Married	.641	.007	.611	.016
3. Widowed	.143	.004	.084	.009
4. Separated	.037	.003	.046	.007
5. Divorced	.077	.003	.095	.009
		Q=59.414 (4 degrees of freedom, P<.001)*		
Poverty Level Status				
1. Poor, below poverty level	.123	.004	.125	.010
2. Near povery level	.051	.002	.039	.006
3. Low income	.142	.004	.140	.011
4. Middle income	.351	.005	.335	.014
5. High income	.333	.008	.362	.017
		Q=4.796 (4 degrees of freedom, P=.3089)*		
Region				
1. Northeast	.221	.015	.185	.024
2. North Central	.262	.017	.290	.026
3. South	.315	.024	.334	.030
4. West	.202	.012	.191	.016
		Q=2.675 (3 degrees of freedom, P=.445)*		
Size of City				
1. 16 largest SMSA	.257	.021	.232	.030
2. SMSA> 500,000 but no⁻ in 1	.263	.030	.274	.036
3. SMSA<500,000	.180	.031	.188	.035
4. Not SMSA, less than 60% or more rural	.181	.022	.188	.026
5. Not SMSA, 60% or more rural	.118	.020	.118	.022
		Q=.593 (4 degrees of freedom, P=.964)*		
Sex				
1. Male	.749	.006	.702	.012
2. Female	.251	.006	.298	.012
		Q=12.272 (1 degree of freedom, P=.0005)*		
Race				
1. White	.882	.008	.874	.017
2. Black and other	.118	.008	.126	.017
		Q=.181 (1 degree of freedom, P=.670)*		

*Testing for homogeneity in distributions across family types and determining the P-value for the Wald Statistic, Q, by comparison to a X^2 distribution with the appropriate degrees of freedom.