TECHNICAL PAPERS ON
HEALTH AND BEHAVIOR MEASUREMENT

TECHNICAL PAPER NO. 35

Monitoring the Health Status and Impact of Treatment on Older Americans: The Medicare Beneficiary Health Status Registry

A. Marshal McBean, Charles F. Turner, Leslye K. Fitterman, Kirk Pate, Thomas Reilly, Timothy K. Smith, Ann Trontell, Michael B. Witt, Lynne Penberthy, Judith T. Lessler, Barbara H. Forsyth, Sara Wheeless, Frank Mierzwa, and Heather G. Miller

Reference Citation

McBean, A.M., Turner, C.F., Fitterman, L.K., Pate, K., Reilly, T., Smith, T.K., Trontell, A., Witt, M.B., Penberthy, L., Lessler, J.T., Forsyth, B.H., Wheeless, S., Mierzwa, F., and Miller, H.G. (1999) Monitoring the Health Status and Impact of Treatment on Americans: The Medicare Beneficiary Health Status. *Medical Care* 37(2):189-203.

MEDICAL CARE
Volume 37, Number 2, pp 189–203

Monitoring the Health Status and Impact of Treatment on Americans

The Medicare Beneficiary Health Status Registry

A. Marshall McBean, MD,* Charles F. Turner, PhD,† Leslye K. Fitterman, PhD,††
Kirk Pate, BA,¶ Thomas Reilly, PhD,‡ Timothy K. Smith, MA,¶ Anne Trontell, MD,**
Michael B. Witt, MA,¶ Lynne Penberthy, MD,# Judith T. Lessler, PhD,¶
Barbara H. Forsyth, PhD,§ Sara Wheeless, PhD,¶ Frank Mierzwa, BA,¶
and Heather G. Miller, PhD†

Objective. A major new survey program, the Medicare Beneficiary Health Status Registry (MBHSR), has been proposed to improve the monitoring of the health status of Medicare beneficiaries. The MBHSR would collect data by mail with telephone follow up of nonrespondents to permit economical assessment of a total Registry of approximately 200,000 Medicare beneficiaries, approximately 54,000 of whom would be surveyed in any given year. (Surveys would be conducted of samples of new enrollees who would be reinterviewed every five years.)

Method. To assess the feasibility of that approach, a field test was conducted with a probability sample (n = 1,922) that comprised approximately equal numbers of new Medicare enrollees (aged, 65) and current beneficiaries (age range, 76–80). The field test was designed to assess the quality of the data that this design would produce.

Findings. Results indicate that the proposed design of the MBHSR could achieve response rates of approximately 80% among both age cohorts using a survey instrument that took 30 minutes to complete. Internal reliability of Activities of Daily Living, Instrumental Activities of Daily Living, Mobility, Mental Health Index, General Health, and Prostate Symptomatology scales ranged from 0.77 to 0.93. When measurements were

*From the Department of Health Management and Policy, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

†From the Program in Health and Behavior Measurement, Research Triangle Institute, Washington, D.C.

‡From the Health Care Financing Administration, Baltimore, Maryland.

¶From the Research Triangle Institute, Research Triangle Park, North Carolina.

#From the Medical College of Virginia, Richmond, Virginia.

§From Westat, Rockville, Maryland.

**From the Food and Drug Administration, Rockville, Maryland.

††From the Outcomes Research and Management, Merck and Co., Inc., West Point, Pennsylvania.

The field test of the Medicare Beneficiary Health Status Registry (MBHSR) and the analyses reported in this article were supported under contract 500-90-0053 between the Health Care Financing Administration (HCFA) and Research Triangle Institute (RTI). Dr. A. Marshall McBean was Chief of the Epidemiology Branch of HCFA's Office of Research during the design, planning, and execution of the field test. Dr. Charles Turner served as Principal Investigator, Dr. Thomas Reilly as Project Officer, and Mr. Kirk Pate as Project Director from 1992 through completion of the project in 1995. Dr. Judith T. Lessler served as Principal Investigator and Project Director, and Dr. Lynne Penberthy served as Project Officer, from the inception of the field test in 1990 through 1992.

Address correspondence to: Charles F. Turner, PhD, Program in Health and Behavior Measurement, Research Triangle Institute, 1615 M Street, NW, Washington DC 20036. E-mail: cft@rti.org.

Received December 9, 1997; initial review completed February 26, 1998; final acceptance July 31, 1998.

repeated approximately 30 days after the initial survey, moderate to high levels of cross temporal correlation (range, 0.64–0.96) were found for most indexes, with the exception of prostate symptomatology. In addition, an earlier comparison of survey responses in the MBHSR field test to Medicare payment records indicated that the MBHSR field test obtained highly accurate reports of most of the major surgeries that were recorded in Medicare claims files.

CONCLUSION. The design proposed for the MBHSR is feasible. If implemented, it should produce acceptably high rates of response and data quality.

Key words: survey measurement; elderly population; Medicare; monitoring systems; health status. (Med Care 1999;37:189–203)

As the Medicare program has matured, increasing amounts of information have accumulated on the use of services by Medicare beneficiaries. The resulting data on service utilization are critical to oversight and evaluation of the Medicare program, as well as to research on its costs, efficiency, and effectiveness. However, both program and research analyses are constrained by a lack of information on the health status of beneficiaries, their need for services, and quality-of-life outcomes of the care they receive through Medicare. That gap in information led to a proposal for establishing the Medicare Beneficiary Health Status Registry (MBHSR). The registry was conceived as a national resource of data on the Medicare population that would help agencies and health researchers to accomplish the following:

- measure access to medical care;
- understand beneficiaries' patterns of use and their need for health services;
- assess the impact of changes in the provision of health care under Medicare;
- provide a basis for the development of improved risk-adjustment procedures in creating reimbursement schedules for health care providers; and
- assess the effectiveness and quality of care provided through the Medicare program.

There is little argument about the value of the data that the registry can provide or about the expectation that the data will have many other important users besides the Medicare program. But most of the uses proposed for the registry depend on surveying large representative samples each year. Current budgetary constraints make it infeasible to conduct a large annual survey if data must be collected mainly through in-person or telephone interviews.

As a result, some researchers have argued for adopting the less costly strategy of using data generated by other ongoing health surveys, such as the National Health and Nutrition Examination Surveys the National Health Interview Survey and the Longitudinal Survey on Aging. Data from those cross-sectional and longitudinal surveys have been used to estimate functional status, as well as the prevalence of a number of disease-related risk factors in the elderly population. However, the number of elderly people who participate in those surveys is not large enough to produce reliable estimates of key indicators for vulnerable Medicare subgroups, such as Hispanics, rural residents, diabetics, or current smokers. Historically, substantive deficiencies of that kind plus significant operational barriers have prevented the linkage of data from other surveys to Medicare's enrollment and administrative data. The inability to draw together all of the available information underlies the proposal to collect new data to generate subpopulation-specific estimates of health and functional status and behavioral risk factors for elderly people.

To produce such estimates, the designers of the MBHSR anticipate the use of large probability samples of Medicare enrollees who would be followed longitudinally using a combination of mail surveys with telephone follow up of nonrespondents. Theoretically, one could consider expanding the sample of an extant survey, eg, the NHIS, to produce such estimates. There are several impediments to such a strategy. Most importantly, the costs of an annual 54,000 in-person supplement to these surveys would be prohibitive. In contrast, the annual costs of the proposed MBHSR data collection are projected to be less than $3 million per year.

After 4 years of design and testing, the MBHSR researchers adopted a data collection design that would, when it reaches a steady state, collect baseline data *each year* from a probability sample of new enrollees who would be reinterviewed every 5 years until death. (This design has gone

hrough several redefinitions; the design described n this article resulted from the final MBHSR lesign contract.) Enrollment in the Registry would :otal over 200,000. Approximately 12,000 persons per year would be surveyed at intake (age 65), and sample mortality would result in smaller numbers completing their 5-year follow-up interviews at age 70, 75, 80, etc., in each year. It is estimated that approximately 6,600 persons would survive to complete follow-up interviews at age 80; a total of approximately 54,000 interviews (intake plus 5-year follow ups) would be conducted each year.[6] Other features of the MBHSR design would include use of a "jump-start" sample to provide an initial Registry of older beneficiaries (for whom there would, of course, be no intake data at the age of 65), annual data collection from a small subsample of the Registry to provide year-to-year longitudinal measures, and a possible mortality follow back survey.

To balance the need for data against the cost of collecting data, the designers of the MBHSR propose to collect the bulk of the data by mail with self-administered questionnaires supplemented by telephone follow up of nonrespondents. Mail surveys offer an inexpensive way to collect data from a large number of individuals, but their economic savings may be offset by lower data quality. To put it bluntly, mail surveys have a bad reputation largely because of the problems investigators have encountered in achieving adequate response rates and the problems respondents have encountered in completing questionnaires. In a mail survey of an older population, it is not difficult to imagine that some respondents might have physical or cognitive limitations that would preclude them from completing a paper-and-pencil, self-administered questionnaire (SAQ). Those limitations could result in a higher nonresponse rate among people with poor health or functional status. Another all too likely scenario is that with the current volume of "junk mail," potential respondents might dismiss and discard the questionnaire if the importance of the project and the bona fides of the sender have not been clearly communicated. Along with possibly poor response rates, researchers must also be concerned about the validity and reliability of the resultant measurements.

To test the feasibility of collecting MBHSR data through a mail survey, we conducted a field test of the registry design during the summer and early fall of 1993. The test was designed to assess whether an MBHSR that used mail surveys, coupled with telephone follow up of nonrespondents, would achieve an adequate response rate and whether the resulting measurements would have acceptable validity and reliability. In this article, we review the evidence obtained on the completeness of survey response, nonresponse bias, and the internal and cross-temporal reliability of MBHSR measurements of health status. Elsewhere, we have reported our positive findings on the validity of the reporting of medical conditions and procedures in the MBHSR Field Test; we do not repeat our presentation of those findings in this paper.

Methods

The following discussion summarizes key design elements and methods that were used to collect and analyze the MBHSR data. More detailed descriptions of the sampling design and execution of the field test can be found elsewhere.

Sample

The sample for the MBHSR field test was drawn in two parts: a national sample and a reinterview sample. To control the costs of in-person data collection required for the reinterview study, we selected the reinterview sample from 3 areas convenient to the survey organization headquarters in North Carolina. The sample design for the field test was a two-stage, deeply stratified design. Each primary sampling unit (PSU) represented a single county or group of adjacent counties in the coterminous United States. At the first stage of sample selection, 3 PSUs (Raleigh, Atlanta, and Richmond) were selected with certainty and 6 non-MSA and 18 MSA PSUs were selected with probabilities of selection that is proportional to the numbers of survey-eligible individuals who reside in the PSU. After the PSUs were selected, a second-stage sample frame was created. The frame consisted of all beneficiaries who resided in the sample PSUs who met the following criteria:

- they were 65 years old in February 1993 with birth dates between June 1927 and February 1928. Sampled beneficiaries from this stratum are referred to as the 65-year-old cohort; and
- they were 76 to 80 years old in February 1993 with birth dates between March 1912 and Feb-

ruary 1917. Sampled beneficiaries from this stratum are referred to as the 76- to 80-year-old cohort.

In addition to the stratification by age groups, the second-stage frame was also stratified by race (white vs. nonwhite), gender, and a health status indicator. The health status indicator was defined only for the 76- to 80-year-old cohort as Medicare claims records were not available for new enrollees. It was created using health information from the current Medicare Provider Analysis and Review (MEDPAR) file. The indicator identified respondents who received the following procedures or diagnoses during a hospital stay that was covered by Medicare: cataract removal; gall bladder removal; coronary artery bypass surgery; pacemaker insertion; coronary angioplasty; breast surgery or breast removal; hysterectomy; hip replacement; prostatectomy; asthma; cancer; chronic obstructive pulmonary disease (COPD); diabetes; stroke; and heart attack.

The nonwhite and less healthy sample strata were oversampled for the field test; allocation of the sample across genders remained roughly proportional to population counts. The final step in the procedures for sample selection involved the random partition of the sample enrollees to the field test's experimental treatments (form length and prenotification). This step was carried out independently for the 2 cohorts and, to the extent possible, independently between the other second-stage strata. Table 1 shows the target-cell sizes by experimental condition and selected sociodemographic factors for the main field test sample and for the reinterview study combined.

Sample Weighting. After the data were collected, sampling weights were constructed to reflect each respondent's likelihood of being selected for the sample. The weights were then adjusted to account for unit nonresponse by using an unconstrained, polytomous extension of the constrained logistic modeling procedures of Folson and Devill and Samdall. Standard errors were calculated using algorithms that take account of unequal weighting and the complex sample design used in this survey.

TABLE 1. Distribution of Target Sample Cases (Unweighted) by Experimental Conditions and Selected Social and Demographic Characteristics

Factor	COHORT 65	COHORT 76–80
Survey prenotification		
Prenotification	622	636
No prenotification	611	641
Survey length		
Short	413	426
Medium	409	427
Long	411	424
Health status*		
0 Conditions	n.a.	638
1+ Conditions	n.a.	639
Gender		
Male	551	492
Female	682	785
Race		
White	836	850
Nonwhite	397	427
Residence		
MSA	537	568
Non-MSA	696	709
TOTAL	1,233	1,277

Note: Characteristics of target respondents as recorded in Medicare enrollment records.

MSA, metropolitan statistical area; n.a., not applicable.

*Based on prior claims for cataract removal, gall bladder removal, coronary bypass surgery, pacemaker insertion, coronary angioplasty, breast surgery or removal, hysterectomy, hip replacement, prostatectomy, asthma, cancer, chronic obstructive pulmonary disease, diabetes, stroke, and heart attack.

Data Collection

Questionnaire. The MBHSR questionnaire was designed to obtain information in the following 5 domains: health behaviors; functional status; medical history; quality of life; and sociodemographics. The questionnaire's designers began by adapting relevant questions and scales from other instruments that had been used successfully in previous research among elderly populations. Items from the Longitudinal Study of Aging questionnaire, the Long-Term Care Survey, the Medical Outcomes Study Health Status Questionnaire, and the Behavioral Risk Factor Surveillance System were included. The content and format of the field test questionnaire were refined through a program of cognitive testing. The testing was undertaken by a team of survey methodologists and cognitive psychologists whose recommendations were reviewed by a working group on instrumentation and staff of the Epidemiology Branch at

the Office of Research of the Health Care Financing Administration. Cognitive testing identified problems that respondents encountered with proposed questions, such as difficulty following "skip" instructions, failure to notice introductory statements that preceded questions, poor comprehension of negative modifiers, and difficulties with technical terms. Modifications were made to reduce these problems. This process did cause MBHSR wordings of some questions to differ from the wordings used in earlier research.[1]

Main Data Collection. Field operations were carried out in several stages beginning in the spring of 1993. Before the questionnaires were mailed, study personnel sent lead letters to the people who were assigned to the prenotification group to explain the purpose of the study, emphasize its importance, highlight the need for information from Medicare beneficiaries, and alert recipients to the impending arrival of the questionnaire. Three weeks later (May 20, 1993), one version (short, medium, or long) of the questionnaire was sent to each beneficiary selected for the field test. People who did not respond within 7 weeks were sent another questionnaire on July 6, 1993. The process was repeated once more on August 9, 1993. Three weeks after the third questionnaire was mailed (September 4, 1993), all nonrespondents were assigned to RTI's Telephone Survey Unit (TSU) for follow up. TSU personnel asked beneficiaries who were contacted by phone to complete a telephone interview. For beneficiaries who did not respond to the mail survey and who could not be reached by telephone, in-person interviews were attempted. To minimize costs, study staff conducted in-person interviews only in the 3 PSUs located near RTI headquarters (Richmond, Raleigh, and Atlanta).

Collection of Reinterview Data. As noted earlier, a subset of beneficiaries was selected from those same 3 PSUs to participate in a reinterview study. The beneficiaries who were involved in that part of the research were contacted and reinterviewed approximately 4 weeks after the initial interview. Field personnel conducted face-to-face interviews with beneficiaries who had completed either a mailed questionnaire or a telephone interview.

Data Analysis

Data analyses were designed to assess whether a large-scale survey combining mail survey procedures with telephone follow up of nonrespondents could produce an adequate rate of response for the population at large and for key subpopulations. Data analyses were also designed to assses whether the resultant measurements would be valid and reliable indicators of the health status of the targeted populations. All data analyses employed weights to adjust for unequal probabilities of selection and nonresponse (see above) so that our results may be projected to the sampled universes: (1) new Medicare enrollees; and (2) Medicare Beneficiaries in the age range, 76 to 80. Percentages reported in this paper are, thus, estimates of the percentages that would be found if the survey were administered to all members of the universe.

Results

Response Rates by Survey Stage and Survey Design

Slightly more than one-third (35.8%) of all beneficiaries responded to the first mailing (Table 2). Each subsequent mailing yielded a sizable increment in response rates. Among the entire sample, the second mailing increased the response rate by 14 percentage points, and the third stage increased it by 9 percentage points. After 3 mailings, more than one-half of the beneficiaries (58.9%) had responded. The addition of telephone interviews for nonresponders increased the response rate substantially to 75.2% across the 3 questionnaire lengths. The smaller study embedded in the field survey, which employed in-person contact for beneficiaries in three PSUs who did not respond to mail or telephone contact, showed a further increase. Among that small sample, the overall response rate increased by an additional 5 to 7 percentage points.

The length of the survey questionnaire had a statistically significant effect on response rates (Table 3). Beneficiaries assigned to the long form were less likely to respond (a response rate of 69.4%) after both mail and telephone contact compared with those assigned to the short (77.2%) and medium-length (79%) forms. In contrast, prenotification through the use of survey "lead letters" had an insignificant effect on rates of response. Overall, the response rate after mail and telephone contacts was 74.4% with prenotification and 76.1% without it.

TABLE 2. Cumulative Response Rates by Stage of Data Collection and Age Cohort

Cohort	Stage of Data Collection			
	Mail-1	Mail-2	Mail-3	Telephone
65-year olds	42.0	54.9	62.4	77.8
(s.e.)	(2.2)	(2.3)	(2.2)	(1.9)
76- to 80-year-olds	33.0	47.5	57.3	74.1
(s.e.)	(1.4)	(1.7)	(1.5)	(2.0)
Both cohorts	35.8	49.8	58.9	75.2
(s.e.)	(1.2)	(1.4)	(1.4)	(1.5)

Note: Estimates of standard errors use algorithms that take account of weighting and complex sample design (Shah, Barnwell, and Bieler, 1996).

Variation in Response by Subpopulation

Overall, the foregoing results indicated that a questionnaire that took as long as 30 minutes (medium length) would be completed by approximately 80% of Medicare beneficiaries in both age cohorts after mail and telephone contacts were completed. Those response rates appeared to be unaffected by prenotification of the impending survey; however, rates varied considerably across subpopulations, and the subpopulations were differentially affected by including a telephone follow-up stage of data collection (Table 3). After the third mailing, the rate of response to mailed questionnaires was 19 percentage points lower for nonWhites (42.5%) than for Whites (61.1%). However, when telephone interviews were added, the 19 percentage-point difference in response rates across racial groups declined to less than 6 percentage points (70.1%–75.9%). That difference, although attenuated, remains statistically significant. Urban residents also had lower rates of response for both mail (55.6%–66.2%) and mail-plus-telephone interview stages (71.7%–83.7%) compared with beneficiaries living in nonmetropolitan areas. Similarly, when both mail and telephone interview stages were complete, response rates in the Northeast region were 9 to 10 percentage points lower than those in the rest of the country.

The MBHSR Field Test sample was designed so that there would be equal numbers of target respondents (age range, 76–80) from the strata without health conditions and from the strata with 1+ conditions (Table 1). Among the 915 MBHSR Field Test participants in this age cohort who responded during the mail and telephone phases of data collection, 439 (48%) did not have any of the 15 conditions or procedures, 198 (22%) had 1 condition, 156 (17%) had 2, 81 (9%) had 3, 23 (3%) had 4, 12 (1%) had 5, and 5 (0.5%) had 6 conditions. (Ns and percentages are unweighted.) The data suggest that among survey respondents, those with more conditions were more likely to respond to mail solicitations than those with fewer conditions. The weighted proportions *of survey respondents* that required telephone follow up were: 24.1% (0 conditions); 25.7% (1 condition); 15.8% (2 conditions); and 19.2% (3+ conditions).

Questionnaire and Item Completion

A review of the questionnaires indicated that respondents generally provided usable responses to most questions. When a "completed questionnaire" is defined as one that contained usable substantive responses (excluding "don't know" answers, refusals, etc.) to at least 80% of 34 key questions for women and 32 key questions for men, we found that 98.6% of the 65-year-old cohort and 96.5% of the 76- to 80-year-old cohorts who returned a mail survey "completed" their questionnaires. Rates were equally high for telephone interviews (98.4%–99.3%, respectively) and were relatively stable across racial groups, gender groups, and different regions of the country.

Overall, rates of response for *individual* questions, which includes nonkey questions, were high. For example, questions on past surgeries, disease prevention activities, nutrition, medical conditions, and disabilities had average item response rates of 98% percent or higher. (Rates are averaged across questions of a particular type using those survey questions that were included in each of the 3 questionnaire lengths.) We found

TABLE 3. Survey Response Rates by Survey Stage, Age Cohort, and Survey Design Factors

	All Respondents				Age 65				Ages 76–80			
	Mail		Mail + Telephone		Mail		Mail + Telephone		Mail		Mail + Telephone	
Survey Design Factors	Response Rate	(s.e.)	Response Rate	(s.e.)	Response Rate	(s.e.)	Response Rate	(s.e.)	Response Rate	(s.e.)	Response Rate	(s.e.)
Total sample	58.9	(1.44)	75.2	(1.50)	62.4	(2.23)	77.8	(1.92)	57.3	(1.45)	74.1	(1.99)
Questionnaire length												
Short	63.0	(2.69)	77.2	(2.27)	70.9	(2.40)	82.0	(1.66)	59.4	(3.78)	75.0	(3.66)
Medium	62.5	(1.84)	79.0	(1.94)	61.6	(3.60)	79.6	(3.21)	62.9	(1.96)	78.8	(2.31)
Long	51.0	(2.12)	69.4	(2.03)	54.8	(3.50)	71.9	(2.44)	49.2	(2.02)	68.2	(2.87)
Pre-notification												
Had prior notification	57.2	(2.65)	74.4	(2.04)	62.4	(2.77)	77.6	(2.56)	54.9	(3.36)	72.9	(2.75)
No prior notification	60.5	(2.10)	76.1	(2.11)	62.4	(2.82)	78.0	(2.14)	59.7	(2.70)	75.3	(2.84)
Race												
White	61.1	(1.43)	75.9	(1.59)	64.4	(2.49)	78.4	(2.21)	59.6	(1.53)	74.9	(2.12)
Nonwhite	42.5	(2.79)	70.1	(1.77)	48.3	(2.76)	73.7	(2.37)	39.7	(3.83)	68.3	(2.55)
Gender												
Male	61.0	(2.19)	75.3	(1.90)	61.6	(2.67)	77.3	(2.14)	60.7	(2.62)	74.2	(2.73)
Female	57.4	(1.89)	75.2	(1.90)	63.0	(2.98)	78.2	(2.49)	55.1	(2.09)	74.0	(2.44)
Health status												
1+ Conditions (age 76–80 only)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	60.8	(3.07)	77.2	(2.25)
0 Conditions (age 76–80 only)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	54.7	(2.14)	71.8	(2.42)
Metropolitan area												
In an MSA	55.6	(1.88)	71.7	(1.80)	59.1	(2.97)	74.8	(2.56)	54.0	(1.90)	70.3	(2.33)
Not in an MSA	66.2	(1.93)	83.1	(2.02)	69.9	(3.13)	84.5	(1.94)	64.5	(1.78)	82.4	(3.28)
Region												
Northeast	56.6	(4.71)	67.8	(5.15)	59.2	(4.00)	72.3	(5.61)	55.6	(5.38)	66.0	(5.69)
North Central	61.1	(1.83)	78.0	(2.33)	66.1	(4.07)	81.2	(2.40)	58.6	(2.39)	76.4	(4.58)
South	53.8	(3.25)	77.4	(3.28)	57.0	(3.93)	79.2	(3.56)	52.2	(3.49)	76.6	(3.42)
West	64.6	(3.27)	76.9	(3.15)	67.0	(4.97)	76.1	(5.81)	63.5	(2.74)	77.2	(2.19)

Note: Respondents were randomly assigned to experimental treatments that (1) varied the length of the survey questionnaire (short, medium, or long) and (2) used or did not use a prenotification letter before the survey was mailed. Estimates of SE use algorithms that take account of weighting and complex sample design (Shah, Barnwell, and Bieler, 1996).

somewhat lower rates for questions asking about alcohol use, the average item response was 93%, socioeconomic status (93%), and income and assets (88%). Neither the length of the questionnaire nor prenotification had a substantial effect on item nonresponse rates.

Proxy Responses

Special analyses assessed the need for and impact of proxy responding in the MBHSR field test. For those analyses, beneficiaries who completed the interview on their own or who had someone else merely *write down* answers that they dictated were considered to have responded for themselves. Beneficiaries for whom someone else completed the questionnaire or who had proxy involvement other than having someone simply write down "dictated" answers were considered to have responded by proxy.

Data on the use of proxy respondents show that the older cohort was significantly more likely than the younger to rely on proxies (Table 4). Proxies completed questionnaires that were returned by approximately 15% of persons aged 76 to 80 years but only by 2% of the 65 year olds. Among the older cohort, children (48%) and spouses (32%)

TABLE 4. Percent of Sample Whose Questionnaires Were Completed by Proxies, by Age, Demographic Subgroups, and Survey Format

	Age 65		Ages 76–80	
Subgroups	% Proxy	(s.e.)	% Proxy	(s.e.)
Total	2.1	(0.37)	14.7	(1.67)
Race				
White	1.3	(0.37)	13.7	(1.88)
Non-White	7.7	(2.91)	22.2	(2.01)
Gender				
Male	2.8	(0.84)	13.1	(2.73)
Female	1.6	(0.54)	15.6	(1.75)
Health status (ages 76–80)				
1+ Medical events	n.a.	n.a.	15.4	(1.86)
0 Medical events	n.a.	n.a.	14.2	(2.07)
Reside in metropolitan area				
MSA	2.3	(0.53)	16.9	(2.27)
Non-MSA	1.6	(0.45)	10.3	(2.16)
Region				
Northeast	3.6	(1.63)	14.2	(3.26)
North Central	0.5	(0.39)	11.1	(2.04)
South	2.4	(0.37)	19.0	(3.65)
West	3.0	(0.64)	14.7	(5.21)
Questionnaire length				
Short	2.4	(0.89)	15.0	(2.67)
Medium	1.8	(0.77)	13.7	(3.21)
Long	2.1	(0.99)	15.4	(2.27)
Notification				
Prior notification	1.9	(0.71)	12.7	(1.88)
No prior notification	2.4	(0.50)	16.7	(2.38)
Mode of interview				
Mail	1.4	(0.49)	10.3	(1.30)
Telephone	5.1	(1.63)	29.1	(3.65)

Note: The table includes respondents to the mail and telephone interviews. Respondents were considered to have responded for themselves if they either completed the questionnaire by themselves or had someone else write down answers the respondent supplied; all other respondents were considered to have responded by proxy. Of 972 new-enrollee responses, 28 were provided by proxies, and of the 916 older enrollee responses, 138 were provided by proxies. Entries in the table are weighted estimates. Estimates of standard errors use algorithms that take account of weighting and complex sample design (Shah, Barnwell, and Bieler, 1996).

served most frequently as proxies. Reliance on nonrelatives (9%) and siblings (3%) was less common.

The use of proxy respondents was greater among nonWhites than among Whites for both age cohorts (7.7% vs. 1.3% for 65-year-olds and 22.2% vs. 13.7% for the 76- to 80-year-old cohort). Variations in the length of the questionnaire, prenotification, and modes of interview had no substantial effect on the use of proxies by beneficiaries in the younger cohort (Table 4). The same was not true for the 76- to 80-year-old cohort. Among 76- to 80-year-olds, the longer form resulted in a slightly higher rate of proxy response (19%) as compared with the other two lengths (14.2% and 11.1% for the short and medium-length forms). Moreover, the use of proxies was most likely to occur on telephone recontacts rather than on the earlier mail survey attempts (29.1% vs. 10.3%).

In general, item completion rates were high when proxy respondents completed the questionnaires, although there were a few exceptions. Questions related to chest pain or discomfort had lower completion rates among female beneficiaries who relied on proxies. Similarly, for men, questions about difficulties with urination had

lower rates of completion when they were answered by proxy respondents.

Reliability of Measurements

A variety of analyses were conducted to assess the reliability of measurements made in the MBHSR field test. Below, we report the results of 2 major types of analyses involving assessments of the internal consistency of MBHSR multiitem measurements and correlation over time of measurements made in the MBHSR. For those analyses, we focused on measurements in 7 areas, as noted in the following sections.

Activities of Daily Living (ADL). The MBHSR's ADL scale was composed of 8 questions based on work by Katz[14] and Spector et al. Respondents were first asked to rate the difficulty they had performing the following 4 activities: dressing or undressing, getting out of bed, feeding oneself, and bathing or showering. The response categories were as follows: (1) unable to perform the activity; (2) a lot of difficulty; (3) some difficulty; or (4) no difficulty. Respondents were then asked whether they needed help doing each of the 4 activities. For these questions, the answer categories were (1) yes and (2) no.

Instrumental Activities of Daily Living (IADL). The IADL scale constituted 9 questions, which were based on work by Katz[14] and Spector et al[15]) related to the following four activities: shopping, handling money, using the telephone, and doing light housekeeping. Respondents were asked to first rate the difficulty they had performing the activities and whether they needed help from another person to perform them (plus one other). (Response categories were the same as for the ADL scale.) A final item in the IADL scale asked respondents to indicate how much help "you need from another person to take your medicine (in the right doses and at the right time)." The following four response categories were given: (1) cannot take medicine without help; (2) need some help; (3) do not need help; or (4) do not take any medicines. For analytic purposes, categories 3 and 4 were collapsed for this question.

Mobility (MOB). The MOB scale, once again based on work by Katz[14] and Spector et al[15] had four items that assessed a respondent's difficulty in walking and using transportation. The questions asked respondents the following questions: how far they could walk without resting; how many flights of stairs they could climb without resting; how much difficulty they had getting places by car, taxi, public transportation, and so forth; and whether they needed help from another person to get to places that were out of walking distance.

Mental Health Index (MHI-5). The MHI-5 scale in the field test was based on work by Ware and Sherbourne[11] and contained 5 items that asked respondents about their experiences in the preceding 30 days. In particular, respondents were asked how much of the time they "were a very nervous person," "felt so down in the dumps nothing could cheer [them] up," "felt calm and peaceful," "felt down or blue," and "had been a happy person." Responses were coded as follows: (1) all or most of the time; (2) some of the time; (3) a little of the time; or (4) none of the time. To create a scale in which all items were coded in the same direction (ie, high scores indicated a nervous, unhappy respondent), responses to the questions about feeling calm and feeling happy were coded in reverse of the questionnaire ordering (ie, response 1 was recoded to "felt calm or peaceful" or "been a happy person" *none* of the time).

CAGE Alcohol Problems Scale. The MBHSR questionnaire included the four-item CAGE scale. Those questions asked a respondent whether (1) "you ever felt you should cut down on your drinking"; (2) "people annoyed you by criticizing your drinking"; (3) "'you ever felt bad or guilty about drinking"; and (4) "you [have] ever taken a drink first thing in the morning to steady your nerves or get rid of a hangover." Response categories were either (1) yes or (2) no. Respondents who reported in an earlier question that they had not had a drink in the past 30 days were not asked the CAGE questions. This group would include people with a history of alcoholism or alcohol abuse who were abstaining from alcohol at the time of the survey.

General Health. This scale was composed of 5 questions derived from the SF-36 developed by Ware and Sherbourne.[11] The questions included a global health item that asked respondents to rate their overall health (response categories: very good, good, fair, and poor), and a second question that asked respondents how their current health compared with their health 1 year ago (response categories: worse; about the same; or better. (During analysis, codes for the general health question were reversed [ie, 4 equaled very good health, etc.], so that for all questions, higher scores corre-

TABLE 5. Standardized Alpha for Internal Reliability of Selected Scales from Main MBHSR Data Collection, by Sample Cohort (Unweighted Ns for Each Analysis Are Shown in Parentheses)

Scale	Items	Age Cohort		
		Both	65	76–80
Activities of daily living (ADL)	8	0.905	0.846	0.908
		(540)	(286)	(254)
Instrumental activities of daily living (IADL)	9	0.928	0.884	0.927
		(496)	(265)	(231)
Mobility	4	0.811	0.727	0.800
		(532)	(283)	(249)
General health (from SF-36)	5	0.779	0.754	0.780
		(1,165)	(600)	(565)
Mental health index (MHI-5)	5	0.860	0.855	0.860
		(1,790)	(940)	(850)
Prostate symptoms	9	0.870	0.863	0.870
		(441)	(251)	(190)
Alcohol problems index (CAGE)*	3	0.597	0.513	0.802
		(203)	(133)	(70)

*Scale excludes item asking whether respondents have felt that they should cut down on their drinking; this item had a poor correlation with other items. Note that the CAGE items were not asked if respondent reported that he or she did not have a drink in the past 30 days. Thus, respondents who had had serious alcohol problems in the past and had stopped drinking would not be asked these scale items. Note also that the unweighted estimate of internal consistency for CAGE items is 0.74; application of sampling weights reduced that estimate to 0.60.

sponded to better health. Note also that a fifth category ["excellent"] used in the SF-36 wording of this question was not used in the MBHSR Field Test.) Three additional questions asked how often physical health problems, emotional health problems, and pain had interfered with respondents' daily or social activities in the past 30 days (response categories: all or most of the time; some of the time; a little of the time; or none of the time).

Prostate Symptoms Index. A battery of items on prostate symptoms was included because prostate cancer and benign prostate disorders are major sources of morbidity and mortality among older men in the United States. The nine questions in the Prostate Symptoms Index were based mostly on the American Urological Association scale. Six questions asked male respondents how often during the past 30 days they had: felt like their bladder was not empty after urinating; had urinated within 2 hours of prior urination; had stopped and started multiple times while urinating; had difficulty postponing urination; had a weak urinary stream; and had to push or strain to begin urinating (response categories: almost every time; more than half the time; about half the time; less than half the time; less than 1 in 5 times; or never). Three additional questions asked respondents how many times per night they had usually gotten up to urinate during the past 30 days, whether they had *ever* been unable to urinate at all, and whether a doctor had ever told them they had a prostate problem.

Internal Consistency. To assess the internal consistency of the scale measurements, we calculated Cronbach's alpha and standardized alphas for each of the scales. (Standardized alphas, unlike Cronbach's alpha, are not affected by the number of items included in a scale.) Table 5 presents the standardized alphas for the whole sample and separately for the 65-year-old and 76- to 80-year-old cohorts. Reliabilities for ADL and IADL scales exceeded 0.90 for the whole sample and 0.84 when the cohorts were analyzed separately. The internal consistency of the Prostate, Mental Health, Mobility, and General Health measures was also impressive (range, 0.78–0.87 for the whole sample). Again, only minor variations were

TABLE 6. Product Moment Correlation From T-1 to T-2 in Scale Scores for Reinterview Sample

SCALE	Items	n	T-1 to T-2 Correlation
Activities of daily living (ADL)	8	108	0.918
Instrumental activities of daily living (IADL)	9	89	0.958
Mobility	4	103	0.940
General health (from SF-36)	5	238	0.674
Mental health index (MHI-5)	5	353	0.643
Prostate symptoms*	7	87	0.407

Note: Alcohol problems index not shown because sample size was only 33 for persons completing items at Time 1 and Time 2.

*Excludes questions on ever being unable to urinate and doctor's diagnosis. Those two items had relatively low correlations with the overall scale (.17 and .35).

seen between the internal reliabilities obtained for the 65-year-old and 76- to 80-year-old cohorts.

Our initial analyses yielded quite low values for the CAGE scale of alcohol problems. Further investigation of responses to the scale items indicated that the initial question that asked whether the respondents ever felt they should cut down on their drinking detracted from the scale's reliability. For that reason, we excluded the initial item and restricted our analyses to a 3-item CAGE scale. However, the three-item version of the CAGE index is a lackluster performer in terms of internal reliability. Its standardized alpha is 0.60 in the combined sample.

Reliability of Measurements Over Time. To assess the reliability of the MBHSR scales over time, we constructed scales for the main data collection (time 1, or T-1) and the reinterview study (time 2, or T-2) using only those respondents who participated in both phases of the study. Initially, we worked with the MHI-5 scale and performed a principal-components factor analysis to extract a single factor. We used the factor scores, means, and SDs from this analysis of T-1 data to score both the T-1 and T-2 measurements. (The purpose of this procedure was to keep our reliability assessments from capitalizing on random perturbations in the response structure between T-1 and T-2.) We then calculated the correlation between scale measurements for respondents who participated in both the T-1 and T-2 phases of the study. This analysis produced no noteworthy differences when compared with a parallel analysis that transformed each item to a standardized score and constructed a scale treating each item as equally weighted. For that reason, we used the second, simpler procedure for the analysis that we present below. This procedure gives equal weight to each item in a scale. Table 6 shows the correlation between T-1 and T-2 scores for 6 of 7 MBHSR scales shown in Table 5. (Results for the CAGE index of alcohol problems are not shown because only 33 respondents completed both the T-1 and T-2 CAGE measurements.) Although the samples for the scales were small (n = 87–108), the results indicate substantial reliability over time in the measurements of the ADL, IADL, and MOB scales. The cross-temporal correlations for all 3 scales exceed 0.90. (It should be note that distributions for all 3 measures were highly skewed, with the majority of respondents reporting no difficulty or no need for help in response to all items included in a scale. Nonparametric analysis of those data, however, sustained the conclusion that high cross-temporal consistency existed in the measurements.[1] Results for the General Health and MHI-5 indices are based on a larger number of cases (n = 238 and 353, respectively); they also evidence substantial although not equivalent cross temporal reliability (0.67 and 0.64). We suspect that the apparent decrement in cross-temporal reliability for the General Health and MHI-5 indices may reflect the more ephemeral nature of these traits. There is good reason, for example, to believe that "being happy" may be a more changeable phenomenon than needing assistance in bathing or dressing. The index of prostate symptoms used in the MBHSR shows quite low cross-temporal correlation (0.41) despite its high internal reliability (0.87). Measurement unreliability may contribute to that finding; however, we suspect that it may also reflect the innate variability of

some prostate symptoms over the 4-week time interval between our measurements.[1]

Discussion

One major reason for conducting the MBHSR field test was to determine whether a mailed, self-administered questionnaire with follow-up telephone interviews for nonrespondents could achieve acceptable response rates among both younger and older Medicare beneficiaries. The results of the field test indicate that a mailed questionnaire with telephone follow up of nonrespondents yielded response rates of 80% for the 65-year-old cohort and 79% for the 76- to 80-year-old cohort with a questionnaire that took approximately 30 minutes to complete.

Those rates approach, but do not equal, response rates obtained in other large-scale surveys of the elderly that used more expensive in-person interviewing. For example, the Medicare Current Beneficiary Survey (MCBS), achieved a response rate of 87% in its initial round of in person interviews, a rate that is 7 to 8 percentage points higher than that achieved with the medium-length form in the field test. But the increase in response comes at a price. Field costs for an in person survey such as the MCBS are many times higher than those for the hybrid "mail with telephone follow up" design proposed for the MBHSR. Planning future data collection efforts raises the question of whether the increased response rate associated with in-person interviews would justify their use for the registry.

As a general strategy for data collection, we think not. First, the increment in response rates would not be overly large. In the field test, we found that a final phase of in-person data collection increased response rates by only 5 percentage points. Second, if the MBHSR sample is drawn from Medicare enrollment files, those files can be used to learn a great deal about nonrespondents. That information, in turn, becomes an important tool for probing nonresponse bias.

We do believe, however, that in-person interviews may be justified for a small, randomly selected subset of MBHSR nonrespondents. A circumscribed effort of that sort to assess the impact of nonresponse could focus on information that is not routinely captured in Medicare files. Similarly, we note concerns that respondents report symptoms differently when interviewed by mail than by telephone. A mixed-mode design, such as that proposed for the MBHSR, will potentially confound any mode effects on response with other factors correlated with nonresponse to initial mail solicitations. Both bias in prevalence estimates for subpopulations that differ in their speed of response and spurious correlations could, thus, be introduced into the resultant data. An appropriately designed experiment that randomly assigned a small subsample of MBHSR respondents to receive telephone interviews in early waves, rather than mail surveys, would permit the identification of any such biases and recalibration of the overall MBHSR estimates to take account of their impact.

Multiple Waves of Follow Up. As the MBHSR survey designers anticipated, each wave of contact recruited additional beneficiaries. However, in the case of mailed questionnaires, each successive wave recruited a smaller percentage of respondents. If one assumes that people who responded to the second and third mailings would have responded to the telephone interview, the cost of telephone interviews after the first or second mailing would have to be weighed against the time lost and administrative effort spent on multiple mailings. And those costs may not be equal for all subgroups of beneficiaries. As noted earlier, response rates for nonWhite beneficiaries after 3 mailings (39.7% in the older cohort and 48.3% in the younger cohort) clearly were not adequate. Adding telephone interviews substantially increased the representation of nonWhites in both age cohorts, yielding response rates of 74% and 68% for the 65-year-old and 76- to 80-year-old cohorts, respectively. Thus, telephone follow up of nonrespondents to the mail survey is particularly crucial to ensure adequate rates of participation among non-White respondents.

Design Alternatives. Data from the field test indicated that prenotification of respondents had no demonstrable positive effect on response rates. Consequently, such results argue against including it in the full implementation of the MBHSR. The elimination of prenotification should reduce operational costs. Among the other alternatives that were assessed in the field test, the length of the questionnaire had an effect on response rates, as evidenced by the uniformly lower rates of response among respondents who received the long questionnaire. However, given the similar response and completion rates for the medium-length and short forms, a reasonable assumption

is that a questionnaire of approximately 100 items could be used without jeopardizing the quality of the data. Interestingly, the length of the questionnaire did not appear to affect questionnaire completion rates, item response rates, or the reliability of the information provided. The lack of effect in those areas indicates that most beneficiaries who begin the process of responding to the instrument will continue to the end. Our results (eg, Table 4) also make it clear that proxy response will be a crucial component in the MBHSR. Several encouraging findings from the field test support the belief that members of the 76- to 80–year-old cohort can provide high-quality data, although people of that age require proxy respondents more frequently than do younger individuals. First, the results show that after telephone interviews had been completed, relatively small differences in response rates were apparent across the two age cohorts. Second, the increment added to the response rate by telephone interviews was essentially the same in the 2 age cohorts (15.4 percentage points in the younger cohort vs. 16.8 percentage points in the older cohort). Third, the overall quality of data provided by proxy respondents was good. Indeed, excluding proxy reports would, however, result in underrepresentation of less healthy beneficiaries, a group that is likely to report more health problems, more utilization of health care services, and worse functional status. Overall, results from the field test indicate that the age of Medicare beneficiaries does not present a significant problem in collecting high-quality data from that population, at least not through the age of 80, if adequate provision is made for the use of proxy respondents.

Validity of Measurements. Previously[7] we have shown that the MBHSR instrument is relatively sensitive in assessing the 12 medical events that could be checked against Medicare records for the 76- to 80-year-old cohort. That result reinforces our conclusion that most respondents did not have serious problems understanding questions related to medical procedures and diagnoses, and they could accurately report on them.

Measurement Reliability. The level of reliability of the MBHSR measurements was generally high in both age cohorts. Measurements from the widely used scales of physical functioning (eg, Activities of Daily Living, Instrumental Activities of Daily Living, and Mobility) evidenced good internal reliability for the whole sample at the initial measurement (alpha = 0.8–0.9). Internal reliability was equally great for the MBHSR's measures of mental health (the MHI-5 scale), general health, and prostate symptomatology. The one exception to these generally encouraging findings was for the CAGE index of alcohol problems. Even after eliminating 1 item to improve the measurement's reliability, the overall reliability of the CAGE was only 0.61 for the whole sample, and it declined to 0.51 for respondents in the 65-year-old cohort. Those results suggest that the CAGE index will not be a reliable indicator of alcohol problems in the elderly population. Repeating measurements (excluding the CAGE) approximately 30 days after the original assessments produced substantial cross-temporal reliability. Correlations over time for the ADL, IADL, and MOB scale measurements exceeded 0.9. Those for the MHI-5 and General Health scales were somewhat attenuated (range, 0.6–0.7). We believe the attenuation accurately reflects the greater cross-temporal variability in mood and global judgments about one's health status that these scales evoke versus the actual physical impairments measured in the ADL and related scales. Our scale of prostate symptoms showed an even higher level of cross-temporal instability. Given the very high level of internal consistency (0.87) obtained in measurements with the prostate scale at the initial data collection, we suspect that this result may reflect the transitory nature of many prostate symptoms. Results for the MBHSR's multiitem scales were consistent with results of detailed analyses of consistency of response to individual items.[13,24,25] Overall, we found that with few exceptions, the responses given to the MBHSR evidenced levels of internal and cross-temporal reliabilities that were more than satisfactory.

Conclusion

If the Medicare Beneficiary Health Status Registry is implemented, it will not be just "another survey" of the elderly. It will differ from other major federal surveys in three important ways:

- MBHSR samples drawn each year will be sufficiently large to permit analysis of even relatively sparse subpopulations of Medicare beneficiaries;
- The MBHSR will be longitudinal to allow tracing of changes within cohorts of beneficiaries as well as changes between cohorts over time; and

- MBHSR questionnaire data can be linked to Medicare claims data, thereby enabling analyses that link historical data on the use and costs of medical services to measurements of beneficiaries' health status, medical history, and risk factors. Such analyses are crucial to the improvement of our understanding of the impact of treatment patterns on the health of the elderly.

The full value of data gathered with the MBHSR questionnaires will be realized only after the information has been linked to the administrative data collected by HCFA and then used by program planners, policymakers, and researchers within and outside government. In parallel with the field test of the registry, HCFA has developed a system to build analytical files composed of the information collected in registry surveys and Medicare's administrative data. The MBHSR's annual data on the health status of large probability samples of new enrollees and its longitudinal follow up of older beneficiaries will provide many unique analytic opportunities, particularly given the linkage of the registry to data on claims in the Medicare record system. Some of those possibilities may become apparent in the future; others can be foreseen now, such as:

- continuous monitoring of the health status of the elderly population using a sample that is sufficiently large to detect changes specific to particular socioeconomic, racial, geographic, and other groupings;
- evaluation of regional and other variations in the types of medical treatment obtained by Medicare enrollees with equivalent symptomatology;
- assessment of the relative costs of medical care that will be required over time for people who report different classes of health conditions when they enroll in Medicare;
- development of research methods for studying the impact of patterns of health care on subsequent morbidity and mortality over the long term and estimating the relationship between health care status and subsequent utilization of health care services; and
- provision of a sampling frame for subsequent studies of specialized subgroups of the population aged 65 and older, such as those who had particular health problems before their enrollment in Medicare.

In summary, the results from the MBHSR field test indicate that the registry is feasible and that it can collect data by using an economical strategy that combines mail survey procedures with telephone follow up of nonrespondents. The field test has shown that this design produces adequate response rates and yields data that achieve more-than-satisfactory levels of reliability and validity.

Acknowledgments

The authors benefitted from the advice of a number of scientists including the members of our Technical Advisory Panel: Drs. Ronald M. Andersen, Laurence G. Branch, Lester Breslow, Charles Hennekens, Sidney Katz, Vince Mor, and Floyd J. Fowler. Our advisors are not, of course, responsible for any errors we may have made. The present overview article draws on an unpublished final report by Turner et al[27] of the major findings from the MBHSR field test. With the exception of this note, we do not indicate the numerous places in which we have drawn on that report. A version of this article[1] containing additional technical details can be obtained from the corresponding author.

References

1. **McBean AM, Turner CF, Fitterman LK, Pate K, Reilly TW, Smith TK, et al.** Expanded version of "Monitoring the health status and impact of treatment on older Americans: The Medicare Beneficiary Health Status Registry." Technical Papers on Health and Behavior Measurement, 35-b, Washington, DC: Research Triangle Institute, 1998.

2. **Health Care Financing Administration, Office of Management and Budget Clearance Package for the Pilot Study of the Medicare Beneficiary Health Status Registry.** Washington, DC: Epidemiology Branch, Division of Research, 1992.

3. **National Center for Health Statistics.** Plan and Operation of the 1988–1994 National Health and Nutrition Examination Survey. Vital and Health Statistics, Series 1, No. 32. Hyattsville, MD: National Center for Health Statistics, 1997.

4. **National Center for Health Statistics.** Current estimates from the National Health Interview Survey, 1994. Vital and Health Statistics, Series 10, No. 193. Hyattsville, MD: National Center for Health Statistics, 1996.

5. **Bureau of the Census.** Longitudinal Study of Aging Questionnaire (OMB No. 0920-0219; expiration May 31, 1991). Washington, DC: Bureau of the Census, 1991.

6. **Pate K, Biemer P, Chromy J, Hawes C, Judkins D, Lessler J, et al.** Design Documentation Report for the Medicare Registry. Report prepared under HCFA contract 500-95-0038. Research Triangle Park: Research Triangle Institute, 1996.

7. **Turner CF, Smith TK, Fitterman LK, Reilly T, Pate K, Witt MB, et al.** The quality of health data obtained in a new survey of elderly Americans: A validation study of the Medicare Beneficiary Health Status Registry (MBHSR). J Gerontol Soc Sci 1997;52B(1):S49.

8. **Wheeless SC, Witt MB.** Medicare Beneficiary Health Status Registry Pilot Study: Sample design. Report to the Health Care Financing Agency, July 28, 1993. Research Triangle Park: Research Triangle Institute, 1993.

9. **Folsom RE.** Exponential and logistic weight adjustments for sampling and nonresponse error reduction. Proc Am Stat Assoc (Social Statistics Section), 1991:197.

10. **Deville JC, Sarndalo CE.** Calibration estimators in Survey Sampling. J Am Stat Assoc 1992;87:376.

11. **Shah B, Barnwell BG, Bieler B.** SUDAAN Users manual, release 7.0. Research Triangle Park: Research Triangle Institute, 1996.

12. **Bureau of the Census.** Longitudinal Study of Aging Questionnaire (OMB No. 0920–0219). Washington, DC: Bureau of the Census, 1991.

13. **Bureau of the Census.** Longitudinal Study of Aging Questionnaire (OMB No. 0607–0635). Washington, DC: Bureau of the Census, 1991.

14. **Ware JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36). Med Care 1992;30:479.

15. **CDC.** Behavioral Risk Factor Questionnaire. Atlanta: 1992.

16. **Forsyth BH, Kirk Pate D, Smith TK, Fitterman L.** Reinterview methods for assessing and improving the quality of data from a Medicare population. In: Warnecke R, ed. Health Survey Research Methods: Conference Proceedings. DHHS Pub No (PHS) 96-1013. Hyattsville, MD: National Center for Health Statistics, 1996.

17. **Katz S.** Assessing self maintenance: Activities of daily living, mobility, and instrumental activities of daily living. J Am Geriat Soc 1983;31:721.

18. **Spector WD, Katz S, Murphy JB, Fulton JP.** The hierarchical relationship between activities of daily living and instrumental activities of daily living. J Chronic Dis 1987;40:481.

19. **Mayfield D, McLeod G, Hall P.** The CAGE questionnaire: validation of a new alcoholism screening instrument. Am J Psychiatry 1974;131:1121.

20. **Ewing JA.** Detecting alcoholism: The CAGE questionnaire. J Am Med Assoc 1984;252:1905.

21. **Barry MJ, Folwer FJ, O'Leary MP, Bruskewitz, Holtgrewe HL, Medbust WK, et al.** The Measurement Committee of the American Urological Association. The American Urological Association Symptom Index for benign prostatic hyperplasia. J Urol 1992;148:1549.

22. **Jacobsen SJ, Girman CJ, Guess HA, Rhodes T, Oesterling JE, Lieber MM.** Natural history of prostatism: Longitudinal changes in voiding symptoms in community dwelling men. J Urol 1996;155:595.

23. **Gibson DM, Aitkenhead W.** The elderly respondent: Experiences from a large-scale survey of the aged. Res Aging 1983;5(2):283.

24. **Landry JA, Smmyer MA, Tubman JG, Lago DJ, Roberts J, Simonson W.** Validation of two methods of data collection of self-reported medicine use among the elderly. Gerontologist 1998;28:672.

25. **Brambilla DJ, McKinlay SM.** A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey. Am J Epidemiol 1987;126:962.

26. **Caserta MS, Lund DA, Dimond MF.** Assessing interviewer effects in a longitudinal survey of bereaved elderly adults. J Gerontol 1987;49:637.

27. **Turner CF, Forsyth BH, Wheeless SC, Witt MB, Pate DK, Smith TK, et al.** Field test of Medicare Beneficiary Health Status Registry: Major findings and recommendations. Final report to the Health Care Financing Administration under Contract 500-9-0053. Research Triangle Park: Research Triangle Institute, 1994.

28. **Adler S.** A profile of the Medicare Current Beneficiary Survey. Health Care Financing Review 1994;15:153.