MEPS

Methodology Report #19

Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey

AHRQ Advancing Excellence in Health Care

Agency for Healthcare Research and Quality
U.S. Department of Health & Human Services

March 2007

ABSTRACT

In the Medical Expenditure Panel Survey (MEPS), expenditures are defined as payments from all sources (including individuals, private insurance, Medicare, Medicaid, and other sources) for health care services during the year. Data on expenditures are collected for sample persons in the Household Component of the survey and from a sample of their health care providers responding to the Medical Provider Component of the survey. In the absence of payment information from either component, expenditure data are completed through weighted hot-deck imputation procedures. The MEPS collects a wide variety of data about individuals and health care events that are correlated with expenditures and, for each event type (e.g., doctor visits, hospitalizations, etc.), a selected set of these variables is used in the imputation processes. Several hot-deck iterations are run for each medical event type category based on factors such as whether partial payment information was reported and whether payments for the event covered multiple visits. This paper provides an overview of the methodological approach to impute MEPS expenditure data and how class variables for the hot-deck procedures were determined.

Suggested Citation:
Machlin, S. R. and Dougherty, D. D. *Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey.* Methodology Report No. 19. March 2007. Agency for Healthcare Research and Quality, Rockville, Md. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr19/mr19.pdf

* * *

The estimates in this report are based on the most recent data available at the time the report was written. However, selected elements of MEPS data may be revised on the basis of additional analyses, which could result in slightly different estimates from those shown here. Please check the MEPS Web site for the most current file releases.

Center for Financing, Access, and Cost Trends
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
www.meps.ahrq.gov

The Medical Expenditure Panel Survey (MEPS)

Background

The Medical Expenditure Panel Survey (MEPS) is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS is cosponsored by the Agency for Healthcare Research and Quality (AHRQ), formerly the Agency for Health Care Policy and Research, and the National Center for Health Statistics (NCHS).

MEPS comprises three component surveys: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC). The HC is the core survey, and it forms the basis for the MPC sample and part of the IC sample. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

MEPS is the third in a series of national probability surveys conducted by AHRQ on the financing and use of medical care in the United States. The National Medical Care Expenditure Survey (NMCES) was conducted in 1977, the National Medical Expenditure Survey (NMES) in 1987. Beginning in 1996, MEPS continues this series with design enhancements and efficiencies that provide a more current data resource to capture the changing dynamics of the health care delivery and insurance system.

The design efficiencies incorporated into MEPS are in accordance with the Department of Health and Human Services (DHHS) Survey Integration Plan of June 1995, which focused on consolidating DHHS surveys, achieving cost efficiencies, reducing respondent burden, and enhancing analytical capacities. To accommodate these goals, new MEPS design features include linkage with the National Health Interview Survey (NHIS), from which the sample for the MEPS-HC is drawn, and enhanced longitudinal data collection for core survey components. The MEPS-HC augments NHIS by selecting a sample of NHIS respondents, collecting additional data on their health care expenditures, and linking these data with additional information collected from the respondents' medical providers, employers, and insurance providers.

Household Component

The 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 HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a two and a half year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two 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

combined with other ongoing panels, will provide continuous and current estimates of health care expenditures.

The sampling frame for the MEPS-HC is drawn from respondents to NHIS, conducted by NCHS. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized population, with oversampling of Hispanics and blacks.

Medical Provider Component

The MEPS-MPC supplements and validates information on medical care events reported in the MEPS-HC by contacting medical providers and pharmacies identified by household respondents. The MPC sample includes all hospitals, hospital physicians, home health agencies, and pharmacies reported in the HC. Also included in the MPC are all office-based physicians:

- Providing care for HC respondents receiving Medicaid.
- Associated with a 75 percent sample of households receiving care through an HMO (health maintenance organization) or managed care plan.
- Associated with a 25 percent sample of the remaining households. Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents, including:
- Diagnoses coded according to ICD-9 (9th Revision, International Classification of Diseases) and DSMIV (Fourth Edition, Diagnostic and Statistical Manual of Mental Disorders).
- Physician procedure codes classified by CPT-4 (Current Procedural Terminology, Version 4).
- Inpatient stay codes classified by DRG (diagnosis related group).
- Prescriptions coded by national drug code (NDC), medication names, strength, and quantity dispensed.
- Charges, payments, and the reasons for any difference between charges and payments.

The MPC is conducted through telephone interviews and mailed survey materials.

Insurance Component

The MEPS-IC collects data on health insurance plans obtained through private and public sector employers. Data obtained in the IC include the number and types of private insurance plans offered, benefits associated with these plans, premiums, contributions by employers and employees, and employer characteristics.

Establishments participating in the MEPS-IC are selected through three sampling frames:

- A list of employers or other insurance providers identified by MEPS-HC respondents who report having private health insurance at the Round 1 interview.
- A Bureau of the Census list frame of private-sector business establishments.
- The Census of Governments from the Bureau of the Census.

To provide an integrated picture of health insurance, data collected from the first sampling frame (employers and other insurance providers) are linked back to data provided by the MEPS-HC respondents. Data from the other three sampling frames are collected to provide annual national and State estimates of the supply of private health

insurance available to American workers and to evaluate policy issues pertaining to health insurance. Since 2000, the Bureau of Economic Analysis has used national estimates of employer contributions to group health insurance from the MEPS-IC in the computation of Gross Domestic Product (GDP).

The MEPS-IC is an annual panel survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone follow-up for nonrespondents.

Survey Management

MEPS data are collected under the authority of the Public Health Service Act. They are edited and published in accordance with the confidentiality provisions of this act and the Privacy Act. NCHS provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports and microdata files. Summary reports are released as printed documents and electronic files. Microdata files are released on CD-ROM and/or as electronic files.

Printed documents and CD-ROMs are available through the AHRQ Publications Clearinghouse. Write or call:

AHRQ Publications Clearinghouse
Attn: (publication number)
P.O. Box 8547 Silver Spring, MD 20907
800-358-9295
703-437-2078 (callers outside the United States only)
888-586-6340 (toll-free TDD service; hearing impaired only)

To order online, send an e-mail to: ahrqpubs@ahrq.gov.

Be sure to specify the AHRQ number of the document or CD-ROM you are requesting. Selected electronic files are available through the Internet on the MEPS Web site: http://www.meps.ahrq.gov/

For more information, visit the MEPS Web site or e-mail mepspd@ahrq.gov.

Table of Contents

Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey

Steven R. Machlin, Agency for Healthcare Research and Quality, and Deborah D. Dougherty, Westat

Introduction

The Medical Expenditure Panel Survey (MEPS) is a complex national probability survey of the U.S. civilian noninstitutionalized population, and has been conducted on an annual basis since 1996 by the Agency for Healthcare Research and Quality (AHRQ). One of the primary purposes of the survey is to collect data that can be used to analyze national medical expenditures (i.e., the amount paid for health care services).

Unfortunately, it is difficult to obtain complete information on medical expenditures from household survey respondents because the type of information being collected is often not straightforward and requires extensive record keeping over time, especially for households with members who frequently use the health care system. Further, in a significant number of instances, respondents are simply not aware of either the total amount billed or how much the provider is paid for the services that were received. Classic examples are individuals enrolled in the Medicaid program, where financial transactions occur only between the provider and the state Medicaid agency, and enrollees of managed care plans or HMOs who only may be aware of paying some predetermined co-payment that is not necessarily related to the total amount the provider receives (Cohen et al., 1997).

As a consequence of these factors, there is a substantial amount of item nonresponse on medical expenses in the Household Component (HC) of MEPS. To compensate for these missing data and to improve accuracy, data on expenses for sample persons are also collected from a sample of their health care providers in the Medical Provider Component (MPC) of MEPS (see description of MPC under MEPS Expenditure Estimation Strategy below). However, expense data are not available from either survey component for a noteworthy proportion of medical events reported in the survey (e.g., roughly one-third in 2001).

A weighted hot-deck approach is used to impute missing expenditure data in MEPS. This approach uses other survey responses to complete missing data and incorporates survey weights to replicate the weighted distribution of the available data in the imputed data (Cox, 1980). The objectives of the imputations are to create data sets for analysis that preserve sample sizes and reduce the potential for nonresponse bias in analyses of MEPS expenditure data. This paper provides a general overview of the MEPS expenditure imputation process.

MEPS Sample Design

The sample of households for the MEPS-HC is a subsample of households that responded to the prior year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (National Center for Health Statistics, 2002). The MEPS sample is drawn from approximately half of the primary sampling units (PSUs) selected for the NHIS. For example, the 1996 MEPS-HC sample was selected from

households that responded to the 1995 NHIS (Cohen S., 1997). This selection was comprised of 195 PSUs and 1,675 sample segments (second-stage sampling units). Over sampling of households with Hispanics and blacks carries over from the NHIS to the MEPS sample design.

The sample design of the Medical Expenditure Panel Survey is an overlapping panel design, with data collected for each new MEPS panel covering a two-year period (Cohen J., 1997). As a result of the overlapping panel design, MEPS annual data for 1997 and beyond are constructed based on data collected from two consecutive panels.

MEPS Expenditures Defined

Total medical expenditures in MEPS are defined as the sum of direct payments for care provided during the year, including out-of-pocket payments and payments by third-party payers (e.g., private insurance, Medicare, Medicaid, and other sources), rather than the amount billed by the provider for the care provided (i.e., charges). Payments for hospital and physician services, ambulatory physician and nonphysician services, prescribed medicines, home health services, dental services, and various other medical equipment and services that were purchased or rented during the year are included. Payments for over-the-counter drugs and phone contacts with providers are not collected in MEPS.

Provider charges for health care are not considered a proxy for payments, primarily due to two important trends that have occurred since the mid 1990s (Zuvekas and Cohen, 2002). First, pressure to contain health care costs by employers has increased insurers' leverage to negotiate substantial discounts with providers. Second, the insurance industry made significant movement toward capitation as a way of increasing the incentive for providers to contain costs by being subjected to financial risk for high levels of utilization. As a result, for a sizeable number of medical events, charges have become virtually meaningless as a measure of payments. Nevertheless, charges are collected in MEPS because they are highly correlated with payments and are incorporated in the imputation process for missing expenditure data wherever possible (see Example 3 below).

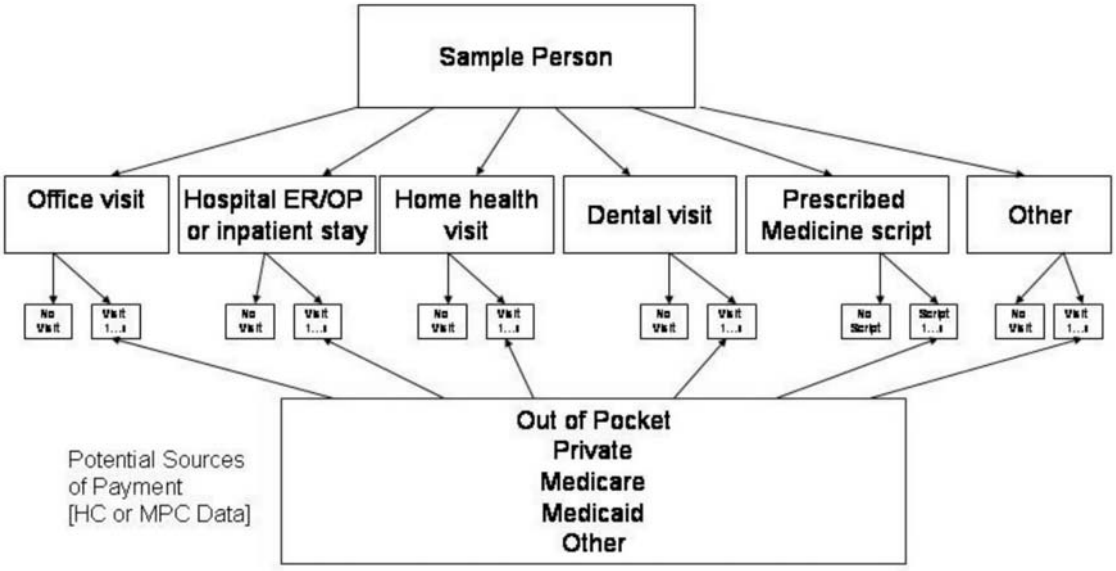
MEPS Household Expenditure Data Collection

Primary data collection in the MEPS-HC employs computer-assisted personal interviewing (CAPI). The HC questionnaire is designed to collect use and expenditure data for two consecutive years through a series of five interviews. In general, annual health care utilization and expenses for sample persons are derived from information collected in three of the five interviews (Cohen J., 1997).

Figure 1 provides a pictorial summary of the data collection process for medical events and expenses in MEPS. For each person in a sample household, the core instrument collects detailed data about medical care received as well as charges and payments for each health care event reported in the utilization section. Medical events reported are grouped into the following categories: office-based medical provider visits, hospital emergency room visits, hospital outpatient visits, hospital inpatient stays, dental visits, home health, prescribed medicines, and other medical expenses. Payments for each event are itemized according to the following 10 source of payment categories: out of pocket, Medicare, Medicaid, private insurance, Veteran's Administration, TRICARE, Other

Federal sources, Other State and local sources, Workers' Compensation, and Other unclassified sources. Payments for a particular medical event can be made across one or a combination of sources (though total payments for a small proportion of events each year are considered to be $0, which occurs when it is reported that no payments were or will be made). Total expenses for a given event are obtained by summing across all payment sources.

Figure 1. Illustration of collection of medical event and source of payment data: MEPS



Nonresponse on payments for a particular medical event may occur for any potential payment source. However, it is not unusual for respondents to report the amount paid out of pocket and that a third-party source(s) paid an unknown amount (i.e., partial item nonresponse).

MEPS Expenditure Estimation Strategy

In addition to the HC, MEPS expenditure data are also collected in the Medical Provider Component (MPC) of the survey. The purpose of the MPC is to collect data directly from a sample of medical providers to reduce the level of missing data and to improve the accuracy of expenditure estimates that would be obtained by relying solely on household responses (Machlin and Taylor, 2000, and Cohen J. et al., 1997). Data from the MPC are considered to be more accurate on average than comparable data reported by household respondents in the HC.

Data obtained in the MPC are linked to medical events reported in the HC based on a probabilistic matching procedure (Winglee et al., 1999). As a consequence of the matching process, each medical event reported in the HC will have expense data from both the HC and MPC, one of these sources, or neither source (i.e., complete missing payment data). A hierarchical approach is used to develop complete data for expenditures as follows: 1) start with household reported medical events, 2) use MPC expense data where available, 3) use HC expense data if no MPC data available, and 4) impute any missing information. Table 1 shows the distribution by source of expenditure

data (i.e., HC, MPC, or imputed) in 2001 for each type of event category, and the subsequent discussion provides an overview of the imputation process.

Table 1. Distribution of source of expenditure data for survey-reported health care events, by type of service, 2001 MEPS

		Hospital events				
	Office visits	**Outpatient visits**	**Emergency room visits**	**Inpatient stays**	**Dental visits[1]**	**Home health[2]**
Number of events	142, 793	15, 763	5, 904	3, 405	26,438	3,155
	Percent distribution by source of data[3]					
Total	100.0	100.0	100.0	100.0	100.0	100.0
MPC	27.9	46.7	47.9	61.4	--	42.3
HC	17.5	6.2	8.1	3.7	47.1	9.4
Imputed: Partial[4]	19.2	8.2	9.7	4.9	11.8	0.1
Imputed: Full	35.3	38.9	34.3	30.0	41.1	48.2

[1]Dental care providers are not surveyed in the MEPS Medical Provider Component, so MPC category is not applicable.
[2]Expense data for home health are collected on a monthly rather than a per visit basis.
[3]Percentages for office visits do not add to exactly 100.0 due to rounding.
[4]Includes events where expense information was imputed for some but not all payment sources.

Imputation Process

Separate imputations are conducted for each event type category because relevant variables and statistically significant correlates of expenditures vary by type of event. However, insurance coverage is utilized for all imputations regardless of event type because generosity of payments is associated with type of coverage. For example, Medicaid payments are typically less generous than private insurance payments for comparable services.

Missing expenditure data for health care events reported in the survey are completed through a weighted hot-deck imputation procedure (Cox, 1980), with data from the MPC used as the primary donor source wherever possible. In general, the hot-deck procedure sorts donor events (complete data) and recipient events (missing data) into imputation cells based on important predictors of expenses available in MEPS. For example, the imputation procedure for hospital inpatient events sorts donors and recipients into cells based on insurance coverage of the sample person, number of nights in the hospital, reason for hospitalization, whether the hospital admission immediately followed an emergency room visit, as well as region and urbanization level of the person's residence. Whenever possible, a donor is selected within the same cell as a recipient to complete a recipient record. However, if there are fewer donors than recipients in a cell, cells are collapsed in a predetermined order until a 1:1 ratio of donors to recipients is achieved. In general, the order used for cell collapsing is determined based on the relative strength of the associations between the classification variables and expenses.

Imputations are handled somewhat differently depending on 1) whether all or some potential sources of payment are missing and 2) whether the total charge for the event was reported or not. Following are examples of three different scenarios for imputation of hospital inpatient expenses. These examples assume that donors and recipients match on the pertinent correlates of expenditures (e.g., insurance coverage, number of nights in the hospital, reason for hospitalization, whether the hospital admission immediately followed an emergency room visit, region, and urbanization).

Example 1. Complete imputation

Payment source	Donor	Recipient (pre-imputation)	Recipient (post-imputation)
Medicare	$1,840	Missing	$1,840
Private insurance	$792	Missing	$792
Total expenses	$2,632	--	$2,632

In Example 1, it was reported that a sample person had a hospital inpatient stay and was covered by Medicare and private insurance but the respondent did not know the amount paid by either source for that stay. The donor record that was selected for this recipient in the hot-deck procedure was an inpatient stay where the hospital was paid a total of $2,632, of which $1,840 was from Medicare and $792 was from a supplemental private insurance policy. These identical values were imputed to the recipient record.

Example 2. Partial imputation

Payment source	Donor	Recipient (pre-imputation)	Recipient (post-imputation)
Out of pocket	$26	$5	$5
Private insurance	$971	Missing	$992
Total expenses	$997	--	$997

In Example 2, it was reported that a sample person had an inpatient hospitalization, was covered by private insurance, and that $5 was paid out of pocket but the respondent did not know the amount paid to the hospital by private insurance. The donor record that was selected for this recipient in the hot-deck procedure was an inpatient stay where the hospital was paid a total of $997, of which $26 was paid out of pocket and $971 was from private insurance. In this situation, the total amount paid for the event from the donor ($997) was imputed to the recipient record, the reported out-of-pocket amount ($5) was retained, and the difference ($992) was imputed to the recipient record as a private insurance payment.

Example 3. Imputation using total charge

Payment source	Donor	Recipient (pre-imputation)	Recipient (post-imputation)
Total charges	$5,171	$4,173	$4,173
Total expenses	$4,248	missing	$3,421
Medicare	$3,411	missing	$2,737
Private insurance	$837	missing	$684

As described earlier (see section on MEPS Expenditures Defined), charges are not identical to but are highly correlated with expenditures (payments) made for health care. In most instances, when there are missing data on payments for a health event reported in the survey there are also missing data on charges. However, in situations where the respondent reports the total charge for an event but does not know the actual payments, the reported information on charges is used to improve the accuracy of the imputation.

To illustrate the use of total charge information when available, in Example 3 the respondent reported there was $4,173 in hospital facility charges for the reported inpatient stay. The donor record selected for the imputation in the hot-deck procedure showed $5,171 in total charges and $4,248 in total expenses. The first step imputes total expenses to the recipient record by applying the ratio of total expenses to total charges

on the donor record (4,248/5,171) to the total charges on the recipient record ($4,173). Then, the imputed total expense on the recipient record ($3,421) is allocated across the two potential sources of payment, Medicare and private insurance, in the same proportion as on the donor record (i.e., 837/4,248 and 3411/4,248 for Medicare and private insurance, respectively).

Summary

MEPS is an ongoing survey that collects data on the utilization and expenditures for health care in the U.S. civilian noninstitutionalized population. Given the complexity of the U.S. health care system and the wide range of public and private financing arrangements, it is difficult to collect complete information on health care expenses.

To maximize the completeness and accuracy of expenditure data, MEPS integrates data on utilization and expenditures from the Household Component of the survey with data from a sample of providers that participate in the Medical Provider Component of the survey. To complete medical expenditure data that were not obtained from either component, a weighted hot-deck imputation procedure is used. The primary advantage of this procedure is that the distribution of data values (including the imputed ones) will look similar to the distribution of the values in the population (Korn and Graubard, 1999).

The hot-deck procedures used to complete missing expenditure data in MEPS are based on statistical as well as substantive considerations regarding the U.S. health care financing system. For example, type of health insurance coverage is used as an auxiliary variable in the imputations for all health service type categories because of differences in average payments between insured and uninsured persons as well as varying generosity of payments by type of insurance coverage. In contrast, length of stay is incorporated as a classification variable in the hot deck only for inpatient stays because it is significantly associated with expenditures for hospital inpatient stays, but is irrelevant when imputing expenses for other types of health care events.

In summary, the dual objectives of imputing missing expenditure data in MEPS are to maximize sample sizes available for analysis and to reduce the risk of nonresponse bias associated with exclusion of cases with missing data. However, the imputation approach used is inherently complex, resource intensive, and leads to underestimation of variances for survey estimates without an additional correction. While it is difficult to assess the impact of imputation on variances, the Center for Financing, Access, and Cost Trends at AHRQ is currently conducting methodological research to estimate the magnitude of the impact. Results of a preliminary investigation of the impact of the expenditure imputations in MEPS have been reported (Baskin, 2004).

References

Baskin, R., Wun, L., Sommers, J., et al. Investigation of the impact of imputation on variance estimation in the Medical Expenditure Panel Survey. *American Statistical Association 2004 Proceedings.* 2004.

Cohen, J. *Design and Methods of the Medical Expenditure Panel Survey Household Component.* MEPS Methodology Report No. 1. AHCPR Pub. No. 97-0026. Rockville, Md.: Agency for Health Care Policy and Research, 1997.

Cohen, J., Monheit, A., Cohen, S., et al. The Medical Expenditure Panel Survey: A National Health Information Resource. *Inquiry,* 33: 373-389 (Winter 1996/97).

Cohen, S. *Sample Design of the 1996 Medical Expenditure Panel Survey Household Component.* MEPS Methodology Report No. 2. AHCPR Pub. No. 97-0027. Rockville, Md.: Agency for Health Care Policy and Research, 1997.

Cox, B. The weighted sequential hot deck imputation procedure. *American Statistical Association 1980 Proceedings of the Section on Survey Research Methods,* 721-726. 1980.

Korn, E. and Graubard, B. *Analysis of Health Surveys.* Wiley Series in Probability and Statistics. 1999.

Machlin, S. and Taylor, A. *Design, Methods, and Field Results of the 1996 Medical Expenditure Panel Survey Medical Provider Component.* MEPS Methodology Report No.9. AHRQ Pub. No.00-0028. Rockville, Md.: Agency for Healthcare Research and Quality, 2000.

National Center for Health Statistics. *Health, United States,* 2002 (361-362). Hyattsville, Md.: 2002.

Winglee, M., Valliant, R., Brick, M., and Machlin, S. Probability matching of medical events. *Journal of Economic and Social Measurement* 23 (1999) 1-12.

Zuvekas, S. H. and Cohen, J. W. A guide to comparing health care expenditures in the 1996 MEPS to the 1987 NMES. *Inquiry,* Spring 2002; 39:76-86.

Acknowledgments

The authors wish to thank Trena Ezzati-Rice, Joel Cohen, and Steven Cohen for their helpful reviews of the paper.