The Utility of the Integrated Design of the Medical Expenditure Panel Survey to Inform Mortality Related Studies

Steven B. Cohen
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

Abstract

The analytic capacity of surveys can be dramatically enhanced through the linkage to existing secondary data sources at higher levels of aggregation as well as through direct matches to additional health and socio-economic measures acquired for the same set of sample units from other sources of survey specific or administrative data. In this presentation, the capacity of an integrated survey design to enhance longitudinal analyses focused on mortality studies is discussed. Examples are drawn from the Medical Expenditure Panel Survey (MEPS), designed to produce estimates of health care utilization, expenditures, sources of payment, and insurance coverage of the US population. Analyses are conducted to examine differentials in pre-dispositional factors that distinguish a representative sample of decedents from their surviving counterparts and to assess the relationship between antecedent health and health care related factors and mortality. The relationship between medical expenditure levels over time and mortality is also examined to illustrate the enhanced set of longitudinal analyses that can be undertaken.

The views expressed are those of the author and no official endorsement by the Agency for Healthcare Research and Quality or the Department of Health and Human Services is intended or should be inferred. All tables can be obtained from the author upon request.

Key Words: Integrated designs, MEPS, mortality studies

1. Introduction

The analytic capacity of surveys can be dramatically enhanced through the linkage to existing secondary data sources at higher levels of aggregation as well as through direct matches to additional health and socio-economic measures acquired for the same set of sample units from other sources of survey specific or administrative data. In this paper, the capacity of one specific integrated survey design to enhance longitudinal analyses focused on mortality studies is discussed. Examples are drawn from the Medical Expenditure Panel Survey (MEPS), an ongoing longitudinal panel survey designed to produce estimates of health care utilization, expenditures, sources of payment, and insurance coverage of the U.S. civilian non-institutionalized population. The first set of analyses are conducted to examine the differentials in pre-dispositional factors that distinguish a representative sample of decedents from their surviving counterparts. Particular attention is given to the capacity to distinguish the health characteristics and the health care experiences of a representative sample of decedents for a time period prior to their deaths. This is followed by a more extensive model-based study to assess the relationship between antecedent health and health care related factors and mortality. The relationship between medical expenditure levels over time and mortality is also examined to illustrate the enhanced set of longitudinal analyses that are possible through this framework. The longitudinal analyses that are highlighted are based on linkages of the MEPS to the National Health Interview Survey and the National Death Index.

2. Background

The MEPS was designed to provide annual and longitudinal estimates at the national level of the health care utilization, expenditures, sources of payment and health insurance coverage for the U.S. civilian non-institutionalized population. The MEPS consists of a family of interrelated surveys, which include a Household Component (HC) and a Medical Provider Component (MPC). In addition to collecting data to yield annual estimates for a variety of measures related to health care use and expenditures, MEPS provides estimates of measures related to health status, demographic characteristics, employment and access to health care1-2. Estimates can be provided for individuals, families and population subgroups of interest. The data collected in this ongoing longitudinal study also permit studies of the determinants of the use of services and expenditures, and changes in the provision of health care in relation to social and demographic factors such as employment or income; the health status and satisfaction with health care of individuals and families; and the health needs of specific population groups such as the elderly and children.

The MEPS research program, broadly defined to encompass data collection, data development, research and the translation of research into practice, is directly tied to the strategic goal of identifying strategies to improve access, foster appropriate use and reduce unnecessary expenditures. Few other surveys provide the foundation for estimating the impact of changes on different economic groups or special populations of interest, such as the poor, elderly, veterans, the uninsured and racial/ethnic groups. The public sector relies upon the MEPS research findings to evaluate health reform policies, the effect of tax code changes on health expenditures and tax revenue, and proposed changes in government health programs such as Medicare. In the private sector, these data are also used to develop economic projections3-4.

The set of households selected for the MEPS HC is a subsample of those participating in the National Health Interview Survey (NHIS), an ongoing annual household survey of approximately 35,000 households (85,000 individuals) conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention, to obtain national estimates of health care utilization, health conditions, health status, insurance coverage and access. The MEPS HC consists of an overlapping panel design in which any given sample panel is interviewed a total of 5 times in person over 30 months to yield annual use and expenditure data for two calendar years. These rounds of interviewing are spaced about 5 to 6 months apart. The interview is administered through a computer assisted personal interview (CAPI) mode of data collection, and takes place with a family respondent who reports for him/herself and for other family members. Initiated in 1996, the 2011 MEPS annual survey consists of approximately 14,000 families and 33,000 individuals, and reflects an oversample of the following policy relevant population subgroups: Hispanics, blacks and Asians. Data from two panels are combined to produce estimates for each calendar year5-7.

To facilitate the conduct of longitudinal cohort analyses using the NHIS and MEPS data in tandem, NHIS/MEPS linkage files have been developed. These NHIS/MEPS linkage files allow users to link persons in the MEPS public use files to the records of the same persons in the previous NHIS public use files. In addition, the documentation clarifies that only a subset of persons from the previous NHIS core person, sample adult and sample child files will match to a subset of the MEPS file. Examples of enhanced longitudinal analyses based on the NHIS-MEPS linked files include studies of the long term uninsured, assessments of prediction models to target

individuals with persistently high expenditures over time, and use of NHIS sociodemographic factors to oversample policy relevant subpopulations into the MEPS. In addition, other longitudinal studies based on the MEPS-NHIS linkage include cohort analyses of individuals with specific medical conditions and the conduct of episodes of illness studies over an extended time interval8-11. Building on this capacity, this study further enhances the capacity for the conduct of longitudinal analyses augmenting the MEPS-NHIS linkages with further matches to the National Death Index to inform mortality related studies. In this paper, attention will be given to enhancing an understanding of the data linkage process, to articulating an estimation strategy to permit longitudinal analyses, and the development of the necessary estimation weight12. In addition, this study provides several illustrative examples of the capacity to conduct health outcome studies that are designed to investigate the association between a set of health and health care related factors with mortality.

3. Building the Analytical File linking MEPS data to the NHIS and the National Death Index

The National Center for Health Statistics (NCHS) has developed a record linkage program designed to enhance the analytic capacity of their population-based surveys. As part of this effort, NCHS links several of its surveys with death certificate records from the National Death Index (NDI). The National Death Index (NDI) is a central computerized index of death record information on file in the State vital statistics offices13. Working in concert with these State offices, NCHS established the NDI as a resource to aid health and medical investigators and researchers with their mortality related efforts. The mortality linkage of the National Health Interview Survey (NHIS) to death certificate data found in the National Death Index (NDI) has been completed for survey years 1986 through 2004. The updated NHIS Linked Mortality Files provide mortality follow-up data from the date of NHIS interview through December 31, 2006. Mortality ascertainment is based primarily upon the results from a probabilistic match between NHIS and NDI death certificate records. There are two versions of the NHIS Linked Mortality Files: public-use files that include a limited set of mortality variables for adult NHIS participants and restricted-use files that include more detailed mortality information and mortality follow-up for children. Information on accessing the NHIS data files linked to the National Death Index is provided on the following weblink14: http://www.cdc.gov/nchs/data_access/data_linkage/mortality/nhis_linkage_public_use.htm

To illustrate the process of file linkage, the 2000 NHIS linked to the NDI will be considered using the publically available files. The linked file consists of 100,618 survey participants. The NDI program will permit a match to be listed if any of the following seven conditions are satisfied: 1) social security number; 2) Exact month and +/- one year of birth, first and last name; 3) exact month and +/- one year of birth, first and middle initials, last name; 4) Exact month and day of birth, first and last name; 5) Exact month and day of birth, first and middle initials, last name; 6) Exact month and year of birth, first name, father's surname; 7) if the subject is female: exact month and year of birth, first name, last name (user's record) and father's surname (NDI record). Of the 100,618 records on the 2000 NHIS-NDI file, 64,514 records (64.12%) are eligible for NDI linkage; 28,495 (28.32%) are not considered for linkage to the NDI on the publically available files since they are under age 18; and 7,609 (7.56%) are ineligible for linkages

based on problematic matching criteria. Of those 2000 NHIS survey participants eligible for linkage to the NDI, 4,032 of the 64,514 records linked to the NDI were determined to be deceased from the time of the NHIS interview through December 31, 2006.

The next step in the process is to link this 2000 NHIS –NDI file to the 2001 MEPS Public Use File which includes 22,701 MEPS sample participants (Panel 6) participating in the first year of the MEPS survey. Using the 2001 MEPS-2000 NHIS link files to facilitate linkage of the MEPS data to the NHIS-NDI files, 20,488 of the 22,701 FY persons can be found in the NHIS 2000 Mortality file. There are 2,213 of the 22,701 MEPS 2001 Panel 6 persons not found in the NHIS 2000 Mortality file. The vast majority of the non-linked cases (1,961) are new individuals in MEPS that have recently joined households that participated in the 2000 NHIS and not eligible for linkage. Only 252 eligible MEPS participants (204 aged 18 or older) could not be linked to the NHIS-NDI analytical file (approximately 1%). A distribution of the eligibility status of the 20,488 2001 MEPS sample participants linked to the 2000 NHIS-NDI file is provided in Table 1 (all tables available from the author on request).

Of the 13,433 linked cases in MEPS eligible for a link to the NHIS-NDI analytic file, 821 aged 18 and older were determined to be deceased from the time of the NHIS interview through December 31, 2006 (Table 2). Of the 130 decedents aged 18 and older in 2001 identified through the link to the National Death Index, 114 were also identified as decedents in 2001 in the MEPS survey. Of the 16 cases not identified as 2001 decedents in MEPS, 4 of the participants entered an institution during 2001 and another 6 were no longer in-scope for the MEPS survey at the end of 2002. Consequently, the level of inconsistency between 2001 MEPS reports of deaths and classification based on the NDI link was low, at less than 5 percent (6 out of 130).

Of the 20,488 MEPS 2001 sample participants linked to the 2000 NHIS-NDI file, 19,544 MEPS sample participants responded for their entire period of eligibility in MEPS over the course of the two years of the panel, 2001 through 2002. These cases were identified as the subset of MEPS participants with positive longitudinal estimation weights (LONGWTP6) on the 2001-2002 MEPS longitudinal file.

4. Developing Estimation Weights to Support Enhanced Longitudinal Analyses based on Linkages of the MEPS to the NHIS and the NDI

Summaries of these cross-sample MEPS-NHIS linkages and non-linkage for a typical year in the administration of these ongoing surveys has been discussed in prior analytical efforts focused on estimation issues that need to be addressed to permit cohort studies. These types of longitudinal cohort analyses treat the population at the time of the NHIS administration as a baseline, with follow-up data on their health care experiences, status and outcomes obtained from the MEPS. By their nature, cohort studies are restricted to those individuals measured at time t-1 in the NHIS that include subsequent observations at time t and t+1 in the MEPS. Separate estimation weights have already been developed to permit cohort analyses using health and health care related data acquired from the NHIS core interview and from the health condition-centric NHIS sample adult interview in concert with the MEPS data. The estimation strategy considered here to permit longitudinal analyses of MEPS data linked to the NHIS and the NDI follows this approach12.

More specifically, the following additional adjustment was implemented to the 2001 MEPS Panel 6 longitudinal analysis estimation weight LONGWTP6, for the 19,544 sample respondents also linked to the 2000 NHIS-NDI analytic file. The adjustment used the following variables in the specification of a raking adjustment to the 2000 population controls available from the NHIS 2000 person file (n=100,618, Population estimate=274,018,975): sex, Hispanic origin, Race (white only; black only; AIAN only; Asian only; other or multiple race), Age (0-4, 5--17, 18-29, 30-44, 45-64, 65+), Region, Health status (excellent; very good; good and DK; fair; poor) and health insurance coverage at the time of the interview (covered; not covered and DK). The variables used to implement the adjustment were the 2000 NHIS measures. The resultant estimation weights used for this NHIS-MEPS-National Death Index cohort analysis was referred to as NDIMEPS.

5. Determining the Pre-dispositional Characteristics of a Representative Sample of Decedents

One of the advantages of using the MEPS data in tandem with linkages to the National Death Index is the capacity to identify the health characteristics and the health care experiences of a representative sample of decedents for a time period prior to their deaths. To illustrate this capacity, attention is focused on a nationally representative cohort of adult decedents that died over the course of the five year interval 2002 through 2006 and were members of the civilian non-institutionalized population in years 2000 and 2001. Based on this new analytical data resource, the socio-demographic and health care characteristics of the representative sample of decedents can be determined and compared with those of their surviving counterparts.

Using data from the 2001 MEPS, the first set of comparisons were directed to examining the differentials in predispositional characteristics that distinguish the representative sample of decedents from their surviving counterparts. Prior studies of medical expenses in the final months of life have demonstrated that these expenditures are high and vary according to geographic region and patient characteristics15-16. A recent study of Medicare patients during the last 6 months of life revealed that higher expenditures were associated with declines in functional status, race and ethnicity, certain chronic conditions, and lack of nearby family support17-18. The linkage of the MEPS to the National Death Index permits the conduct of related analyses which look back over longer periods in time. Controlling for age, the mean medical care expenditures in 2001for the cohort of 2002-2006 decedents was significantly higher than their counterparts that survived through 2006 (Tables 3.1 and 3.2). Similarly, adult decedents over the period 2002-2006 were significantly more likely to be represented in the top 10 percent of the health care expenditure distribution in 2001 relative to those alive at the end of 2006. In terms of hospitalizations, the representative sample of decedents under study was also significantly more likely to have experienced an in-patient stay in 2001. (Tables 3.1 and 3.2)

When examining a set of measures that assess the health status of the population, the five year representative sample of decedents in years 2002-2006 were consistently more likely to be in fair or poor health status in 2001, to have significantly lower physical component functioning scores as measured by the Medical Outcomes Study SF-1219, to have lower self assessed ratings of their health based on the EuroQol-5D20, and to have

more chronic conditions on average in 2001 than their counterparts who were alive at the end of the same five year period (Tables 3.1 and 3.2).

The Household Component of the Medical Expenditure Panel Survey (MEPSHC) also contains a series of self-administered questions that discern individual attitudes regarding their health preferences. Adults age 18 and over are asked whether they strongly agree, agree, are uncertain, disagree, or strongly disagree with each of the following statements: "I'm more likely to take risks than the average person." and "I can overcome illness without help from a medically trained person". For this analysis, strongly agreed and agreed responses were combined into an "agreed" category. When significant differences in risk taking behavior were observed, individuals who were decedents in 2002-2006 were both less likely to take risks and less likely to believe they could overcome illness without medical help in 2001 than individuals alive on 12/31/2006 (Table 3).

With respect to socio-demographic characteristics, when differences were observed, the representative sample of decedents in 2002-2006 were more likely to be older, male, white non-Hispanic, widowed, residing in non-MSA areas and in the southern region of the country, and to have not completed high school in 2001 than their counterparts that survived the five year interval (data not shown). They were also significantly more likely to be poor, and for those aged 18-64, more likely to be publically insured in 2001 than individuals alive at the end of 2006 (Table 3).

Since 2000, MEPS respondents have been asked to provide a summary rating of all the health care they received in the last 12 months from all their doctors and other health providers. Using a scale from 0 to 10 where 0 indicates the worst health care possible and 10 represents the best health care possible, the MEPS self-administered questionnaire permits the survey participants to provide this quality of care assessment. When controlling for age, no significant differentials in the ratings of the health care received in 2001 distinguished the representative sample of decedents in 2002-2006 from their counterparts.

6. Determinants of Mortality for a Representative Sample of Decedents

The linkage of the MEPS to the NDI also permits more extensive model based studies to assess the relationship between antecedent health and health care related factors and mortality. To demonstrate the capacity to identify a set of salient factors associated with greater probabilities of mortality over a five year period in the future, the following logistic regression model based on the 2001 MEPS linkage to the NDI was specified. The logistic model under consideration classified individuals who were identified as decedents in 2002-2006 through the NDI in addition to 19 deaths identified in the 2002 MEPS as Y=1, with all other individuals classified as Y=0. The predispositional variables included as potential correlates were based on an individual's 2001 profile. This modeling effort builds off related efforts that attempted to identify individual characteristics associated with a higher likelihood of incurring high levels of medical expenditures in the future, an outcome strongly correlated with the likelihood of morbidity and mortality. Based on prior studies that have assessed the relationship between an individual's characteristics at an initial time period, (year 1, t(1)) relative to their significant association with future year expenditures (year 2, t(2)), a detailed set of pre-dispositional factors were given consideration in this study. The measures under consideration included demographic, socio-economic and geographical characteristics,

health insurance coverage, health status and health conditions, health care utilization indices and total health care spending (Table 4). More specifically, the model included the following measures: gender; race/ethnicity; marital status (married, widowed, divorced, separated, never married (excluded category); level of education (less than high school, general education degree (GED), high school, college, advanced degree (excluded category); poverty status, insurance coverage (for those ages 18-64:full year uninsured, public only, some private (excluded category); for those ages 65 and older: Medicare only, Medicare + some private, Medicare + public only (excluded category)); health status(excellent/very good, good, fair/poor (excluded cat)) ; self-assessed ratings of their health based on the EuroQol-5D (0-10 scale); the number of chronic conditions they had; the presence of limitations in activity; the number of inpatient stays in 2001; the number of prescribed medicine purchases in 2001; the number of office based visits in 2001; and their total medical expenditures in 2001. Separate models were run for those ages 18-64 and for those ages 65 and older.

The final models excluded several of the above measures under consideration that were not determined to be significant factors when testing at the .05 level of significance. The standard errors of the survey estimates have been adjusted for the complex survey design of the MEPS and the test statistics used to test for statistical significance have also been adjusted to control for survey's complex multi-stage probability design with unequal weighting. When attention is directed to the results for adults under the age of 65 in 2001, many of the same predispositional factors associated with increased likelihood of future mortality over a five year interval in the bivariate analyses remained significant when controlling for other socio-economic and health care measures (Table 5.1). For those aged 18-64 in 2001, individuals who were males, black non-hispanic, not completing high school, having limitations in activity, with lower self assessments of their health state, having public only health insurance coverage, and having higher levels of health care expenditures were more likely to have a higher probability of death over the subsequent five year interval relative to their respective counterparts (Table 5.1). Similarly for those aged 65 and older in 2001, individuals who were males, black nonhispanic, having limitations in activity, in fair or poor health status, with lower self assessments of their health state, and having higher levels of medical expenditures were significantly more likely to have a higher probability of death over the subsequent five year interval relative to their respective counterparts (Table 5.2). The analysis revealed a general convergence in the set of predisposition measures identified as significant factors that distinguished an individual's likelihood of death of over a future five year interval both in direction and scale.

7. The Relationship Between Medical Expenditure Levels over Time and Mortality

In 2008, health care expenses among the U.S. community population totaled $1.15 trillion. Medical care expenses, however, are highly concentrated among a relatively small proportion of individuals in the community population21. As previously reported in 1996, the top 1 percent of the U.S. population accounted for 28 percent of the total health care expenditures and the top 5 percent for more than half. More recent data have revealed that over time there has been some decrease in the extent of this concentration at the upper tail of the expenditure distribution22. Furthermore, the top 10 percent of the population accounted for 65.2 percent of overall health care expenditures in 2007, and 42.7 percent of this subgroup retained this top decile ranking with respect to

their 2008 health care expenditures. Studies that examine the persistence of high levels of expenditures over time are essential to help discern the factors most likely to drive health care spending and the characteristics of the individuals who incur them. The MEPS-HC data are particularly well suited for measuring trends in concentration and persistence8. Alternatively, individuals ranked in the bottom half of the health care expenditure distribution accounted for only 3.0 percent of medical expenditures. Similar to the experience of the top half of the population based on their medical expenditure rankings, 74.6 percent of those in the lower half of the expenditure distribution retained this classification in 2008.

This analysis demonstrates the capacity to conduct enhanced longitudinal analysis with the MEPS data linked to the NDI. Using expenditure data from the 2001 and 2002 Medical Expenditure Panel Survey linked to the National Death Index, this study investigates the relationship of the persistence in medical expenditures over a two year interval and subsequent mortality between 2003-2006. Restricting the analysis to individuals aged 18 and older that survived through 2002, the population is divided into the following mutually exclusive and exhaustive groups: 1) top 10th percentile of medical expenditure distribution for both years; 2) bottom 50th percentile of medical expenditure distribution for both years; 3) top 10th -50th percentile of medical expenditures first year - top 10th percentile of medical expenditures in second year; 4) other. Based on these medical expenditure classifications for 2001-2002, an estimate of the probability of subsequent mortality between 2003 through 2006 is obtained (Table 6).

During the period 2001-2002, 3.7 percent of the adult population aged 18 or older were classified in the top 10th percentile of medical expenditure distribution for both years; 37.2 percent were classified in bottom 50th percentile of medical expenditure distribution for both years, and 4.9 percent were classified into the top 10th -50th percentile of medical expenditures first year and the top 10th percentile of medical expenditures in second year (Table 6). When examining the estimated probability of death during the period 2003-2006, a clear pattern in the relationship between the level of medical expenditures in the prior two year period and mortality. As expected, individuals in the top decile of the medical care expenditure distribution both years were the most likely group to die over the next four years (estimated likelihood of death=.222 (.019), Table 6). This relationship between expenditure levels and mortality was most pronounced for the elderly (estimated likelihood of death = .353 (.034)). Alternatively, adults in the bottom half of the health care expenditure distribution both years were most likely to survive through 2006 (likelihood of death=.012, Table 6). Individuals that were in the top half of the medical expenditure distribution, though not in the top decile in 2001 that experienced a shift into the top decile in the subsequent year also exhibited a higher likelihood of mortality relative to their counterparts with lower expenditures in both years. This incremental relationship between the level of medical expenditure spending and mortality held when the population was further controlled by age (ages 18-64 and the elderly ages 65+).

8. Summary

Promoting data collection coordination, integration and future alignment of surveys, administrative data and electronic health record systems is a core component of the Department of Health and Human Services data strategy. It recognizes the enhanced analytical capacity of surveys that may be achieved though connectivity to other surveys and administrative data sources23-26. This paper attempts to further advance this premise

by demonstrating the capacity of the integrated survey design of the Medical Expenditure Panel Survey to enhance longitudinal analyses focused on mortality studies. The longitudinal analyses that are highlighted are based on linkages of the MEPS to the National Health Interview Survey and the National Death Index. The process of accessing these publically available datasets and the steps involved in data linkage through application of cross-walk identifiers is also illustrated. An estimation strategy is also specified to support enhanced longitudinal analyses based on these newly linked and augmented analytic files. Particular attention is given to assessments of the relationship between antecedent health and health care related factors and mortality. The relationship between medical expenditure levels over time and mortality is also examined to illustrate the enhanced set of longitudinal analyses that are possible through this framework.

One recent study that focused on arthritis, occupational class, and the aging US workforce compared age- and occupational class–specific quality adjusted life years (QALYs) between workers with and without arthritis by merging data from the National Health Interview Survey, Medical Expenditure Panel Survey, and the National Death Index into a single analytic database27. By giving more visibility to these types of analytic studies and demonstrating the accessibility and the relative ease of data linkages across these invaluable national health and health care related data resources, the potential for enhanced analytic capacity through their integration should be more likely to be realized.

Acknowledgements

The author wishes to thank Zhengyi Fang, Social and Scientific Systems, Inc., for his excellent statistical programming support.

References

1.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 Healthcare Policy and Research, 1997. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr1/mr1.pdf

2.Machlin, SR and A.K. Taylor, AK.(2000). Design, methods, and field results of the 1996 Medical Expenditure Panel Survey Medical Provider Component. AHRQ Pub. 2000: 28.

3. Cohen, J., Cohen, S. and Banthin, J. The Medical Expenditure Panel Survey: A National Information Resource to Support Healthcare Cost Research and Inform Policy and Practice. Medical Care. 2009; 47(7): S1:44-50.

4. Cohen, S. *Design Strategies and Innovations in the Medical Expenditure Panel Survey*. Medical Care. 2003: 41(7) S3:5-12.

5.Ezzati-Rice, T.M., Rohde, F., Greenblatt, J. *Sample Design of the Medical Expenditure Panel Survey Household Component, 1998-2007*. Methodology Report No. 22. March 2008. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr22/mr22.pdf

6.Cohen, SB. Sample Design of the 1997 Medical Expenditure Panel Survey Household Component. AHRQ Pub. 1997-01.

7.Cohen, S.B. Sample Design of the 1996 Medical Expenditure Panel Survey: Medical Provider Component. The Journal of Economic and Social Measurement 1998; 26: 1-29.

8. Cohen, S. and Yu, W. *The Concentration and Persistence in the Level of Health Expenditures over Time: Estimates for the U.S. Population, 2007-2008*. Statistical Brief #309. December 2010. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st309/stat309.pdf
9. Cohen, S, Makuc, D and Ezzati-Rice, T. Health insurance coverage during a 24-month period: a comparison of estimates from two national health surveys. Journal of Health Services and Outcomes Research Methodology 2007; 7(3-4):125-144.
10. Cohen, S. Attitudes toward Health Insurance and Their Persistence over Time, Adults 2006-2007 "MEPS Statistical Brief #260 AHRQ 2009.
http://www.meps.ahrq.gov/mepsweb/data_files/publications/st260/stat260.pdf
11.Cohen, S. and Yu, W. The Utility of Prediction Models to Oversample the Long Term Uninsured. Medical Care, 2009; 47(1)1: 80-87.
12. Cohen, S. An Estimation Methodology to Permit Longitudinal Cohort Analyses Based on the National Health Interview Survey and Medical Expenditure Panel Survey". Agency for Healthcare Research and Quality Working Paper No. 11002, October 2010, http://gold.ahrq.gov.
13. NCHS. About the National Death Index [NCHS\CDC web site]. October 21, 2009. Available at http://www.cdc.gov/nchs/data_access/ndi/about_ndi.htm . Accessed August 3, 2011.
14. NCHS. NHIS Linked Mortality Public-use Files. [NCHS\CDC web site]. March 22, 2010. Available at http://www.cdc.gov/nchs/data_access/data_linkage/mortality/nhis_linkage_public_use.htm. Accessed August 3, 2011
15. Cohen, SB, Carlson, BL, and Potter, D. Health Care Expenditures in the Last 6 Months of Life. Health Policy Review. 1995; 1 (2), 1-13.
16. Cohen, SB, Carlson, BL. and Potter, D. Health Care Expenditures in the Last Six Months of Life. Report to the Senate Committee on Labor and Human Resources and to the House of Representatives Committee on Energy and Commerce. 1994. U.S.Department of Health and Human Services, Public Health Service.
17. Kelley AS, Ettner SL, Morrison RS, et al. Determinants of medical expenditures in the last 6 months of life. Ann Intern Med., 2011; 154(4):235-42.
18. Adler, G. A Profile of the Medicare Current Beneficiary Survey. Health Care Financing Review, 1994; 15(4): 153-163.
19. Hays, R., Sherbourne, C. and Mazel, R. User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life. Santa Monica, CA: RAND Corporation, 1995. http://www.rand.org/pubs/monograph_reports/MR162.
20. McCrone P, Patel A, Knapp M et al., A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia. J Ment Health Policy Econ; 2009. 12(1):27-31.
21. Machlin S, J. Cohen, S. Zuvekas et al. (2001). Health Care Expenses in the Community Population. AHRQ Pub. 2001:27
22. Yu, W. and Ezzati-Rice, T. *Concentration of Health Care Expenditures in the U.S.Civilian Noninstitutionalized Population*. Statistical Brief #81. May 2005. Agency for Healthcare Research and Quality, Rockville, MD.
http://www.meps.ahrq.gov/mepsweb/data_files/publications/st81/stat81.pdf
23. Cohen, S., Ezzati-Rice, T. and Yu, W. Integrated Survey Designs: A Framework for Nonresponse Bias Reduction. Journal of Economic and Social Measurement, 2005; 30 (2-3), 101-114.
24. Cohen, S. Integrated Health Care Survey Designs: Analytical enhancements achieved through linkage of surveys, administrative and secondary data. Statistical Policy Working Paper: Seminar on Achieving Statistical Quality in Diverse and Changing Environments.

2005. Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.
25. Moeller, J., Cohen, S. Mathiowetz, N. et al. Regression-Based Sampling for Persons with High Health Expenditures: Evaluating Accuracy and Yield with the 1997 MEPS. Medical Care. 2003. 4(7):44-52.
26. Cohen, S. and Wun, L. A Comparison of Household and Medical Provider Reported Health Care Utilization and Estimation Strategies to Correct for Response Error. Journal of Economic and Social Measurement. 2005. 30, (2-3), 115-126.
27. Caban-Martinez, A.J., Lee, D., Fleming, L. et al. Arthritis, Occupational Class, and the Aging US Workforce, American Journal of Public Health, 2009. http://ajph.aphapublications.org/cgi/reprint/AJPH.2011.300173v1

Table 3: 2001 Characteristics of a Representative sample of decedents in 2002-2006, aged 18 and older

Population Estimates

Age (Yrs)	Decedents in 2002-2006	Alive 12/31/2006
	Population (x1000)	Population (x1000)
18+	9,692	192,571
18-64	2,935	166,520
18-54	1,520	143,657
55-64	1,414	22,863
65+	6,758	26,051
65-79	3,885	21,377
80+	2,873	4,674

Medical Expenditures

Age (Yrs)	Decedents in 2002-2006		Alive 12/31/2006			
	Mean	S.E.	Mean	S.E.	Z	P
18+	10,800.37	919.46	2,661.89	60.88	8.8320	<0.0001
18-64	11,382.48	1,673.36	2,271.71	56.56	5.4415	<0.0001
18-54	8,501.32	1,732.49	2,005.52	57.84	3.7473	0.0004
55-64	14,479.45	2,994.49	3,944.31	188.29	3.5112	0.0008
65+	10,547.58	928.41	5,155.98	223.58	5.6459	<0.0001
65-79	10,370.54	929.31	4,953.06	254.73	5.6222	<0.0001
80+	10,786.98	1,715.58	6,084.10	636.73	2.5700	0.0147

In top 10th percentile of medical expenditures (Percent)

18+	38.76	2.09	8.56	0.28	14.3289	<0.0001
18-64	33.05	3.66	7.06	0.28	7.0878	<0.0001
18-54	30.06	5.43	6.03	0.30	4.4185	<0.0001
55-64	36.26	5.28	13.53	0.87	4.2511	<0.0001

65+	41.24	2.59	18.13	1.02	8.2869	<0.0001
65-79	39.78	3.17	17.47	1.01	6.6993	<0.0001
80+	43.21	4.43	21.14	2.97	4.1359	0.0001
Percent with inpatient stays						
18+	30.38	2.02	8.05	0.32	10.9100	<0.0001
18-64	27.86	2.99	7.05	0.30	6.9355	<0.0001
18-54	27.60	4.78	6.70	0.32	4.3605	<0.0001
55-64	28.14	4.27	9.29	0.84	4.3349	<0.0001
65+	31.47	2.60	14.42	1.10	6.0311	<0.0001
65-79	30.51	2.91	13.60	1.12	5.4247	<0.0001
80+	32.77	4.31	18.18	3.03	2.7686	0.0086

Table 4: Measures considered as potential correlates of likelihood of mortality over five year interval

Measures	**Description**
Age	Age at end of a year
Sex	Male, Female
Race/ethnicity	Hispanic, Black/not Hispanic, White and Other
Marital Status	Married, Widowed, Divorced/Separated, Never Married
Family Size	One, 2 or more
Family income classification	Poor, income at or below the federal poverty level; near poor, income over the poverty level through 125% of the poverty level; low income, over 125% through 200% of the poverty level; middle income, over 200% through 400% of the poverty level; high income, over 400% of the poverty level.
Health Insurance Coverage	For those ages 18-64:Full-year insured, part-year insured, uninsured For those ages 65 and older: Medicare Only; Medicare + some private; Medicare + public only
Health Status	Excellent, Very good, Good, Fair, Poor
Health Ratings	Self-assessed ratings of health based on the EuroQol-5D: Scale form 0-10
Limitation in Activity	Presence of activity limitation in work, housework, or employment; no limitation in activity
Health Conditions	Number of Chronic Conditions
Inpatient Events	Frequency in year
Number of Prescribed Medicine Purchases	Frequency in year
Number of Ambulatory Visits (office based)	Frequency in year
Total Health Care	Continuous measure for expenditures in 2001

Expenditures	

Source: 2001 Medical Expenditure Panel Survey Household Component. Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality

Table 5.1: Logistic Regression Model to Identify Factors Associated with Higher Mortality Rates in 2002-2006, based on 2001 Profiles (2001 MEPS, ages 18-64)

Independent Variables and Effects	Beta		P-value T-Test		Wald Statistic	
	Coeff.	SE Beta	B=0	Wald F	P-value	
Intercept	-8.52676	0.94993	<0.00001			
AGE	0.06506	0.01190	<0.00001	29.89	<0.0001	
SEX						
Male	0.62542	0.19482	0.00151	10.31	0.0015	
RACE/Ethnicity						
HISPANIC	-0.35578	0.54841	0.51715	3.07	0.0285	
BLACK, NH	0.56937	0.54492	0.29717			
WHITE, NH	0.23372	0.50675	0.64508			
MARITAL STATUS						
MARRIED	-0.31203	0.37567	0.40706	1.98	0.0989	
WIDOWED	0.38058	0.47629	0.42508			
DIVORCED	0.28600	0.40235	0.47791			
SEPARATED	-0.64644	0.63881	0.31262			
EDUCATION						
<HS	1.53487	0.52179	0.00360	5.45	0.0003	
GED	2.03297	0.57966	0.00054			
HS	0.95694	0.50888	0.06129			
College	0.58529	0.52070	0.26215			
POVERTY STATUS						
near poor	0.45025	0.35680	0.20824	0.90	0.4652	
low income	0.48951	0.31624	0.12301			
middle income	0.42316	0.28519	0.13922			
high income	0.51560	0.36895	0.16361			
HEALTH INSURANCE						
ANY PRIVATE (<65)	0.44460	0.31462	0.15896	8.07	0.0004	
PUBLIC ONLY (<65)	1.14949	0.29286	0.00011			
HEALTH STATUS						
Excellent/VG	-0.14339	0.36627	0.69580	0.10	0.9077	
Good	-0.05564	0.32050	0.86233			
HEALTH STATE EuroQol 5D	-0.01387	0.00527	0.00904	6.93	0.0090	
FAMILY SIZE						
1	-0.09792	0.26620	0.71334	0.14	0.7133	
ACTIVITY LIMITATION						
LIMITATION	0.77342	0.24533	0.00183	9.94	0.0018	
MEDICAL EXPENDITURES	0.00002	0.00001	0.00133	10.56	0.0013	

Sample size 10,263
Pseudo R-Square: 0.035541

-2 * Normalized Log-Likelihood with Intercepts Only	:	1687.98
-2 * Normalized Log-Likelihood Full Model	:	1316.58
Approximate Chi-Square (-2 * Log-L Ratio)	:	371.40
Degrees of Freedom	:	25

Source: 2001 Medical Expenditure Panel Survey, Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, 2000 National Health Interview Survey, NCHS/CDC, 2001-2006 National Death Index, NCHS/CDC

Table 5.2: Logistic Regression Model to Identify Factors Associated with Higher Mortality Rates in 2002-2006, based on 2001 Profiles (2001 MEPS, ages 65+)

Independent Variables and Effects	Beta		P-value T-Test		Wald Statistic
	Coeff.	SE Beta	B=0	Wald F	P-value
Intercept	-9.00874	1.21523	<0.00001		
AGE	0.09706	0.01361	<0.00001	32.35	<0.0001
SEX					
Male	1.10448	0.15860	<0.00001	13.45	<.0001
RACE/Ethnicity					
HISPANIC	0.35186	0.44666	0.43179	2.69	0.0472
BLACK, NH	0.93301	0.45761	0.04279		
WHITE, NH	0.82075	0.39443	0.03873		
MARITAL STATUS					
MARRIED	0.03707	0.41514	0.92893	1.44	0.2210
WIDOWED	0.62154	0.41508	0.13588		
DIVORCED	0.22043	0.49790	0.65846		
SEPARATED	0.07658	0.73569	0.91720		
EDUCATION					
<HS	0.10173	0.34445	0.76805	0.11	0.9793
GED	0.09401	0.49233	0.84877		
HS	0.14212	0.33498	0.67183		
College	0.20621	0.35838	0.56567		
POVERTY STATUS					
near poor	-0.01140	0.32466	0.97203	0.74	0.5678
low income	0.12673	0.21754	0.56086		
middle income	-0.09533	0.22032	0.66571		
high income	-0.17700	0.22763	0.43774		
HEALTH INSURANCE					
MEDICARE ONLY (65+)	-0.32823	0.22447	0.14526	1.15	0.3198
MEDICARE+PRV (65+)	-0.19532	0.23465	0.40620		
HEALTH STATUS					
Excellent/VG	-1.09155	0.21773	0.00000	13.10	<0.0001
Good	-0.71000	0.19599	0.00037		
HEALTH STATE EuroQol 5D	-0.01050	0.00418	0.01279	6.31	0.0128
FAMILY SIZE					
1	-0.30889	0.21647	0.15516	2.04	0.1552
ACTIVITY LIMITATION					
LIMITATION	0.49704	0.18878	0.00913	6.93	0.0091
MEDICAL EXPENDITURES	0.00002	0.00001	0.02829	4.88	0.0283

Sample size 2,050
Pseudo R-Square: 0.179176

-2 * Normalized Log-Likelihood with Intercepts Only : 2030.95
-2 * Normalized Log-Likelihood Full Model : 1626.18
Approximate Chi-Square (-2 * Log-L Ratio) : 404.76
Degrees of Freedom : 25

Source: 2001 Medical Expenditure Panel Survey, Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, 2000 National Health Interview Survey, NCHS/CDC, 2001-2006 National Death Index, NCHS/CDC

Table 6: Probability of Death in 2003-2006 Conditioned on Medical Expenditure Experience in 2001-2002

GROUP	N	Population	Probability of Death in 2003-2006	S.E.
AGECAT=Age 18+				
Overall	13,918	202,680,569	.0378	.0019
Top 10 % of expenditures both years	504	7,444,887	.2221	.0193
Bottom 50% both years	5,455	75,360,309	.0121	.0015
Top 10%-50% in Y01 and top 10% in Y02	680	9,906,070	.1243	.0169
Remaining persons	7,279	109,969,303	.0352	.0024
AGECAT=Age 18-64				
Overall	11,659	169,718,650	.0124	.0012
Top 10 % of expenditures both years	276	3,954,720	.1064	.0205
Bottom 50% both years	5,170	71,615,968	.0067	.0012
Top 10%-50% in Y01 and top 10% in Y02	405	5,762,923	.0414	.0095
Remaining persons	5,808	88,385,039	.0109	.0017
AGECAT=Age 65+				
Overall	2,259	32,961,919	.1687	.0094
Top 10% of expenditures both years	228	3,490,167	.3532	.0342
Bottom 50% both years	285	3,744,341	.1148	.0213
Top 10%-50% in Y01 and top 10% in Y02	275	4,143,147	.2396	.0349
Remaining persons	1,471	21,584,264	.1347	.0095

Note: Restricted to MEPS Panel 6 persons with AGE01X=18+ and NDIMEPS>0 who are alive as of 12/31/2002 and are in both 2001 full year and 2002 full year files

Source: 2001 and 2002 Medical Expenditure Panel Survey, Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, 2000 National Health Interview Survey, NCHS/CDC, 2001-2006 National Death Index, NCHS/CDC