Cognitive Function and the Concordance Between Survey Reports and Medicare Claims in a Nationally Representative Cohort of Older Adults

Fredric D. Wolinsky, PhD, Michael P. Jones, PhD,† Fred Ullrich, MS,* Yiyue Lou, MS,† and George L. Wehby, PhD**

Background: While age-related cognitive decline may affect all stages in the response process—comprehension, retrieval, judgment, response selection, and response reporting—the associations between objective cognitive tests and the agreement between self-reports and Medicare claims has not been assessed. We evaluate those associations using the Survey on Assets and Health Dynamics among the Oldest Old (AHEAD).

Methods: Eight waves of reinterviews (1995–2010) were linked to Medicare claims for 3661 self-respondents yielding 12,313 person-period observations. Cognitive function was measured by 2 episodic memory tests (immediate and delayed recall of 10 words) and 1 mental status test (backward counting, dates, and names). Survey reports on 12 diseases and 4 health services were mapped to Medicare claims to derive counts of concordant reports, underreports, and overreports, as were the numbers of hospital episodes and physician visits. GEE negative binomial and logistic regression models were used.

Results: Better mental status was associated with more concordant reporting and less underreporting on disease history and the number of hospital episodes. Better mental status and delayed word recall were associated with more concordant reporting and less underreporting on health services use. Better delayed recall was significantly associated with less underreporting on the number of physician visits. These associations were not appreciably altered by adjustment for demographic characteristics, socioeconomic status, self-rated health, or secular trends.

Conclusion: We recommend that future surveys of older adults include an objective measure of mental status (rather than memory), especially when those survey reports cannot be verified by access to Medicare claims or chart review.

Key Words: survey reports, medicare claims, cognition, concordance, older adults

(*Med Care* 2015;53: 455–462)

From the Departments of *Health Management and Policy; and †Biostatistics; College of Public Health, University of Iowa, Iowa City, IA.
Supported by PCORI Pilot Project IP2 PI000659.
The authors declare no conflict of interest.
Reprints: Fredric D. Wolinsky, PhD, Department of Health Management and Policy, College of Public Health, University of Iowa, CPHB N211, 145 Riverside Drive, Iowa City, IA 52242. E-mail: fredric-wolinsky@uiowa.edu.
Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

ISSN: 0025-7079/15/5305-0455

Gerontologic and geriatric research often rely on the collection of information on disease history and the use of health services.[1–3] This process is expensive and time consuming, especially for regional or national samples, regardless of whether it involves surveying older adults or reviewing their medical records.[4–6] Moreover, the agreement between (validity of) survey reports and administrative data (eg, Medicare claims) is not well documented.[7–21]

Perhaps the most important concern associated with the validity of survey reports among older adults is the potential influence of age-related decline in cognitive function.[8,18–21] Cognitive function is important because it may affect all 5 stages in the response process—comprehension, retrieval, judgment, response selection, and response reporting.[22–24] Comprehension reflects whether the older adult understands the question being asked, retrieval involves the recall of information relevant to the question, judgment integrates the retrieved information and fills-in any remaining knowledge gaps, response selection maps the retrieved and integrated information into a response option, and response reporting represents the recording (if self-administered) or verbalization (if interviewed) of the response option.[22–24]

Although concerns about the potential effects of age-related declines in cognitive function on the validity of survey reports have been raised for 5 decades,[8,21,25–28] the problem remains empirically unresolved. The 2 main reasons are the general absence of objective cognitive assessments from surveys of older adults and the absence of an appropriate reference standard for comparative purposes.[8,21] Here, we used a nationally representative, prospective cohort study—the Survey on Assets and Health Dynamics among the Oldest Old (AHEAD)—to evaluate the associations of cognitive function with the agreement between survey reports and Medicare claims.[29,30] Cognitive functioning was objectively assessed using 3 tests—a 6-item mental status test taken from the Telephone Interview for Cognitive Status

(TICS-6), and 2 episodic memory tests that use the same 10 words for immediate recall and delayed recall.[31,32] Participants were asked whether a physician had ever told them that they had any of the 12 diseases (or conditions), or if they had used 4 health services in the last 2 years, as well as the numbers of hospital episodes and physician visits in that period.[21] Survey reports were compared with information retrieved from Medicare claims, with each comparison coded as concordant reporting, underreporting, or overreporting by participants relative to the claims. Generalized estimating equation (GEE) methods with negative binomial and logistic regression models evaluated the associations with the 3 cognitive tests, with each participant contributing up to 8 person-period observations.[33,34]

Three hypotheses derived from the literature framed our study.[7–29] The first was that better cognitive function would lead to more concordant reporting and less underreporting, which is the most frequent form of reporting errors.[8,21] The second hypothesis was that although both the immediate and delayed word-recall (episodic memory) tests tap verbal encoding and retrieval strategies, the latter would be more predictive of concordant reporting and underreporting because it provides a better indication of organic cognitive deficits.[31,32] The third hypothesis was that adjusting for demographic characteristics, socioeconomic status, self-rated health, and survey period, all known to be associated with reporting accuracy,[7–21] would only partially reduce the unadjusted associations with the cognitive tests, with the greatest mediation coming from the demographic characteristics, which are often used as proxies for unobserved cognitive function when evaluating the validity of self-reports.[8,17–21]

METHODS

Data

Complete documentation for the AHEAD is readily available online (http://hrsonline.isr.umich.edu) and elsewhere.[29,30] AHEAD was funded by the National Institute on Aging to serve as the longitudinal source of data on the health and retirement patterns of older Americans. Participants were 70 years old or older at their baseline interviews in 1993, and were reinterviewed biennially thereafter. Participants came from 2 sources: household screening conducted during a multistage cluster sampling process, or persons 80 years old or older from the Medicare Master Enrollment File. To ensure a sufficient number of minority elders, African Americans and Hispanic Americans were oversampled by 1.8 times the probability of general selection. The same oversampling rate was applied to Floridians to ensure that statewide estimates for Florida were possible. Survey weights were used in all analyses to account for the oversampling and design effects. The baseline (1993) response rate was 80.4%, yielding 7447 participants. Data from the 1995–2010 reinterviews are used.

We linked survey reports to Medicare claims [the Denominator, Inpatient, Outpatient, and Carrier Standard Analytic Files (SAFs)] for 6730 participants (90%).[21] To facilitate sufficient lookback periods for mapping to the claims data, 911 individuals were excluded who did not complete their 1995 reinterviews (most had died). We further excluded 873 participants who were not continuously enrolled in Medicare fee-for-service plans with both Part A and Part B coverage before their 1995 reinterview. This reduced the analytic sample to 4946 for 1995 or 66%. Participants were censored after 1995 when they were no longer continuously enrolled in Medicare fee-for-service with both Part A and Part B coverage, were lost to follow-up, or died, whichever came first, yielding 17,033 person-period observations on 4791 participants. The final exclusion restricted the sample to self-respondents, because non–self-respondents were not administered the 3 cognitive tests, resulting in 3661 participants contributing 12,313 person-period observations. These data facilitate intraindividual and secular trend analyses of the associations between cognitive status and the concordance of self-reports with Medicare claims in a large nationally representative sample.

Concordance Outcomes

We constructed 3 summary counts separately for the 12-item disease history (range, 0–12 each) and 4-item health services measures (range, 0–4 each) reflecting the number of (a) concordant reporting (agreement) between the survey reports and claims data, (b) survey underreporting (participant did not report what was found in the claims), and (c) survey overreporting (participant reported what was not found in the claims). Thus, the means for the 3 reporting types (underreporting, concordant reporting, and overreporting) must add to the sum of the number of items in the outcome measure (either 12 for disease history, or 4 for health services). The exact wording of the 12 disease history and 4 health service use questions [see table 1 in Wolinsky et al[21]; available online as Supplemental Table 1 (ST1), Supplemental Digital Content 1, http://links.lww.com/MLR/A884] and the corresponding ICD9-CM (International Classification of Diseases, Ninth Edition Clinical Modification) and/or CPT/HCPCS (Current Procedural Terminology/Healthcare Common Procedure Coding System) codes (see table 2 in Wolinsky et al[21]; available online as ST2, Supplemental Digital Content 2, http://links.lww.com/MLR/A885) has been described in detail elsewhere.[21] The 12 disease history questions were whether a physician had ever told the participant that she had arthritis, angina, cancer, congestive heart failure, diabetes, glaucoma, heart attack, heart disease, hip fracture, hypertension, lung disease, or stroke. Evidence of these diseases from the Inpatient SAF required 1 indication after January 1, 1991 (the earliest year of available claims) up until the current interview. Evidence from the Outpatient or Carrier SAFs required ≥ 2 indications at least 2 but no more than 365 days apart with evaluation and management (E&M) codes (indicating that the physician had seen the patient).

The 4 health services questions were whether during the 2 years before the current interview any hospital episode had occurred, whether any physician visit had occurred, and whether the participant had cataract or outpatient surgery. Mapping the ICD9-CM codes for the hospital utilization measure required an overnight (or longer) stay from the

Inpatient SAFs in the 2 years before the current interview. The physician visit measure was similarly based on ICD9-CM and/or CPT/HCPCS codes during the 2 years before the current interview from the Outpatient or Carrier SAFs for visits with E&M codes, as long as that visit was not on the second or subsequent days of a hospital episode. For cataract and outpatient surgery, the appropriate CPT/HCPCS codes were mapped for the 2 years before the current interview, as long as they did not coincide with an overnight hospital stay.

Concordance on the numbers of hospital episodes and physician visits were considered separately from the above counts. Given the amount of concordant reports on the number of hospital episodes, a trichotomous measure contrasted any underreporting and any overreporting with concordant reporting. Because of their greater volume, the number of physician visits was considered concordant if the self-report and claims data were within ± 2 visits. For the number of overreported physician visits, those overreporting by >100 visits were excluded to minimize the distributional skew.

Cognitive Function

Each AHEAD interview included 3 cognitive tests taken from the TICS,[31] for which test-retest reliability, validity, and comparability to the Mini-Mental Status Examination is well established,[32] with high diagnostic validity for the identification of dementia versus nondementia (83.3% sensitivity and 81.6% specificity), and cognitive impairment versus normal cognition (83.3% sensitivity and 78.3% specificity).[33–35] The first test was the immediate word recall which taps episodic memory. Participants were read a set of 10 words and asked to freely recall as many as possible over a 2-minute period. The score was the number of words (0–10) correctly recalled. The second component was the mental status (TICS-6) test which asked participants to count backward from 20 (1–2 points if it took 2 or just 1 attempt), orient themselves to time (month, day, year, and day of the week; 0–4 points), and name (a) the tool that cuts paper (scissors; 1 point), (b) the prickly plant found in the desert (cactus; 1 point), and (c) the current President and Vice President of the United States (1 point each). Thus, mental status scores ranged from 0 to 10 points (correct answers). The third test was delayed word recall which asked participants to freely recall as many of the 10 words from the earlier list as possible over the next 2-minute period. It was also scored as the number of correctly recalled words (0–10). Over time, modest changes occurred to these 3 cognitive tests to minimize practice effects. These involved 4 alternative 10-word lists that were randomly assigned at each follow-up interview, changing the starting number for counting backward, and substituting alternative objects for the scissors and cactus naming questions.

We use the 3 cognitive tests separately because 4 prior psychometric analyses of these data[36–39] identified 2 factors—episodic memory (the immediate and delayed word-recall tasks) and mental status [the TICS-6 tasks (plus serial 7s subtraction, which are not used here due to missing data)]. Principal components analysis of the data used here revealed the same simple structure with the 2 word-recall tests loading (0.92 and 0.92, respectively) on the former, and the 3 mental status (TICS-6) elements (backward counting, dates, and names) loading (0.61, 0.83, and 0.52, respectively) on the latter, with the factor intercorrelation being 0.30. The eigenvalues were 2.0 and 1.1 before rotation, and 1.9 and 1.5 after rotation. Coefficient α was 0.63. A principal components analysis using just the backward counting, dates, and names elements produced a simple 1-factor solution, with an eigenvalue of 1.4, explaining 48.1% of the variance in the 3 data elements. The factor loadings were 0.69, 0.67, and 0.72, respectively. The ratio of the first eigenvalue to the second was 1.8, and coefficient α was 0.46. These results support the factor structure of the 3 measures, especially in terms of concordant and discriminant validity. The low α results from only 3 data elements (backward counting, dates, and names) with narrow observed ranges.

Covariates

To obtain the net associations of each of the 3 cognitive tests with the concordance between self-reports and claims data, we used fixed and time-dependent covariates for demographic characteristics (age, sex, race, and marital status), socioeconomic status (education, wealth, and MediGap coverage), and self-rated health, memory, and depressive symptoms as considered in previous studies.[7–21] The demographic characteristics included centered age and its quadratic counterpart along with indicator variables for men, African Americans and Hispanic Americans (non-Hispanic whites were the reference), and being married. Socioeconomic status included indicator variables for elementary and college education (high school was the reference), the lowest and highest wealth quintiles (the 3 middle quintiles were the reference), and having MediGap insurance coverage. Self-rated health and self-rated memory had 5 response options (poor to excellent), and depressive symptoms was the number of symptoms endorsed from a list of 8. Seven indicators captured secular trends associated with the timing of the reinterviews (1998, …, or 2010; 1995 was the reference).

Analysis

The associations of the 3 cognitive tests with the 3 summary (concordant reporting, underreporting, and overreporting) counts were estimated separately for the 12-item disease history and 4-item health services measures with SAS version 9 (SAS, Cary, NC) using GEE negative binomial models[40,41] with m-dependent correlation structures and m set at 7 for the number of reinterview waves after 1995. Because SAS (and other commercially available statistical packages) can only estimate GEE multinomial logistic regression under the restrictive assumption of ordered categories (which is not appropriate here), GEE binomial logistic regression was used to evaluate the associations of the 3 cognitive tests by contrasting those who concordantly reported the number of hospital episodes first with those who underreported, and then by contrasting those who concordantly reported the number of hospital episodes with those who overreported the number of hospital episodes. Similarly, GEE binomial logistic regression was used to evaluate the associations of the 3 cognitive tests by contrasting those who concordantly (± 2 visits) reported the number of physician visits first with those who underreported, and then by contrasting those who concordantly reported the number of physician visits with those who overreported. Finally, among those underreporting the number of physician visits, GEE negative binomial regression was used to estimate the associations of the 3 cognitive tests with the number of physician visits that were underreported, with similar models among those overreporting the number of physician visits.

RESULTS

Descriptive

Table 1 contains descriptive statistics across all person-period observations for the concordance measures, the 3 cognitive tests, and the demographic factors, socioeconomic status, health status, and survey period covariates. Most reports on the 12-item disease measure were concordant (mean = 10.2), with few underreports (mean = 1.2), and even fewer overreports (mean = 0.6). Concordant reporting (mean = 3.5) was also more likely on the 4-item health services measure than underreporting (mean = 0.3) or overreporting (mean = 0.2). Three fourths of the participants had concordant reports on the number of hospital episodes, with 12.5% underreporting and 11.7% overreporting. Concordance on the number of physician visits (± 2 visits) was observed for only 19.2%, with 68.1% underreporting (median = 10 visits), and 12.4% overreporting (median = 8 visits). The mean on the mental status test was 9.1, the mean on the immediate word-recall test was 4.6, and the mean on the delayed word-recall test was 3.3. Mean age was 81.5, 35% were men, 9.7% were racial or ethnic minorities, and 39% were married. Sixteen percent had elementary and one third had college educations. Nearly three fourths had MediGap coverage. Self-rated health and self-rated memory were "good" (means = 3.0), and the mean number of endorsed depressive symptoms was 1.6.

GEE Negative Binomial Models of the 12-Item Disease and 4-Item Health Services Concordance Measures

Table 2 contains the incident rate ratios (IRRs) from the GEE negative binomial models for concordant reporting (top panel), underreporting (middle panel), and overreporting (bottom panel) on the 12-item disease measure, and Table 3 contains similar information on the 4-item health services measure. The IRRs reflect the ratio of the counts of the concordant reports, underreports, or overreports per 1-point improvement on each of the 3 cognitive tests. For example, the IRR of 0.958 for the mental status test in model 1 of Table 2 for underreporting indicates that, for example, a 4-point difference in scores between 2 participants on the mental status test would be associated with the higher scoring participant having 15.8% fewer underreports than the lower scoring participant ($1-[\exp\beta]^4 = 10.000 - 0.958^4$). The first columns of the tables represent the unadjusted associations of the cognitive tests, whereas the second through fifth columns serially adjust the IRRs of the cognitive tests for the demographic, socioeconomic, health status, and interview period effects, respectively. The IRRs for the covariates are available online as supplementary tables ST3–ST8.(Supplemental Digital Content 3, http://links.lww.com/MLR/A886, Supplemental Digital Content 4, http://links.lww.com/MLR/A887, Supplemental Digital Content 5, http://links.lww.com/MLR/A888, Supplemental Digital Content 6, http://links.lww.com/MLR/A889, Supplemental Digital Content 7, http://links.lww.com/MLR/A890, Supplemental Digital Content 8, http://links.lww.com/MLR/A891).

TABLE 1. Descriptive Statistics for Concordance Measures, the 3 Cognitive Tests, and All of the Covariates

Variables	Mean or %	SD
Summary counts for the 12-item disease history measures		
Underreporting	1.2	1.3
Concordant reporting	10.2	1.4
Overreporting	0.6	0.9
Summary counts for the 4-item health services measures		
Underreporting	0.3	0.5
Concordant reporting	3.5	0.7
Overreporting	0.2	0.5
No. hospital episodes measures (%)		
Underreporting	12.5	
Concordant reporting	75.8	
Overreporting	11.7	
No. physician visits measures (%)		
Underreporting	68.1	
Concordant reporting	19.6	
Overreporting	12.4	
Cognitive tests		
Mental status (TICS-6)	9.1	1.4
Immediate word recall	4.6	1.8
Delayed word recall	3.3	2.1
Covariates		
Age	81.5	5.5
Sex (%)		
Men	35.1	
Women	64.9	
Race (%)		
Non-Hispanic whites	90.3	
African Americans	7.2	
Hispanic Americans	2.6	
Married (%)	38.8	
Education (%)		
Less than high school graduate or GED	16.0	
High school graduate or GED	50.4	
≥ 1 y of college	33.6	
Wealth (%)		
Lowest quintile	13.9	
Middle quintiles	61.6	
Highest quintile	24.5	
MediGap insurance coverage	71.8	
No. depressive symptoms	1.6	1.9
Self-rated health (1 = poor, …, 5 = excellent)	3.0	1.1
Self-rated memory (1 = poor, …, 5 = excellent)	3.0	1.0
Interview period (%)		
1995	29.7	
1998	20.7	
2000	16.1	
2002	12.3	
2004	9.5	
2006	6.4	
2008	3.8	
2010	1.5	

As shown in Table 2, only the associations with the mental status test were significant after adjustment for the demographic characteristics, and further adjustment for the socioeconomic status, health, and period covariates did not appreciably alter these associations. Higher scores on

TABLE 2. GEE Negative Binomial Models of Concordant Reporting, Underreporting, and Overreporting on the 12-Item Disease History Measures Among Self-Respondents in the AHEAD

Cognitive Measures	Model 1	Model 2	Model 3	Model 4	Model 5
Incident Rate Ratios for the Count of Concordant Reporting on the Disease History Measure					
Mental status (TICS-6)	1.006***	1.004***	1.004***	1.003***	1.003***
Delayed word recall	1.002***	1.001	1.000	1.000	1.000
Immediate Word-Recall	1.003***	1.001	1.001	1.001	1.001
Incident Rate Ratios for the Count of Underreporting on the Disease History Measure					
Mental status (TICS-6)	0.958***	0.979***	0.978***	0.979***	0.978***
Delayed word recall	0.987**	0.997	0.997	0.998	0.997
Immediate word recall	0.978***	0.992	0.992	0.992	0.994
Incident Rate Ratios for the Count of Overreporting on the Disease History Measure					
Mental status (TICS-6)	0.997	0.992	0.996	1.002	1.003
Delayed word recall	1.001	0.999	1.000	1.003	1.002
Immediate word recall	0.999	0.998	1.001	1.002	1.001

Model 1 contains the unadjusted associations for the 3 cognitive tests, and models 2–5 serially adjust for the demographic factors, socioeconomic status, health status, and the reinterview periods (y).

**$P < 0.01$.

***$P < 0.001$.

the mental status test were associated with greater counts of concordant reports on the disease history measure (IRR = 1.003, $P < 0.001$), and lesser counts of underreports (IRR = 0.978, $P < 0.001$), although the magnitude of these associations were small. None of the cognitive tests were significantly associated with overreporting on the disease history measure in any model.

Table 3 reveals that the mental status test also had significant associations with the counts of both concordant reporting and underreporting on the health services measure, and that these were not appreciably altered by adjustment for any covariates. Higher scores on the mental status test were associated with greater counts of concordant reports on the health services measure (IRR = 1.008, $P < 0.001$), and lesser counts of underreports (IRR = 0.927, $P < 0.001$). While these associations were relatively small, their magnitude was larger than that shown in Table 2 for the disease history measure. The delayed word-recall test also had significant associations with the concordant reporting and underreporting on the health services measure, although after adjustment for the covariates, the significance of the former was marginal. The associations of the delayed word-recall test with underreporting on the health services measure were about the same magnitude as those for the mental status test, with higher scores associated with smaller counts of underreports (IRR = 0.944, $P < 0.001$). None of the cognitive tests

TABLE 3. GEE Negative Binomial Models of Concordant Reporting, Underreporting, and Overreporting on the 4-Item Health Services Measures Among Self-Respondents in the AHEAD

Cognitive Measures	Model 1	Model 2	Model 3	Model 4	Model 5
Incident Rate Ratios for the Count of Concordant Reporting on the 4-Item Health Services Measure					
Mental status (TICS-6)	1.009***	1.009***	1.009***	1.008***	1.008***
Delayed word recall	1.005***	1.004**	1.004**	1.003*	1.003*
Immediate word recall	0.999	0.997	0.997	0.996*	0.996*
Incident Rate Ratios for the Count of Underreporting on the 4-Item Health Services Measure					
Mental status (TICS-6)	0.920***	0.918***	0.924***	0.926***	0.927***
Delayed word recall	0.937***	0.942***	0.945***	0.945***	0.944***
Immediate word recall	0.999	1.015	1.020	1.020	1.021
Incident Rate Ratios for the Count of Overreporting on the 4-Item Health Services Measure					
Mental status (TICS-6)	0.998	0.988	0.991	0.999	0.996
Delayed word recall	1.006	1.011	1.012	1.016	1.016
Immediate word recall	1.009	1.025	1.027	1.033	1.033

Model 1 contains the unadjusted associations for the 3 cognitive tests, and models 2–5 serially adjust for the demographic factors, socioeconomic status, health status, and the reinterview periods (y).

*$P < 0.05$.

**$P < 0.01$.

***$P < 0.001$.

were significantly associated with overreporting on the health services measure in any models.

GEE Logistic Regression Models of the Number of Hospital Episodes

Table 4 contains the results [adjusted odds ratios (AORs)] of the GEE binary logistic regression models of underreporting versus concordant reporting, and overreporting versus concordant reporting of the number of hospital episodes (the AORs for the covariates are available online as ST9 (Supplemental Digital Content 9, http://links.lww.com/MLR/A892) and ST10 (Supplemental Digital Content 10, http://links.lww.com/MLR/A893). Only the mental status test was significantly associated with underreporting and overreporting the number of hospital episodes. The AORs indicate that for each additional point on the mental status examination, the odds of underreporting were reduced by 12.4% when all covariates were in the model, and the odds of overreporting were reduced by 8.3%. The AORs for the covariates are available online as ST9 (Supplemental Digital Content 9, http://links.lww.com/MLR/A892) and ST10 (Supplemental Digital Content 10, http://links.lww.com/MLR/A893).

GEE Logistic Regression Models of the Number of Physician Visits

Table 5 contains the results of the GEE binary logistic regression models of underreporting versus concordant reporting (± 2), and overreporting versus concordant reporting of the number of physician visits (the AORs for the covariates are available online as ST11, Supplemental Digital Content 11, http://links.lww.com/MLR/A894 and ST12, Supplemental Digital Content 12, http://links.lww.com/MLR/A895). Only the delayed word-recall test was significantly associated with underreporting of the number of physician visits. The AORs indicate that for each additional point on the delayed word-recall test, the odds of underreporting the number of physician visits were reduced by 4.1% when all covariates were in the model. None of the 3 cognitive tests had statistically significant effects on overreporting the number of physician visits. The AORs for the covariates are available online as ST11 (Supplemental Digital Content 11, http://links.lww.com/MLR/A894) and ST12 (Supplemental Digital Content 12, http://links.lww.com/MLR/A895).

GEE Negative Binomial Models of the Volume of Physician Underreporting and Overreporting

The GEE negative binomial models of the associations of the 3 cognitive tests with the magnitude of underreporting physician visits (conditional on underreporting), and with the magnitude of overreporting physician visits (conditional on overreporting), were not informative (data not shown). That is, none of the cognitive tests was significantly associated in any model with the magnitude of either underreporting or overreporting the number of physician visits.

Sensitivity Analyses

Because 52.5% of the AHEAD participants had “perfect” scores (10s) on the mental status measure, we conducted sensitivity analyses to evaluate whether effects were nonlinear, in which case the estimates could be especially influenced by the large proportion of perfect scores. Those analyses (data not shown) replaced the metric mental status score (0–10) with a set of dummy variables using perfect scorers (10s) as the reference, with indicators for scores of 0–6 (5.4%), 7 (5.2%), 8 (11.1%), and 9 (25.8%). For all outcomes, those results revealed a monotonic, stair-stepped pattern with comparably sized risers and no discernable threshold or cutoff point(s), suggesting that the effects are linear.

DISCUSSION

This is the first study of which we are aware to use standard, objective cognitive tests (mental status, immediate word recall, and delayed word recall)[31,32] to evaluate the associations between cognitive function (mental status and episodic memory) and the agreement between self-reports and Medicare claims. The results were generally consistent with our hypotheses. We found that only the mental status test was significantly associated with concordant reporting and underreporting on the 12-item disease history measure, with higher scores leading to more concordant reporting and less underreporting, although the magnitude of these associations was small. On the 4-item health services measure, we found that higher scores on both the mental status and the delayed word-recall tests were significantly associated with more concordant reporting and less underreporting, although these associations were also small. Mental status was the only cognitive test significantly associated with the underreporting and overreporting of the number of hospital episodes after adjustment, with higher scores associated with lower odds of both. Only the delayed word-recall test was significantly associated with underreporting on the number of physician visits, with higher scores associated with lower odds. All of these associations persisted after extensive covariate adjustment. None of the cognitive tests were associated with the volume of either underreporting or overreporting physician visits.

Our study, however, had 5 limitations. First, the analyses were restricted to self-respondents, because proxies were not asked to complete the cognitive tests for the target person. Second, the cognitive status of the AHEAD self-respondents declined over time, but this led to censoring when they could no longer complete the interviews on their own. Third, the mental status scores for AHEAD participants were high, and the sensitivity analyses notwithstanding, the results may not be applicable to those with lower scores. Fourth, although the diagnostic validity of the TICS is well established[31,33–35] and the factor structure was consistent with expectations,[36–39] α for the mental status measure was low. Fifth, the range of health outcomes for which concordance could be assessed was limited by the ability to crosswalk the AHEAD survey questions with the Medicare claims data.

Although our focus was on the independent associations between cognitive function and the concordance be-

TABLE 4. GEE Logistic Models of Underreporting Versus Concordant Reporting, and Overreporting Versus Concordant Reporting on the Number of Hospital Episodes Among Self-Respondents in the AHEAD

Cognitive Measures	Model 1	Model 2	Model 3	Model 4	Model 5
Adjusted Odds Ratios of Underreporting vs. Concordant Reporting on the Number of Hospital Episodes					
Mental status (TICS-6)	0.849***	0.867***	0.879***	0.879***	0.876***
Delayed word recall	0.946**	0.958*	0.962	0.968	0.966
Immediate word recall	0.938*	0.960	0.971	0.987	0.988
Adjusted Odds Ratios of Overreporting vs. Concordant Reporting on the Number of Hospital Episodes					
Mental status (TICS-6)	0.910***	0.910***	0.917***	0.921***	0.917***
Delayed word recall	0.953*	0.959*	0.960*	0.968	0.967
Immediate word recall	0.989	1.008	1.017	1.044	1.047

Model 1 contains the unadjusted associations for the 3 cognitive tests, and models 2–5 serially adjust for the demographic factors, socioeconomic status, health status, and the reinterview periods (y).
*$P<0.05$.
**$P<0.01$.
***$P<0.001$.

tween survey reports and Medicare claims, and only those results are shown in Tables 2–5, the full models with all the covariates and their associations are available online in supplemental tables ST3–ST12 (Supplemental Digital Content 3, http://links.lww.com/MLR/A886, Supplemental Digital Content 4, http://links.lww.com/MLR/A887, Supplemental Digital Content5, http://links.lww.com/MLR/A888, Supplemental Digital Content6, http://links.lww.com/MLR/A889, Supplemental Digital Content 7, http://links.lww.com/MLR/A890, Supplemental Digital Content 8, http://links.lww.com/MLR/A891, Supplemental Digital Content 9, http://links.lww.com/MLR/A892, Supplemental Digital Content 10, http://links.lww.com/MLR/A893, Supplemental Digital Content 11, http://links.lww.com/MLR/A894, Supplemental Digital Content 12, http://links.lww.com/MLR/A895). The word constraints of the journal, however, only permit us to briefly mention here those respondent characteristics with generally consistent associations. Older adults and men were less likely to be concordant, and were more likely to both underreport and overreport disease history counts and all of the health services use measures. African Americans were less likely than whites to underreport the health services counts and the number of hospital episodes, but were more likely than whites to overreport the number of physician visits. Hispanic Americans were more likely than whites to underreport disease history counts. And those with better self-rated health were more likely to be concordant reporters of disease history counts and all of the health services use measures.

Notwithstanding the study limitations, our results are informative for health services research, of which a nontrivial portion relies on self-reports of disease histories and health services use from older adults. Cognitive function was directly associated with greater accuracy and fewer errors in self-reports relative to Medicare claims, even after adjustment for demographic characteristics, socioeconomic status, self-rated health, memory, depressive symptoms, and secular trends. Most importantly, it was primarily mental status (the TICS-6) rather than short-term or long-term (episodic) memory (word recall) that had the largest and most consistent associations with improved accuracy and reduced error. Indeed, only mental status was significantly associated with accurate reporting on the disease history counts and the number of hospital episodes, both of which are extensively used as either focal and/or outcome measures in health services research.

Therefore, we recommend that future surveys of older adults include the TICS-6 or a similar mental status test (one

TABLE 5. GEE Logistic Models of Underreporting Versus Concordant Reporting, and Overreporting Versus Concordant Reporting on the Number of Physician Visits Among Self-Respondents in the AHEAD

Cognitive Measures	Model 1	Model 2	Model 3	Model 4	Model 5
Adjusted Odds Ratios of Underreporting vs. Concordant Reporting on the Number of Physician Visits					
Mental status (TICS-6)	0.962	0.970	0.955*	0.952*	0.958
Delayed word recall	0.946**	0.960*	0.956**	0.962*	0.959*
Immediate word recall	0.986	1.009	1.002	1.009	1.011
Adjusted Odds Ratios of Overreporting vs. Concordant Reporting on the Number of Physician Visits					
Mental status (TICS-6)	0.930*	0.948	1.054	0.954	0.946
Delayed word recall	1.009	0.930	0.971	1.041	1.043
Immediate word recall	0.955	1.141**	1.127	0.980	0.980

Model 1 contains the unadjusted associations for the 3 cognitive tests, and models 2–5 serially adjust for the demographic factors, socioeconomic status, health status, and the reinterview periods (y).
*$P<0.05$.
**$P<0.01$.

that generates more variation in scores than the TICS-6 would be particularly desirable), especially when survey reports cannot be confirmed by linkage to Medicare claims or chart review. We make this recommendation based on the consistency of the direct associations between the mental status test and the greater accuracy and lower underreporting in self-reports on disease history and the types and volumes of health services used, as well as the relative ease with which the mental status test can be used in either mail, telephone, or personal health surveys. Although the associations of the mental status test with the 12-item disease history and 4-item health services measures were small, its association with underreporting of the number of hospital episodes was notable, with each additional point associated with a 12% reduction in the odds of underreporting. We also encourage research using surveys of older adults other than the AHEAD that can be linked to Medicare claims and in which objective measures of cognitive function are available to replicate these findings.

REFERENCES

1. Wolinsky FD, Arnold CL. A different perspective on health and health services utilization. *Annu Rev Gerontol Geriatr*. 1988;8:71–101.
2. Ferraro KF, Farmer MM. Utility of health data from social surveys: is there a gold standard for measuring morbidity? *Am Sociol Rev*. 1999;64:303–315.
3. Ferraro KF. Health and aging. In: Binstock RH, George LK, eds. *Handbook of Aging and the Social Sciences*. 6th ed. New York, NY: Academic Press; 2006:238–256.
4. Feld K. Survey costs In: Lavrakas PJ, ed. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications; 2008: 862–866.
5. Shine D, Sundaram P, Torres DM, et al. Can computerized cost data substitute for chart review? *J Healthc Qual*. 2002;24:26–33.
6. Hogg W, Gyorfi-Dyke E, Dahrouge S, et al. Conducting chart audits in practice-based primary care research. *Can Fam Physician*. 2010;56: 495–496.
7. Cleary PD, Jette AM. The validity of self-reported physician utilization measures. *Med Care*. 1984;22:796–803.
8. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev*. 2006; 63:217–235.
9. Day HR, Parker JD. Self-report of diabetes and claims-based identification of diabetes among Medicare beneficiaries. *Natl Health Stat Rep*. 2013;69:1–15.
10. Fiscella K, Holt K, Meldrum S, et al. Disparities in preventive procedures: comparisons of self-report and Medicare claims data. *BMC Health Serv Res*. 2006;6:1–8.
11. Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. *J Ambul Care Manage*. 1998;21:24–34.
12. Miller DR, Rogers WH, Spiro A, et al. *Evaluation of Disease Status Based on Patient Self-Report in the Medicare Health Outcomes Survey*. Bedford, MS: Department of Veterans Affairs; 2003.
13. Raina P, Torrance-Rynard V, Wong M, et al. Agreement between self-reported and routinely collected health-care utilization data among seniors. *Health Serv Res*. 2002;37:751–774.
14. Ritter PL, Stewart AL, Kaymaz H, et al. Self-reports of health care utilization compared to provider records. *J Clin Epidemiol*. 2001; 54:136–141.
15. Roberts RO, Bergstrahl EJ, Schmidt L, et al. Comparison of self-reported and medical records health care utilization measures. *J Clin Epidemiol*. 1996;49:989–995.
16. Tisnado DM, Adams JL, Liu H, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care*. 2006;44:132–140.
17. Wallihan DB, Stump TE, Callahan CM. Accuracy of self-reported health services use and patterns of care among urban older adults. *Med Care*. 1999;37:662–670.
18. Wolinsky FD, Miller TR, An H, et al. Hospital episodes and physician visits: the concordance between self-reports and Medicare claims. *Med Care*. 2007;45:300–307.
19. Zuvekas SH, Olin GL. Validating household reports of health care use in the Medical Expenditure Panel Survey. *Health Serv Res*. 2009;44:1679–1700.
20. Zuvekas SH, Olin GL. Accuracy of Medicare expenditures in the Medical Expenditure Panel Survey. *Inquiry*. 2009;46:92–108.
21. Wolinsky FD, Jones MP, Ullrich F, et al. The concordance of survey reports and Medicare claims in a nationally representative longitudinal cohort of older adults. *Med Care*. 2014;52:462–468.
22. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press; 2000.
23. Podsakoff PM, MacKenzie SB, Lee J-Y, et al. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol*. 2003;88:879–903.
24. Brener ND, Billy JOG, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *J Adolesc Health*. 2003;33:436–457.
25. Andersen R, Kasper J, Frankel MR. *Total Survey Error*. San Francisco, CA: Jossey-Bass; 1979.
26. Andrews FM, Herzog AR. The quality of survey data as related to age of respondent. *J Am Stat Assoc*. 1986;81:403–410.
27. Kempen GIJM, Steverink N, Ormel J, et al. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. *J Gerontol B Psychol Sci Soc Sci*. 1996;51B:P254–P260.
28. Herr KA, Garand L. Assessment of pain in older adults. *Clin Geriatr Med*. 2001;17:457–465.
29. Myers GC, Juster FT, Suzman RM. Assets and Health Dynamics among the Oldest Old (AHEAD): initial results from the longitudinal study. *J Gerontol B Psychol Sci Soc Sci*. 1997;52B(SI):v–viii.
30. Juster FT, Suzman RM. An overview of the Health and Retirement Study. *J Hum Resour*. 1995;30:S7–S56.
31. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatr Neuropsychol Behav Neurol*. 1988;1:111–117.
32. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive sate of patients for the clinician. *J Psychiatr Res*. 1975;12:189–198.
33. Manly JJ, Schupf N, Stern Y, et al. Telephone-based identification of mild cognitive impairment and dementia in a multicultural cohort. *Arch Neurol*. 2011;68:607–614.
34. Plassman BL, Newman TN, Welsh KA, et al. Properties of the Telephone Interview for Cognitive Status: application in epidemiological and longitudinal studies. *Neuropsychiatr Neuropsychol Behav Neurol*. 1994;7:235–241.
35. Ferruci L, Del Lungo I, Guralnik JM. Is the Telephone Interview for Cognitive Status a valid alternative in persons who cannot be evaluated by the Mini-Mental State Examination? *Aging*. 1998;10:332–338.
36. Herzog AR, Wallace RB. Measures of cognitive functioning in the AHEAD study. *J Gerontol B Psychol Sci Soc Sci*. 1997;52B(SI):37–48.
37. Wolinsky FD, Bentler SE, Hockenberry J, et al. A prospective cohort study of long-term cognitive changes in older Medicare beneficiaries. *BMC Public Health*. 2011;73:1–11.
38. McArdle JJ, Fisher GG, Kadlec M. Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992-2004. *Psychol Aging*. 2007;22:525–545.
39. Blankson AN, McArdle JJ. A brief report on the factor structure of the cognitive measures in the HRS/AHEAD studies. *J Aging Res*. 2014;5:1–5.
40. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13–22.
41. Diggle PJ, Heagerty P, Liang KY, et al. *Analysis of Longitudinal Data*. New York, NY: Oxford University Press; 2002.