An evaluation of the collection and production of medical condition estimates from the MEPS

Frances M. Chevarley[1]
[1]Agency for Healthcare Research and Quality, Department of Health and Human Services, 540 Gaither Road, Rockville, MD 20850

The views in this paper are those of the author and no official endorsement by the Department of Health and Human Services, or the Agency for Healthcare Research and Quality (AHRQ) is intended or should be inferred.

Abstract
The prevalence and costs of medical conditions are important to public health. This paper analyzes condition estimates from the Medical Expenditure Panel Survey (MEPS), a nationally representative panel survey studying health care use, access, expenditures, source of payment, insurance coverage, and quality of care. Each year a new panel begins and each panel has 5 rounds of data collection over 2 ½ years that covers a two-year period.

There are alternative ways to produce condition estimates based on MEPS data. The first is based on responses to condition-specific questions in the household CAPI instrument. Condition estimates associated with health events such as doctor visits (treated prevalence) can also be made. In addition, estimates can be made based on conditions that bothered the person and occurred between the rounds or since the last interview. The respondent along with other family members who may be present during the interview report the information.

Starting with Panel 12 (2007 Panel) in an attempt to reduce respondent burden and to improve the reporting of conditions, the enumeration of priority conditions was moved to the first round of data collection and earlier in the CAPI instrument. The questions are then asked in subsequent rounds in certain situations. Several other changes were made at this time in the data collection procedures/tabulations/editing. In addition, there was a change in the sample design in 2007. This paper focuses on the effects of changes to the priority condition-specific questions of the MEPS CAPI instrument.

Key Words: MEPS, conditions, CAPI, sample design, paradata

1. Introduction

The prevalence of and expenditures for medical conditions are important to public health. Studies using the Medical Expenditure Panel Survey (MEPS) have demonstrated high health care utilization and expenses for certain conditions and for persons with multiple chronic conditions.[1,2,3] Because of the importance of condition data in determining health care access, utilization, costs, and condition prevalence there have been a number of studies evaluating the quality of condition data reported on health surveys such as MEPS and the National Health Interview Survey (NHIS); a few references are provided.[4,5,6,7]

There are alternative ways of producing condition estimates using MEPS data: Responses to condition-specific questions in the household CAPI instrument; condition estimates associated with health events such as doctor visits; and conditions reported as bothering the person during the MEPS reference period. This paper focuses on the condition-specific questions asked in the CAPI instrument.

Changes were made to when and how often the priority condition-specific questions are asked starting with MEPS Panel 12 (2007 Panel) in order to: enhance the analytic utility of the survey; improve reporting of conditions associated with events; make the interview process smoother; and, improve the reporting of conditions using the condition-specific questions. The objective of this paper is to evaluate effects of changes to when and how often the priority condition-specific questions are asked starting with MEPS Panel 12 on burden and on estimates.

1.1 Background/Data

The Medical Expenditure Panel Survey (MEPS) is a nationally representative longitudinal survey that collects detailed information on health care utilization and expenditures, health insurance, and health status, as well as on a wide variety of social, demographic, and economic characteristics for the U.S. civilian noninstitutionalized population. MEPS's main sponsor is the Agency for Healthcare Research and Quality. The MEPS has three components—Household, Medical Provider, and Insurance. The MEPS Household component (MEPS-HC) uses the National Health Interview Survey as it's sampling frame; MEPS has an overlapping panel design—5 interviews over 2 ½ years covering a 2-year reference period. Rounds 1, 3 and 5 are fielded in the first half of the year and rounds 2 and 4 are fielded in the second half of the year. Data for a full year file is based on data for rounds 1-3 of the panel that began that year and data for rounds 3-5 of the panel that began the year before.

The data used in this paper are from the following MEPS-HC files: the 2006, 2007, and 2008 MEPS full-year consolidated files; and the 2006 MEPS panel 11 longitudinal file. Also used are data from the CDC Vital and Health Statistics Series 10, Numbers 235, 242 (NHIS) reports.[8,9]

1.2 Condition Questions in MEPS

Specific high prevalent "ever" condition questions have been asked in MEPS since 2000. The questions begin with "Has (PERSON) ever been told by a doctor or other health professional that (PERSON) had..." The conditions asked about include: diabetes (ages 18+), asthma (all ages), hypertension/high blood pressure (ages 18+); high cholesterol (ages 18+) (since 2005), coronary heart disease (ages 18+), angina (ages 18+), heart attack/myocardial infarction (MI) (ages 18+), any other kind of heart condition (ages 18+), stroke or transient ischemic attack (TIA) (ages 18+), emphysema (ages 18+) and arthritis (ages 18+). For this paper analysis is restricted to ages 18 and over for all the conditions even though asthma was collected for all ages.

Before MEPS panel 12, the "ever" questions were asked in the priority conditions quality supplement section (PC) towards the end of the interview in rounds 3 and 5. The Diabetes Care Supplement was given out right after a person was reported as ever having diabetes. Follow-up asthma questions were asked right after a person was identified as ever having had asthma, and either still has asthma or had an episode of asthma or an asthma attack in the past 12 months.

Starting with MEPS panel 12 (the panel starting in the year 2007) the high prevalent "ever" condition questions were moved from the priority conditions quality supplement section (PC) that was asked towards the end of the interview in rounds 3 and 5 to the priority conditions enumeration section (PE) at the beginning of the interview in round 1 and subsequent rounds in certain situations. Starting with panel 12 the "ever" questions are asked: in round 1 of everyone; in round 2 of new people to the survey; in round 3 of persons whose response up to this point is not "yes" for the condition; in round 4 of new people to the survey; and, in round 5 of persons whose responses up to this point were not "yes" for the condition. Starting with Panel 12, follow-up asthma questions are attempted later in the priority conditions quality supplement section (PC) asked toward the end of the round 3 and 5 interviews of those who were reported to have ever had asthma in that or a previous round. Also an attempt is made in the PC toward the end of the round 3 and 5 interviews to distribute the diabetes care supplement, a paper-and-pencil self-administered questionnaire (SAQ), to those reported as ever having diabetes in that or a previous round.

2. Analysis on Burden

2.1 Is it worth asking new people in the survey condition questions in rounds 2 and 4?

Starting with Panel 12 (2007 Panel), the condition questions in the priority conditions enumeration section (PE) are asked of new people in rounds 2 and 4. We wanted to know whether it was worth asking new people to the survey in rounds 2 and 4 the PE condition questions. To answer this question the 2007 FY file which contains data for rounds 1-3 of Panel 12 and rounds 3-5 of Panel 11 was analyzed. By analyzing the Panel 12, rounds 1-3 data, we learned that only 2 new people in round 2 (not in round 1) would not have subsequently been in round 3 to get the condition questions in round 3. By analyzing the Panel 11, rounds 3-5 data, we learned that only 1 new person in round 4 (not in round 3) would not have been in round 5 of the survey and therefore been asked the questions in round 5. Since only a very few people new to the survey in rounds 2 and 4 would not get the questions in the next round, the rounds 2 and 4 interviews can be streamlined by not including the Priority conditions enumeration section for new people in those rounds without losing information.

2.2 What is additional burden if condition questions get asked of everyone in round 3 when they have already been asked in round 1 (or new people in round 2).

Starting with Panel 12, the condition questions are asked of everyone in round 1 and of only those who have not already reported the condition previously in rounds 3 and 5. To get an idea of the additional burden of asking the question of everyone, not just those who hadn't already reported the condition in a previous round, the Panel 11 longitudinal file where the condition questions are asked of everyone in both rounds 3 and 5 was used. For each condition, we compared the additional burden if the round 5 questions were asked of everyone versus if the round 5 condition questions were only asked of persons who didn't report yes in Round 3.

Table 1

Condition	Additional Burden if asked of people who responded yes in R3 and are eligible for responding in R5. (number of people)	Burden if asked of everyone eligible in R5 (number of people)	Burden by only asking those who did not say yes in R3 and are eligible for R5 (number of people)
Diabetes	1,059	11,469	10,410
Asthma	1,097	11,469	10,372
High BP 1+ times	3,129	11,469	8,340
High BP 2+ times	2,792	11,469	8,677
High cholesterol	2,747	11,469	8,722
Coronary Heart disease	328	11,469	11,141
Angina	213	11,469	11,256
Myocardial infarction	304	11,469	11,165
Other heart disease	651	11,469	10,818
Stroke	298	11,469	11,171
Emphysema	148	11,469	11,321
Joint pain	3,858	11,469	7,791
Arthritis	2,266	11,469	9,203

As shown in Table 1, it is clear that the burden is reduced more for higher prevalent conditions than for lower prevalent conditions by asking the questions only of those who did not say yes versus when the conditions are asked of everyone (column 2 versus column 3). The average reduction in burden across these conditions is 12.7 percent. One can speculate about the additional interview time resulting from the need of the CAPI instrument to store all of the previous rounds responses in the CAPI instrument for each condition for each person in the RU. One can also speculate upon the possible confusion of the interviewer and household respondent that may be created – for example there may be a household after the condition change in Panel 12, where in round 5, the Asthma questions are asked of everyone, the Diabetes questions are asked of everyone but say Jim, the cholesterol questions are asked of everyone but Mary, etc. Even though CAPI keeps track of this, it still adds to the complexity of the interview for the both the interviewer and the respondent.

2.3 Does moving the priority conditions to the beginning of round 1 before questions about events and prescription medicines reduce the interviewer burden?

The priority condition questions are now asked in the priority conditions enumeration section (PE) at the beginning of round 1 before the medical event and prescription medicine questions are asked. Conditions identified for a person from the PE are then added to a condition pick list from which the interviewer can select the condition as being associated with medical events and/or prescriptions asked about later in the interview. Prior to Panel 12, the interviewer needed to type out each of the conditions associated with the medical events and/or

prescriptions the first time the condition was identified. Selecting conditions from a pick list reduces the interviewer burden since the interviewer can select the condition from the pick list rather than having to type it out.

3. Analysis on estimates

3.1 Do priority condition estimates starting with Panel 12 (2007 Panel) differ from earlier estimates?

To try to answer whether the priority condition estimates starting with panel 12 differ from estimates from earlier panels, I compared estimates using the 2006 MEPS full-year file (contains data for panels 10 (rounds 3 to 5) and 11 (rounds 1 to 3)) with estimates using the 2008 MEPS full-year file (contains data for panels 12 (rounds 3 to 5) and 13 (rounds 1 to 3)).

Table 2

Condition	2006 MEPS (Percent)	2008 MEPS (Percent)	Z-test
Diabetes	7.9	9.7	5.33
Asthma	9.6	9.0	-1.60
High blood pressure 1+ times	26.6	31.9	7.70
High blood pressure 2+ times	23.6	25.9	3.49
High cholesterol	24.9	31.2	10.11
Coronary heart disease[1]	5.0	8.2	9.25
Angina	1.9	3.2	6.00
Myocardial infarction	2.9	4.0	4.43
Other heart disease	5.9	11.1	14.11
Stroke or TIA	2.4	3.8	6.85
Emphysema	1.4	2.5	6.31
Arthritis	20.0	24.9	7.94

[1]Coronary heart disease includes coronary heart disease, angina, or myocardial infarction.

As shown in Table 2,MEPS condition estimates increased significantly at the .05 level from 2006 to 2008 for eleven of the 12 conditions: Diabetes, High Blood Pressure, High cholesterol, Coronary Heart Disease, Angina, Myocardial Infarction, Other Heart Disease, Stroke, Emphysema, and Arthritis. Asthma estimates did not change significantly from 2006 to 2008.

3.2 Do MEPS changes in priority condition estimates from 2006 to 2008 differ from NHIS changes in condition estimates from 2006 to 2008?

To try to answer whether the MEPS condition changes from 2006-2008 signify actual changes in condition prevalence or whether they may have resulted from the changes in the way condition questions are asked and the estimates are made, MEPS condition changes from 2006 to 2008 (before and after the MEPS condition changes) for the conditions also asked about in the NHIS, are compared with the NHIS condition changes from 2006 to 2008 during which period NHIS condition collection procedures did not change. As in section 2.4 above, MEPS estimates for 2006 to 2008 were produced using the 2006 and 2008 MEPS full-year files. The NHIS condition estimates are from the CDC Vital and Health Statistics Series 10, Numbers 235 and 242 (NHIS) reports.[8,9]

MEPS and NHIS changes are shown for 8 conditions common to both surveys during this time period in Table 3.

Table 3

Condition	MEPS 2006 Percent	MEPS 2008 Percent	MEPS 2006-2008 Z-test	NHIS 2006 Percent	NHIS 2008 Percent	NHIS 2006-2008 Z-test
Diabetes	7.9	9.7	5.33	7.9	8.4	1.61
Asthma	9.6	9.0	-1.60	11.0	12.6	4.11
High Blood Pressure 2+ times	23.6	25.9	3.49	23.5	25.0	2.87
Coronary heart disease[1]	5.0	8.2	9.25	6.4	6.4	0.00
All heart disease[2]	9.6	14.5	10.96	11.0	11.8	2.21
Stroke	2.4	3.8	6.85	2.6	2.9	1.70
Emphysema	1.4	2.5	6.31	1.8	1.7	-0.74
Arthritis	20.0	24.9	7.94	21.2	21.7	1.08

[1]Coronary heart disease includes Coronary heart disease, Angina or Myocardial infarction.
[2]All heart disease includes Coronary heart disease, Angina, Myocardial infarction, or Other heart disease.

Because MEPS condition estimates increased from 2006 to 2008 for 7 of the 8 conditions (Table 3) and NHIS condition estimates increased for only 3 of the 8 conditions, we conclude that the MEPS 2006-2008 estimate changes may have resulted from changes in the way the conditions are asked starting with Panel 12, in addition to any actual changes in prevalence.

4. Conclusion

4.1 Summary

In summary MEPS changed the way condition questions are asked starting with panel 12 (2007 panel). The high prevalent "ever" condition questions were moved from the priority conditions quality supplement section (PC) asked towards the end of the rounds 3 and 5 interviews in panels prior to panel 12, to the priority conditions enumeration section (PE) asked at the beginning of the interview in round 1 and in certain situations subsequent rounds starting with panel 12.

The interview process could be streamlined without losing information by not asking the PE of new people in rounds 2 and 4 since all but a few of the new people in rounds 2 and 4 would get the condition questions in the immediately following round.

The burden was reduced marginally for most conditions by asking about conditions in rounds 3 and 5 only those who did not report the condition up to that point. This change only reduced the burden marginally and may have added complexity to the interview process for both the interviewer and respondent.

Interviewer burden was reduced by asking the condition questions at the beginning of round 1 and by adding the conditions identified by the condition questions to a pick list

which interviewers can select as reasons for medical events or prescriptions asked about later in the interview.

It appears that MEPS condition estimates from condition questions have changed starting with Panel 12. It is hard to tell if the changes in estimates are due to condition question changes since other changes were made at the same time. In addition to the condition question changes, MEPS converted from a DOS to a window's based CAPI instrument in 2007 and the sample changed in 2007.

4.2 Future work

More work is needed to determine whether condition estimate trends were affected by changes in how the condition questions are asked. Condition estimates were compared for 2006 and 2008. More years of data could be analyzed to determine whether there was a change in trend starting with Panel 12 data. Additional work could be done in assessing whether current edits are behaving as intended in producing condition estimates starting with Panel 12. Analysis of trend data using only round 1 estimates (when the questions are asked of everyone) could be undertaken to determine whether trends using only round 1 data differ from those calculated when all the rounds of data are used. Time-stamp paradata could be used to assess whether the time it takes to ask the condition questions in round 1 of everyone is substantially longer than it takes to ask the condition questions in rounds 3 and 5 of only those who have not reported the condition previously. Additional analyses of changes to the reporting of conditions from medical events and the questions on conditions bothering the person could be done to see if the number of conditions or distribution of conditions differs starting with Panel 12 data.

4.3 Streamlining the questionnaire/Conclusion

There are a number of ways to streamline the process of collecting condition information. One possibility is using a handcard with a list of conditions and having the interviewer ask family-style questions about whether they or anyone else in their family have ever been told by a doctor or other health professional that they had the conditions, where the interviewer would read aloud each of the conditions on the handcard. For longitudinal surveys such as MEPS, there are different ways of reducing the number of times the condition questions get asked to streamline the interview and reduce burden. Perhaps asking the condition questions once of everyone in round 1 would suffice in producing unbiased prevalence estimates for the full year files. If asking the condition questions only in round 1 was not sufficient, then the condition questions could be asked of everyone in round 3. For those conditions with additional treatment questions or SAQs in rounds 3 and 5, then only those questions could be asked of everyone in rounds 3 and 5 thereby reducing burden. In rounds 3 and 5, there are currently two sections in which condition specific questions are asked: the priority conditions enumeration section (PE); and the priority conditions quality supplement section (PC). Combining these two sections for rounds 3 and 5 might also streamline and simplify the interview.

In conclusion, evaluating and revising condition questions and how they are asked in a longitudinal survey to improve the reporting of conditions and to reduce respondent burden is exemplary. However, collecting condition information can be complicated and an iterative process may be needed to get the best results in data collection efficiency while maintaining and improving data quality for conditions.

References

[1] Machlin S, Cohen J, Beauregard K. *Health Care Expenses for Adults with Chronic conditions 2005*. Statistical Brief #203. May 2008. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st203/stat203.pdf.

[2] Cohen J., Krauss N. Spending and Service Use among People with the Fifteen Most costly medical conditions, 1997, *Health Affairs*, March/April 2003.

[3] Soni A. *The five Most Costly Conditions, 2000 and 2004: Estimates for the U.S. Civilian Noninstutioinalized Population*. Statistical Brief #167. March 2007. Agency for Healthcare Research and Quality. Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st167/stat167.pdf.

[4] Machlin S, Cohen J, Elixhauser A, Beauregard K, Steiner C. Sensitivity of Household Reported Medical conditions in the Medical Expenditure Panel Survey. *Medical Care*. 2009;47(6):618-625.

[5] Cox B, Cohen S. A comparison of household and provider reports of medical conditions. In: Cox B, Cohen S, eds. *Methodological Issues for Health Care Surveys*. New York, NY: Marcel Dekker Inc; 1985:150-189.

[6] Krauss N, Kass B. Comparison of household and medical provider reports of medical conditions. Paper presented at: The Joint Statistical Meetings; august 2000; Indianapolis, Indiana.

[7] Edwards WS, Winn DM, Kurlantzik V, et al. Evaluation of Natonal Health interview Survey Diagnositc Reporting. National Center for Health Statistics. VitalHealth Stat 2(120). 1994.

[8] Pleis JR, Lethbridge-Cejku M. Summary health statistics for U.S. adults: National Health Interview Survey, 2006. National Center for Health Statistics. Vital Health Stat 10(235). 2007.

[9] Pleis JR,Lucas JW, Ward BW. Summary health statistics for U.S. adults: National Health Interview Survey, 2008. National Center for Health Statistics. Vital Health Stat 10(242). 2009.