Missing Data for Children's Height, Weight, and BMI in the Medical Expenditure Panel Survey

Frances M. Chevarley
Center for Financing Access and Cost
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850-6649

Keywords: MEPS, height, weight, BMI, item nonresponse

1. Introduction

The Medical Expenditure Panel Survey (MEPS) is designed to provide nationally representative annual estimates of health care use and expenditures, access to care, patient and customer satisfaction, health status, and insurance coverage for the U.S. civilian noninstitutionalized population. It is co-sponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS) and is an annual survey that has been fielded since 1996.

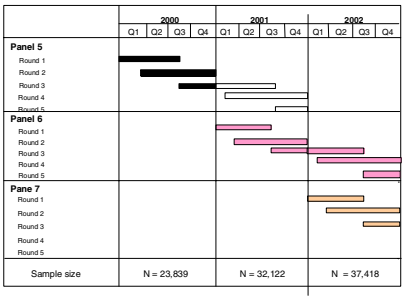
Children's height and weight questions and subsequently Body Mass Index Information (BMI) (which is calculated as weight in kilograms divided by height in meters squared) appear to have high item nonresponse rates. This paper analyzes missing height and weight information from 1996-2002 in an attempt to describe the extent of this problem. Several suggestions are made concerning how these data may be improved.

2. MEPS Household Component

The MEPS actually comprises a family of surveys that cover three different major components of the U.S. health care system: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC) (Cohen J, 1997; Cohen SB, 2000). In this paper we focus on children's data based on the MEPS-HC.

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. This material was presented at the May 2005 AAPOR Annual Meeting in Miami.

Figure 1: MEPS Panel Design: Data Reference Periods



The MEPS-HC collects data through an overlapping panel design (Figure 1). In this design, two calendar years of information are collected from each household through five in-person interviews over approximately 2 ½ years. Each of these interviews is called a Round. Since not all interviews can be conducted on exactly the same day, reference periods for the 5 interviews or Rounds vary in length and can range from 3-6 months. Round 1,2, and part of 3 contain year 1 of data for a panel; part of Round 3, 4, and 5 comprise year 2 data for a panel. This series of data collection activities is repeated each year on a new sample of households resulting in overlapping panels of survey data. For example, data from the second year of Panel 6 can be combined with the first year of Panel 7 to yield a sample size of 37,418 for the 2002 calendar year. Children's sample sizes from the MEPS-HC range from 6,400 from the 1996 file to 11,500 in the 2002 file (Table 1).

Table 1: MEPS – Sample Size

Year	Households	Persons	Children
1996	9,400	23,500	6,400
1997	13,500	33,000	10,000
1998-2000	10,000	25,000	7,000
2001	13,500	33,000	10,300
2002	15,000	37,000	11,500

3. Source of Data

Children's height and weight data have been collected by MEPS since 1996 and are collected in Rounds 2 and 4 which are fielded during the second half of the year. Children's height and weight data are on the MEPS public use files, or PUFs, from 1996-2000 (Table 2). Starting in 2001, children's body mass index is computed and is included in the MEPS PUFs. Starting in 2001, children's height and weight information are no longer on the MEPS PUFs but are available in the MEPS data center.

Table 2: Availability of MEPS Children's height, weight, and BMI

Year	Height & weight collected	Height & weight on PUFs	BMI on PUFs
1996-2000	X	X	
2001-2002	X	1_/	X

1_/ Available in MEPS data center.

For each child in the family, respondents are asked to report about how tall the child is without shoes and also about how much the child weighs without shoes. These questions remained the same since 1996 with only slight differences in CAPI edit checks that were introduced in 2001.

However, the placement of these questions in the questionnaire changed in 2001. From 1996-2000, these children's height and weight questions had been asked after the children's general health status questions. Starting in 2001, the children's height (weight) question is asked after ascertaining when a doctor or other health professional last measured the child's height (weight).

The distributions of BMI values differ for children than for adults because BMI changes as children get older. For children, the 2000 CDC age- and gender-specific BMI growth charts were used to ascertain whether or not a child was overweight or at risk for being overweight. Information on the 2000 CDC Growth charts are available from the CDC's National Center for Health Statistics website www.cdc.gov/growthcharts. In these charts expert panel recommendations were used in classifying as "overweight" BMI-for-age at or above the 95th percentile and as "at risk of overweight" BMI-for-age between the 85th and 95th percentile (Himes & Dietz, 1994; Barlow & Dietz, 1998; Kuczmarski, Ogden Grummer-Strawn, et al., 2000). The CDC BMI growth charts are not intended to be used as a sole determinant of being at risk or of being overweight, but rather as screening tools that contribute along with other information to an overall evaluation of a child's weight.

4. Results

As a measure of quality of children's weight and height data, sample item nonresponse for both weight and height were calculated from 1996 to 2002 for the age groups: 3-4, 5-9, 10-14, 15-17, and 18 and over (Tables 3 & 4, respectively). The age group 18 and over is included to assess how item nonresponse for children's weight and height compares to that of adults.

Table 3: Sample Item Nonresponse for weight

Year	Ages 3-4	Ages 5-9	Ages 10-14	Ages 15-17	Ages 18+
1996	7.5	7.8	4.9	3.7	
1997	8.4	9.0	5.7	5.2	
1998	8.1	8.0	5.4	5.8	
1999	8.3	8.8	5.7	5.1	
2000	6.6	7.2	4.6	3.5	4.3
2001	6.9	8.8	7.2	6.4	4.2
2002	9.9	9.7	8.3	6.7	4.2

Based on Table 3, children's item nonresponse for weight seems to be higher for children ages 3-4 and 5-9 then for ages 10-14 and 15-17. It also looks like children's item nonresponse for weight may show an increase starting in 2001. There does not seem to be a corresponding increase in weight item nonresponse for adults starting in 2001.

Similar to the pattern of item nonresponse for weight, item nonresponse for height is higher for younger children (ages 3-4 and 5-9) than for older children (ages 10-14 and 15-17) (Table 4). Height item nonresponse for younger children (ages 3-4 and 5-9) also tends to be about double the corresponding weight item nonresponse for these children. As with the pattern of weight item nonresponse, height item nonresponse may show an increase for children but not for adults starting in 2001.

Table 4: Sample Item Nonresponse for Height

Year	Ages 3-4	Ages 5-9	Ages 10-14	Ages 15-17	Ages 18+
1996	18.8	14.5	5.4	2.0	
1997	18.9	17.1	6.0	3.5	
1998	19.6	13.9	6.2	3.6	
1999	16.3	15.2	6.3	3.5	
2000	15.2	11.3	4.7	1.9	1.1
2001	21.8	19.5	8.7	4.4	1.2
2002	27.1	23.5	10.8	5.2	1.3

For both height and weight, almost all of the nonresponse or missing data is "Don't Know" as opposed to "Refused" or "Not ascertained". In 2002, for children ages 3-17 "Don't Know" made up 99.74% of the missing cases of height item nonresponse and 99.39% of the missing cases of weight item nonresponse. Item nonresponse was calculated for the other 71 questions in the Child health and preventive care section of the MEPS survey in 2002 to ascertain whether height and weight item nonresponse were the exception. Almost all of the other questions in this section have item nonresponse less than 1%. The highest item nonresponse for a few other questions was at most 3.2%.

Because children's height item nonresponse had larger values than weight item nonresponse, the rest of the paper focuses mainly on height item nonresponse. In order to examine the potential effect of height item nonresponse on estimated statistics and analysis, weighted estimates of height item nonresponse are used for the rest of the paper. Using the SUDAAN software package, chi-square tests and corresponding p-values were calculated using a p=.05 cut-off value to determine whether or not height item nonresponse varied by different variables.

As seen in Table 5, height item nonresponse varied by age, race/ethnicity, poverty status, health status, and region. It was not significantly associated with gender or MSA status. When the more refined urban influence codes were used (Ghelfi & Parker, 2004), height item nonresponse did vary by the four collapsed categories of the 2003 urban influence codes of Large metropolitan (metro), Small metro, Micropolitan (Micro), and Noncore.

In order to determine whether children's height and weight item nonresponse increased starting in 2001, as suggested by the data in Tables 3 & 4, weighted height and weight item nonresponse were calculated by age for 2000 and 2001 and associated z-tests were performed (Tables 6 & 7). Based on these z-tests using a 0.05 cut-off value, item nonresponse for height increased significantly from 2000 to 2001 for children ages 3-4, 5-9, and 10-14. The increase in height item nonresponse for children ages 15-17 was not significant.

According to Table 7, item nonresponse for weight increased significantly for children ages 10-14 and 15-17. Because the children's height and weight questions for both 2000 and 2001 are exactly the same with only slight differences in CAPI edits, it is thought that these increases may be due to the change in location of these questions in the MEPS CAPI questionnaire from 2000 to 2001. As mentioned earlier, the height and weight questions were asked in 2000 after a couple of children's general health status questions which are from the General Health Subscale of the Child Health Questionnaire. In particular they were asked after a question about whether the "child usually catches something that is going around". Starting in 2001, these questions were asked within a series of questions on receipt of specific clinical preventive services. These questions begin by asking about when was the most recent time a doctor other health professional measured the child's height and then the question is asked about how tall the child is without shoes. Next, questions are asked about when was the most recent time that a doctor or other health professional measured the child's weight and then the question is asked About how much does the child weigh without shoes. By grouping the height and weight questions with the questions on when the doctor or other health professional last measured the child's height and weight, may have created an order effect. In generating a response, the common model is to divide the respondent's response process into four major components (Tourangeau, Rips, & Rasinski, 2000): comprehension of the item; retrieval of relevant information; use of that information to make required judgments; and selection and reporting of an answer. In our situation, respondents may be misinterpreting the question as asking for the height and weight as last measured by a doctor or other health professional which may lead to larger unknown values depending on how long ago a doctor or other health professional had measured the child's height and/or weight. This also could lead to biased estimates since presumably height and weight would be increasing over time for most children. Even if the respondent realizes they are to report current height and weight (and not last measured in a doctor's visit), there may be errors in retrieving the correct information because the earlier questions about the doctor's measuring the child's height and weight may bias their retrieval to the

height and weight as last measured by a doctor, rather than the current height and weight of the child.

Table 5: Weighted Item Nonresponse Children's Height (ages 3-17), 2002 MEPS

Item		Non-response	p-value		Item		Non-response	p-value
Age	3-4	23.6%	<0.0001		Health Status	Excellent	10.7%	<0.0001
	5-9	19.6%				Very Good	16.6%	
	10-14	7.9%				Good	19.6%	
	15-17	3.9%				Fair/Poor	16.8%	
Race/ ethnicity[a]	Hispanic	24,3%	<0.0001		Region	Northeast	15.4%	0.0307
	Black, single race, NH[b]	15.0%				Midwest	10.9%	
	White, single race, NH[b]	10.4%				South	13.3%	
Gender	Male	13.6%	0.1024			West	16.9%	
	Female	14.5%			MSA status	MSA	14.6%	0.0630
Poverty Status	<200% poverty level	19.7%	<0.0001			non-MSA	11.5%	
	>=200% poverty level	10.5%	<0.0001		Urban Influence Category	Large metro	15.8%	0.0024
						Small metro	11.6%	
						Micro	13.6%	
						Noncore	8.8%	

[a]Data for the Other, non-Hispanic category is not shown.
[b]NH=non-Hispanic.

Table 6: Comparing weighted height item nonresponse from 2000 to 2001

Ages	Height item nonresponse (%)			
	2000	2001		Z-test
3-17	6.4	11.0		5.6
3-4	12.7	19.7		3.4
5-9	9.4	16.8		4.8
10-14	3.6	7.0		4.2
15-17	1.6	2.7		1.8

Table 7: Comparing weighted weight item nonresponse from 2000 to 2001

Ages	Weight item nonresponse (%)			
	2000	2001		Z-test
3-17	4.6	6.5		2.8
3-4	6.1	5.8		-0.2
5-9	5.8	7.7		1.9
10-14	4.0	6.2		2.6
15-17	2.7	5.3		3.0

5. Conclusion

Children's height and weight item nonresponse within MEPS is high and this results in a large amount of missing BMI information for children. Even if height and weight item nonresponse were reduced, there is still the issue of measurement error. According to a telephone conversation with Cynthia Ogden from the National Center for Health Statistics, there may be over reporting of BMI for children because parents may have more recent weight than height information and therefore tend to underreport height more than weight. Since BMI is (weight/height2) (in kilograms and meters, respectively) this could lead to higher reported BMI.

At this point we have guarded optimism that improvements can be made in the quality of the height and weight information for children collected in MEPS. As a first step, we would want to move the children's height and weight questions back to where they had been prior to 2001, that is after the children's general health status questions. There would still be a large amount of missing height and weight data, especially for younger children, but it would more than likely be reduced to the level it was prior to 2001. In addition, we could consider changing the CAPI questionnaire so that respondents would have the option of reporting height and weight in metric units. Having the option of reporting height and weight in metric units might have the potential of reducing missing information for Hispanics and/or other groups which have differentially more item nonresponse as seen in Table 5.

In addition to missing data, measurement error is still a real issue and direct measurement would improve the results. A possibility would be to contact the parents prior to the Round 2 and 4 interviews and ask that they measure and record their children's height and weight on a small form before the Round 2 and 4 interviews. In considering this option, we would need to weigh any possible negative consequences of asking respondents to do this additional chore on top of all of the other tasks they are already asked to do for the survey over 2 ½ years with 5 interviews and numerous supplemental activities.

References

Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. Pediatrics 102(3)29; 1-11. 1998. URL: http://www.pediatarics.org/cgi/content/full/102/3/e29.

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

Cohen SB. Sample design of the 1997 Medical Expenditure Panel Survey Household Component. Rockville (MD): Agency for Healthcare Research and Quality, 2000. MEPS Methodology Report No. 11, AHRQ Pub. No. 01-0001.

Ghelfi L, Parker T. Developing a county-level measure of urban influence. Amber Waves. April 2004. Available at http://www.ers.usda.gov/Amberwaves/Arpil04/Indicators/BehindData.htm. Accessed October 24, 2005.

Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: Recommendations from an expert committee. Am J Clin Nutr 59:307-16. 1994.

Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, et al. CDC growth charts: United States. Advance data from vital and health statistics; no. 314. Hyattsville, Maryland: National Center for Health Statistics. 2000.

Tourangeau R, Rips LJ, Rasinski K. The psychology of survey response. Cambridge University Press, 2000.