Published in final edited form as:

Med Care. 2007 April ; 45(4): 300–307.

Hospital Episodes and Physician Visits: *The Concordance Between Self-Reports and Medicare Claims*

Fredric D. Wolinsky, PhD^{*,†}, Thomas R. Miller, MBA[†], Hyonggin An, PhD[†], John F. Geweke, PhD[†], Robert B. Wallace, MDMS[†], Kara B. Wright, MS[†], Elizabeth A. Chrischilles, PhD[†], Li Liu, MS[†], Claire B. Pavlik, PhD[†], Elizabeth A. Cook, MA[†], Robert L. Ohsfeldt, PhD[‡], Kelly K. Richardson, PhD^{*,†}, and Gary E. Rosenthal, MD^{*,†}

* Iowa City VA Health Care System, Iowa City, Iowa

† University of Iowa, Iowa City, Iowa

‡ Texas A&M University Health Science Center, College Station, Texas.

Abstract

Background—Health services use typically is examined using either self-reports or administrative data, but the concordance between the 2 is not well established.

Objective—We evaluated the concordance of hospital and physician utilization data from self-reports and claims data, and identified factors associated with disagreement.

Methods—We performed a secondary analysis on linked observational and administrative data. A national sample of 4310 respondents who were 70 years old or older at their baseline interviews was used. Self-reported and Medicare claims-based hospital episodes and physician visits for 12 months before baseline were examined. Kappa statistics were used to evaluate concordance, and multivariable multinomial logistic regression was used to identify factors associated with overreporting (self-reports > claims), underreporting (self-reports < claims), and concordant-reporting (self-reports ~ claims).

Results—The concordance of hospital episodes was high ($\kappa = 0.767$ for the 2×2 comparison of none vs. some and $\kappa = 0.671$ for the 6×6 comparison of none, 1, ..., 4, or 5 or more), but concordance for physician visits was low ($\kappa = 0.255$ for the 2×2 comparison of none versus some and $\kappa = 0.351$ for the 14×14 comparison of none, 1, ..., 12, and 13 or more). Multivariable multinomial logistic regression indicated that over-, under-, and concordant-reporting of hospital episodes was significantly associated with gender, alcohol consumption, arthritis, cancer, heart disease, psychologic problems, lower body functional limitations, self-rated health, and depressive symptoms. Over-, under-, and concordant-reporting of physician visits were significantly associated with age, gender, race, living alone, veteran status, private health insurance, arthritis, cancer, diabetes, hypertension, heart disease, lower body functional limitations, and poor memory.

Conclusions—Concordance between self-reported and claims-based hospital episodes was high, but concordance for physician visits was low. Factors significantly associated with bidirectional

Reprints: Fredric D. Wolinsky, the John W. Colloton Chair in Health Management and Policy, College of Public Health, the University of Iowa, 200 Hawkins Drive, E-205 General Hospital, Iowa City, Iowa 52242. E-mail: fredric-wolinsky@uiowa.edu.

Supported by NIH grants R01 AG-022913 and R03 AG027741 to Dr. Wolinsky. Dr. Wolinsky is the Associate Director of the Center for Research in the Implementation of Innovative Strategies in Practice (CRIISP) at the Iowa City VA Medical Center, Dr. Rosenthal is the Director of CRIISP, and Dr. Richardson is a CRIISP Statistician. CRIISP is funded through the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service (HFP 04-149).

The opinions expressed here are those of the authors and do not necessarily reflect those of any of the funding, academic or governmental institutions involved.

(over- *and* underreporting) and unidirectional (over- *or* underreporting) error patterns were detected. Therefore, caution is advised when drawing conclusions based on just one physician visit data source.

Health care costs have increased annually at or near the double-digit level for 3 decades.¹ By 2008, health care costs will be \$2.5T, or one-sixth of the GDP.² Health care costs for older adults are 3 times larger than those for younger adults, and most of these costs accrue from hospital episodes and physician visits paid for by public funds.³ Indeed, 40% of Medicare claims dollars are for hospital inpatient expenses, and the next largest outlay (18%) is for managed care, of which a major proportion is also for inpatient expenses.⁴ Furthermore, substantial social and cultural inequalities exist in the use of health services among older adults, as well as in the quality of the health services they receive.⁵ The elimination of these inequalities is one of the main goals in *Crossing the Quality Chasm*.

If health care costs are to be constrained, and if social and cultural inequalities in service consumption are to be eliminated, further research on health services use among older adults is needed. In general, studies of health services use rely either on self-reports *or* administrative data. The difference between these data sources has been considered for decades. In the 1960s and 1970s, the concern was whether sufficiently accurate information could be obtained directly from respondents, because administrative records were not readily accessible. Interest in the 1980s shifted to the abilities of administrative records from a given care source to capture out-of-plan use, especially in health maintenance organizations and other managed care plans. By the 1990s, the focus had shifted to the ability to rely solely on claims data for modeling purposes. The concordance between self-reports and administrative data, however, is not well established, especially among older adults.^{6–11}

It has been assumed and demonstrated that (1) the more salient the health event is to the individual, the more accurate the match between their self-reports and administrative claims, and (2) the longer the recall period, the less accurate the match.^{9,11–18} Because health events requiring hospitalization are generally regarded as the most salient to individuals, and because recall accuracy is known to decay with volume, the least accurate self-reported recall should involve the number of physician visits during the last year, whereas the most accurate should exist for whether any hospital episodes occurred.^{11,17}

In this article, we use data from a large, nationally representative sample of older adults to achieve 2 goals. First, we evaluate the concordance of hospital and physician utilization data obtained from self-reports and Medicare claims data. Second, we use multivariable multinomial logistic regression to examine the factors associated with overreporting (self-reports > claims), underreporting (self-reports < claims), and concordant-reporting (self-reports ~ claims) between these 2 informational sources.

METHODS

Sample

Data were taken from the Survey on Assets and Health Dynamics among the Oldest Old (AHEAD).¹⁹ Respondents were identified either from household screening conducted during the 1992 multistage cluster sampling process for the companion Health and Retirement Study²⁰ of preretirement-aged adults, or a supplemental sample of persons 80 years or older identified from the CMS Medicare Master Enrollment File. Oversampling increased the number of black, Hispanic, and Floridian subjects. Thus, all analyses presented here are weighted to adjust for the unequal probabilities of selection due to the multistage cluster and oversampling designs.

Baseline AHEAD in-home interviews were conducted in 1993/1994 with 7447 respondents who were 70 years old or older. The response rate was 80.4%. Complete linkage to Medicare Part A and B claims was accomplished for 4,697 individuals (63%). Of these, we excluded 101 respondents who had any evidence of being in Medicare managed care during the 2-year prebaseline period because managed care plans are not required to report complete data. We also excluded 286 individuals for whom baseline data was provided by a proxy, because our purpose was to compare *self*-reports with claims data. Thus, our analytic sample involved 4310 men and women (58% of the original AHEAD cohort).

Self-Reported Hospital Episodes and Physician Visits

At baseline, AHEAD respondents were asked 2 questions about their hospital utilization. The first was: “During the last 12 months, since (month) of (1992/1993), have you been a patient in a hospital overnight?” The response options were “yes” or “no.” Respondents who said “yes” were then asked: “How many different times were you a patient in a hospital overnight in the last 12 months?” The response options were 1 through the highest integer reported, which was 20. Two similar questions were asked about physician visits. The first was: “(Aside from any hospital or nursing home stays), during the last 12 months, since (month) of (1992/1993), have you seen a medical doctor about your health?” Again, the response options were “yes” or “no.” Respondents who said “yes” were then asked: “How many times have you talked to a medical doctor (about your own health) in the last 12 months?” Again, the response options were 1 through the highest integer reported, which was 50.

Claims-Based Hospital Episodes and Physician Visits

Claims-based hospital episodes and physician visits were obtained as follows. First, the exact date of each AHEAD respondent’s baseline interview was determined. Then, all hospital episodes and physician visits in the Medicare Part A and B claims files for the 12 months prior to each respondent’s interview date were identified. Determining the number of hospital episodes was straightforward, and simply involved retaining all Part A claims episodes that lasted for at least one night, to be comparable to the self-report questions posed to respondents. We note that as an added safeguard for respondent anonymity, CMS selected a random integer from -14 to $+14$ for each AHEAD respondent, and added that random integer consistently to the Julian dates for all of the Part A and Part B claims for that subject. For example, for Mary Jones the random integer of -6 was selected. As a result, -6 was added (ie, subtracted) from the Julian dates for all of Mary Jones’ Part A and Part B Medicare claims. Although this creates the potential for discrepancies between the self-reported and claims-based utilization totals, that potential is marginal and random, and thus completely ignorable.²¹

Determining the number of physician visits was not as straightforward, and involved 2 phases. First, it required defining a “visit.” Simply put, Part B data are structured as “lines” (billable goods, services, or procedures) performed under (within) a specific “claim.” To restrict our measure to the outpatient setting (and thus achieve comparability with the self-reported data), we first deleted all inpatient-related line items and claims (ie, hospital, hospital discharge, hospital consultation, nursing facility, and care plan oversight services). We then deleted all Part B claims for which the “from and through” (service start and stop) dates completely overlapped with Part A hospital stays (admission to discharge dates), with one exception: physician claims that occurred on the day of admission were included, because these most likely reflect outpatient or emergency department encounters during which the decision to hospitalize was made. To restrict the measure to physician services provided directly to patients (analogous to what AHEAD respondents would report as a physician visit), we deleted line items or claims that were not primary or specialty care (ie, anesthesiology, drugs, supplies, radiology, labs, pathology). To further ensure that the patients were “seen” (ie, that “visits” occurred), we then deleted line items and claims without evaluation and management (E&M)

codes. After deleting duplicates (same day, same provider), all remaining line items with E&M codes qualified as “visits.”

Covariates

Consistent with *Crossing the Quality Chasm*,⁵ our approach focuses on a comprehensive set of covariates known to be associated with health services use among older adults. These covariates may be classified into 5 categories—socio-demographic characteristics, socioeconomic factors, lifestyle, disease history, and functional limitations.^{22–26} Sociodemographic characteristics included age (measured in years), sex (men vs. women), race (2 dummy variables contrasting black and Hispanic subjects with non-Hispanic Whites), and whether the respondent lived alone (yes vs. no). Socioeconomic factors included education (2 dummy variables contrasting only grade school or at least some college with high school), income (2 dummy variables contrasting less than \$7K or more than \$50K with incomes in-between), veteran status (yes vs. no, reflecting potential access to Veterans Health Administration services as a source of any observed discordance between self-reports and Medicare claims), and private health insurance (yes vs. no; recall that all respondents had Medicare).

Lifestyle factors included smoking (ever having smoked cigarettes vs. never), weight (2 dummy variables contrasting overweight or obese respondents based on the National Institutes of Health body mass index thresholds with normal and underweight), alcohol consumption (3 dummy variables contrasting <1 drink daily, 1–2 drinks daily, and 3 or more drinks daily with no alcohol consumption), and whether the respondent never drove a motor vehicle (2 dummy variables contrasting those who never drove and those who currently drive with those who have had to give up driving, which we consider a functional limitation). Disease history included 8 indicator variables for whether the respondent reported having (yes versus no), arthritis, cancer, diabetes, hypertension, lung disease, a heart condition, hip fracture, or psychologic problem at baseline. Functional limitations included the number (0–5) of activities of daily living (ADLs) with difficulty, the number (0–5) of instrumental ADLs (IADLs) with difficulty, and the number (0–5) of lower body functional limitations, as well as 4 indicator variables for self-reports of fair or poor (versus excellent, very good, or good) hearing, vision, memory, or overall health. Also included were current ability to drive a motor vehicle, depressive symptoms (2 dummy variables contrasting none, or 3–8 symptoms with 1–2 symptoms),²⁷ and cognitive status (2 dummy variables contrasting 0–10 [low] and 14–15 [high] with in-between scores).²⁸

Analytic Methods

Concordance on hospital episodes and physician visits was evaluated using simple and weighted kappa (κ) statistics, as appropriate for 2×2 and larger (ie, $N \times N$) tables, respectively.²⁹ Multivariable multinomial logistic regression was used to identify covariates associated with overreporting (self-reports > claims), underreporting (self-reports < claims), and concordant-reporting (self-reports ~ claims) of their number of hospital episodes.^{30–34} Multivariable multinomial logistic regression was also used to identify covariates associated with respondents overreporting, underreporting, or concordant-reporting of their number of physician visits. Because the correspondence between self-report and claims-based totals of physician visits was expected to be less robust, we performed sensitivity analyses in which we estimated these multivariable multinomial logistic regressions using 3 bandwidth criteria for determining discordant-reporting: ± 1 or more visits, ± 2 or more visits, and ± 3 or more visits. Although all multivariable models entered the covariates serially, starting with the most distal to the most proximal in time sequence to trace decomposition effects, only the final models are shown here to enhance clarity and due to space constraints.

RESULTS

Descriptive Data

Among the 4310 AHEAD subjects in the analytic sample, the mean age was 77.3 years, 35.1% were men, 8.7% were black, 4.0% were Hispanic, and 39.1% lived alone. One-fourth had only been to grade school, and 27.6% had been to college. Twenty-three percent reported incomes less than \$7K, and 14.9% reported incomes of \$50K or more. There were 21.4% veterans, and 78.6% had private health insurance. Half were or had been smokers, 35.7% were overweight, 14.1% were obese, 42.9% did not drink alcohol, 2.0% averaged ≥ 3 alcohol drinks per day, and 11.2% had never driven a motor vehicle. One-fourth reported arthritis, 13.0% reported cancer, 11.5% reported diabetes, 46.4% reported hypertension, 8.8% reported lung disease, 28.1% reported a heart condition, 4.2% reported a fractured hip, and 7.3% reported psychologic problems. The mean number of ADLs was 0.29, the mean number of IADLs was 0.38, and the mean number of lower-body functional limitations was 1.33. Fair/poor hearing, vision, memory, and health were reported by 24.7%, 25%, 25.2%, and 33%, respectively, and 70.4% were able to drive. No depressive symptoms were reported by 37.7%, and 25.7% had 3 or more. Twenty-seven percent had low scores on cognitive status and 39.9% had high scores.

The Prevalence of Hospital Episodes and Concordance

Table 1 contains the cross-classification of the numbers of self-reported versus claims-based hospital episodes in the year prior to baseline for 4229 AHEAD respondents (81 subjects did not provide self-reports). As shown, one or more hospital episodes were identified from the Part A claims for 16.4% of respondents, and 21.1% of the respondents reported having been hospitalized. The mean number of hospital episodes based on claims was 0.21, and the mean number based on self-reports was 0.31. Concordance between these data sources was high, with simple $\kappa = 0.767$ for the 2×2 comparison of none versus ≥ 1 , and weighted $\kappa = 0.671$ for the 6×6 comparison of none, 1, ..., 4, or ≥ 5 . In sensitivity analyses to explore the possible effect of telescoping (data not shown), we lengthened the claims-based look-back period by 3, 6, 9, and 12 months. Concordance, however, was not affected, with simple $\kappa = 0.765, 0.735, 0.697$, and 0.662 , respectively, for the 2×2 comparison of none versus >1 . Overall, there were only 475 divergent cases on the number of hospital episodes, most of which (79.4%) involved over-reporting by respondents. Two-thirds (66.6%) of all of the overreports occurred among respondents with no claims-based evidence of any hospital episodes.

Factors Associated with Over- or Underreporting Hospital Episodes

Table 2 contains the adjusted odds ratios (AORs) from the weighted multivariable multinomial logistic regression of the 4201 AHEAD respondents for whom complete data on all of the covariates were available. These models predict whether the respondent was among the 398 individuals who overreported, or among the 101 individuals who underreported their number of hospital episodes, compared with the 3702 individuals who concordantly-reported. Note that if a covariate has AORs of the same sign (ie, >1 or <1) and comparable magnitude, that covariate identifies respondents prone to reporting bidirectional errors or general errors in reporting, rather than respondents prone to unidirectional errors (ie, specific errors involving either over- or underreporting, but not both).

Three bidirectional errors were identified. Men were more likely to over- and underreport their number of hospital episodes, although the latter was marginally insignificant given the smaller number of underreporters. Respondents with heart disease also were more likely to over- and under-report but were noticeably more likely to underreport. Those with lower-body functional limitations were also more likely to over- and underreport, although the latter was again marginally insignificant given the smaller number of underreporters.

Six unidirectional errors were identified. Respondents who reported having cancer, psychologic problems, or poor self-rated health were more likely to overreport their number of hospital episodes, whereas respondents who had modest alcohol consumption habits were less likely to overreport. Underreporting the number of hospital episodes was more likely to occur among respondents having arthritis, and less likely to occur among respondents without depressive symptoms.

The Prevalence of Physician Visits and Concordance

Table 3 contains the cross-classification of the numbers of self-reported versus claims-based physician visits in the year prior to baseline for 4182 AHEAD respondents (128 subjects did not provide self-reports). The mean number of physician visits based on claims was 5.8, and the mean number based on self-reports was 4.8. Although these means differ by just one visit, they do not indicate comparability. For example, no physician visits were reported by only 10.8% of the respondents, but no physician visits were found in the claims for 22.2%. Moreover, simple $\kappa = 0.255$ for the 2×2 comparison of none versus ≥ 1 , and weighted $\kappa = 0.351$ for the 14×14 comparison of none, 1, ..., 12, or ≥ 13 . Given the markedly lower concordance between the self-reported and claims-based physician visit totals (compared with the concordance for hospital episodes), sensitivity analyses using 3 bandwidth criteria for discordant-reporting were conducted. These included ± 1 or more visits, ± 2 or more visits, and ± 3 or more visits. Table 4 shows the distributions of over-, concordant-, and underreporting of physician visits using these 3 bandwidth criteria. Even under the most relaxed bandwidth criterion of ± 3 or more visits, only about half of the AHEAD respondents were classified as concordant-reporters. Further relaxation of the bandwidth criterion is inappropriate in light of the mean number of visits.

Factors Associated with Over- or Underreporting Physician Visits

Table 5 contains the AORs from the weighted multivariable multinomial logistic regression among the 4154 AHEAD respondents for whom complete data on all of the covariates were available. These models identified covariates associated with respondents who were over-concordant-, or underreporters of their total number of physician visits using the 3 bandwidth criteria described above. As with Table 3, note that if a covariate has AORs of the same sign (ie, >1 or <1) and comparable magnitude, that covariate identifies respondents prone to reporting bidirectional errors or general errors in reporting, rather than respondents prone to unidirectional errors (ie, over- or underreporting, but not both).

As shown in Table 5, the pattern of covariates associated with the over- and underreporting of physician visits is remarkably similar across the 3 bandwidth criteria. Given the robustness of these results, we focus on the broadest (ie, most relaxed) bandwidth criterion, which is shown in the first 2-column panel of the table. In this analysis, 10 bidirectional errors were identified. Hispanics, those who live with others, have private health insurance, and report having arthritis, cancer, diabetes, hypertension, heart disease, lower-body functional limitations, or poor memory were significantly more likely to both over- and underreport their number of physician visits. Four unidirectional effects also were identified. Older adults were more likely to underreport the number of their physician visits, men were less likely to overreport, and black respondents and veterans were more likely to overreport their number of physician visits.

DISCUSSION

There are 3 important aspects of these findings that warrant further discussion. The first involves hospital episodes. Our results demonstrated that (1) the congruence of self-reports versus claims-based data for hospital episodes was high, (2) errors between these 2 data sources generally involved overreporting by the AHEAD respondents, (3) 3 covariates were associated

with bidirectional or general error patterns (ie, men, those with heart disease, or those with lower body functional limitations were more likely to both over- and underreport), and (4) 6 covariates were associated with unidirectional error patterns (those with cancer, psychologic problems, or poor self-rated health were more likely to overreport; modest drinkers of alcohol were less likely to overreport; and those with arthritis or without depressive symptoms were less likely to underreport).

For the most part, these findings are consistent with previous reports^{6–15} and are intuitively plausible. On the basis of these findings, we conclude that self-reports and claims-based data for hospital episodes for a 12-month recall period are readily substitutable. Indeed, although not shown here, multivariable logistic regression and multivariable multinomial logistic regression separately using either self-reports or claims-based data yielded equivalent predictive models of the demand for and volume of hospital utilization. This is not surprising, given the high kappa statistics between the self-reported and claims-based measures.

The second important aspect of our findings involves physician visits. In stark contrast to the situation with hospital episodes, (1) the congruence between self-reports and claims-based data on physician visits was low, (2) the covariates associated with most of the errors between these 2 data sources were bidirectional (except for older adults being more likely to underreport, and women, blacks, and veterans being more likely to overreport), (3) the bidirectional errors appeared rationally based (eg, the likelihood of over- and underreporting was greatest for those with diseases, lower-body functional limitations, and poor memory abilities), and (4) most of the unidirectional errors appeared rationally-based (older adults are known to underreport, and veterans are more likely to overreport given their access to the Veterans Health Administration, for which visits would not show up in Medicare claims). On the basis of these findings, we conclude that self-reports and claims-based data for physician visits for a 12-month recall period are *not* readily substitutable, and that caution must be exercised when relying on just one of these data sources.

The third important aspect of these findings that warrants further discussion involves the self-reported poor memory marker. When the congruence between self-reports and claims-based data is low, as is the case for physician visits over the past 12 months, poor self-reported memory plays an important role in predicting bidirectional error. As shown in Table 5, the increased odds of over- and underreporting physician visits were 34.1% and 34.5%, respectively, among those with poor self-reported memory. This result is entirely consistent with a growing body of work that underscores the importance of self-reported memory as an efficient marker for current clinical memory deficits, and as an effective predictor of subsequent declines in memory performance.^{35–37} On the basis of these findings, we recommend that future studies using self-reported data on physician visits include the self-reported memory question whenever possible for adjustment purposes.

Despite the important contributions that this article makes to the literature, this study is not without its own limitations. Three of these warrant special mention here. First, no medical charts were available for use in reconciling discrepancies between the self-reported and claims-based numbers of hospital and physician visits. As a result, we know more about the epidemiology of discordance than its etiology, and this is especially the case with regard to physician visits, where the concordance between self-reports and claims data is so much lower. Second, neither self-reports, nor administrative claims, nor the medical record are gold standards. Indeed, they are just different measures of the same latent constructs, and thus attributing one or the other as the source of the error that accounts for the discordance between them is arbitrary and inappropriate. Third, self-reported survey data and Medicare claims could only be linked for 58% of the 7447 AHEAD respondents, creating the potential for selection bias. Previous analyses of these data, however, have failed to identify any meaningful evidence

of such selection bias.^{38,39} Therefore, the main contributions of this study represent significant contributions to the literature on the concordance of self-reported and claims-based health services utilization data using methods consistent with the current state of the art.

References

1. Levit K, Smith C, Cowan C, et al. Trends in US health care spending, 2010. *Health Affairs* 2003;22:154–164. [PubMed: 12528847]
2. Heffler S, Smith S, Keehan S, et al. Health spending projections for 2002–2012. *Health Affairs* 2003;W3:54–65.
3. National Center for Health Statistics. DHHS Pub. No. 1232. Hyattsville, MD: US GPO; 2002. Health, United States, 2002, with Chartbook on Trends in the Health of Americans.
4. Centers for Medicare and Medicaid Services. 2002 Data Compendium. June 2002. Baltimore, MD: Centers for Medicare and Medicaid Services; 2002.
5. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
6. 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 Mgmt* 1998;21:24–34.
7. Raina P, Torrance-Rynard V, Wong M, et al. Agreement between self-reported and routinely collected health-care utilization data among seniors. *Health Services Res* 2002;37:751–774.
8. 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. [PubMed: 11166528]
9. 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. [PubMed: 8780606]
10. 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. [PubMed: 16434912]
11. 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. [PubMed: 10424637]
12. Andersen, R.; Kasper, J.; Frankel, MR. Total Survey Error. San Francisco, CA: Jossey-Bass; 1979.
13. Cleary PD, Jette AM. The validity of self-reported physician utilization measures. *Med Care* 1984;22:796–803. [PubMed: 6492908]
14. Coleman EA, Wagner EH, Grothaus LC, et al. Predicting hospitalization and functional decline in older health plan enrollees: Are administrative data as accurate as self-report? *JAGS* 1998;46:419–425.
15. Glandon GL, Counte MA, Tancredi D. An analysis of physician utilization by elderly persons: systematic differences between self-report and archival information. *J Gerontol* 1992;47:S245–S252. [PubMed: 1512446]
16. Jobe JB, Mingay DJ. Cognitive research improves questionnaires. *AJPH* 1989;79:1053–1055.
17. Jobe JB, Tourangeau R, Smith AF. Contributions of survey-research to the understanding of memory. *Applied Cog Psych* 1993;7:567–584.
18. Jobe JB, White AA, Kelley CL, et al. Recall strategies and memory for health-care visits. *Milbank Q* 1990;68:171–189.
19. Myers GC, Juster FT, Suzman RM. Asset and Health Dynamics Among the Oldest Old (AHEAD): initial results from the longitudinal study. *J Gerontol: Psychol Sci Soc Sci* 1997;52B(Special Issue):v–viii.
20. Juster FT, Suzman RM. An overview of the health and retirement study. *J Human Resources* 1995;30:S7–S56.
21. Allison, PD. Missing Data. Thousand Oaks, CA: Sage; 2002.
22. Andersen, RM. A Behavioral Model of Families' use of Health Services. Chicago, IL: Center for Health Administration Studies; 1968.
23. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav* 1995;36:1–10. [PubMed: 7738325]

24. Miller JE, Russell LB, Davis DM, et al. Biomedical risk factors for hospital admission in older adults. *Med Care* 1998;36:411–421. [PubMed: 9520964]
25. Weissman JS, Stern R, Fielding SL, et al. Delayed access to health care: risk factors, reasons, and consequences. *Annals Intern Med* 1991;114:325–331.
26. Wolinsky FD. Health services utilization among older adults: conceptual, measurement, and modeling issues in secondary analysis. *Gerontologist* 1994;34:470–475. [PubMed: 7959103]
27. Kohout FJ, Berkman LF, Evans DA. Two shorter forms of the CES-D depression symptoms index. *J Aging Health* 1993;5:179–193. [PubMed: 10125443]
28. Herzog AR, Wallace RB. Measures of cognitive functioning in the AHEAD study. *J Gerontol: Psychol Sci Soc Sci* 1997;52B(Special Issue):37–48.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174. [PubMed: 843571]
30. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36. [PubMed: 7063747]
31. Hosmer, DW.; Lemeshow, S. *Applied Logistic Regression*. New York, NY: Wiley; 1989.
32. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Annals Intern Med* 1993;118:201–210.
33. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387. [PubMed: 8668867]
34. Famoye F. Restricted generalized Poisson regression models. *Communication Stat: Theory Methods* 1993;22:1335–1354.
35. Brewer WF, Sampaio C. Processes leading to confidence in sentence recognition: a metamemory approach. *Memory* 2006;14:540–552. [PubMed: 16754240]
36. Valentijn SA, Hill RD, Van Hooren SA, et al. Memory self-efficacy predicts memory performance: results from a 6-year follow-up study. *Psychol Aging* 2006;21:165–172. [PubMed: 16594801]
37. Chua EF, Schacter DL, Rand-Giovannetti E, et al. Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage* 2006;29:1150–1160. [PubMed: 16303318]
38. Wolinsky FD, Miller TR, Geweke JF, et al. An interpersonal continuity of care measure for Medicare Part B claims analyses. *J Gerontol: Soc Sci* 2007;62BIn press
39. Wolinsky FD, Miller TR, An H, et al. Dual use of Medicare and the Veterans Health Administration: are there adverse health outcomes? *BMC Health Serv Res* 2006;6:131,1–11. [PubMed: 17029643]

TABLE 1
 Weighted Comparison of Self-Report vs. Claims-Based Hospital Episodes*

Self-Reported	Claims-Based						Total
	0	1	2	3	4	5+	
0	3285	43	8	0	0	0	3335
1	193	392	25	6	1	0	617
2	47	67	63	8	3	1	189
3	4	14	17	9	0	1	45
4	3	10	2	7	2	2	26
5+	3	3	4	3	0	3	16
Total	3536	529	119	32	7	7	4229
Agreement	92.8%	74.2	53.0%	27.2%	30.3%	41.4%	88.8%

* Numbers may not total because of rounding.

TABLE 2

Adjusted Odds Ratios and 95% Confidence Intervals from Weighted Multinomial Logistic Regressions of Concordance Categories on Hospital Admissions (n = 4201; Reference = Concordance)

Variables	Overreport		Underreport	
	AOR	(95% CI)	AOR	(95% CI)
Sociodemographic				
Age	0.986	(0.964–1.010)	1.040	(0.999–1.083)
Men	1.495*	(1.059–2.110)	1.835	(0.972–3.467)
Race				
Black	1.164	(0.782–1.732)	1.056	(0.534–2.087)
Hispanic	1.359	(0.803–2.298)	1.220	(0.477–3.121)
Living alone	1.076	(0.833–1.389)	1.276	(0.800–2.036)
Socioeconomic				
Education				
Grade school	0.999	(0.745–1.340)	0.942	(0.571–1.553)
Some college	1.170	(0.876–1.564)	0.523	(0.260–1.053)
Income				
≤ \$7000	0.846	(0.597–1.200)	1.045	(0.590–1.849)
≥ \$50,000	1.099	(0.807–1.496)	0.995	(0.493–2.008)
Veteran	1.075	(0.751–1.541)	0.990	(0.487–2.015)
Private insurance	0.943	(0.696–1.280)	1.029	(0.593–1.788)
Lifestyle				
Smoker (ever)	1.140	(0.887–1.465)	0.784	(0.487–1.263)
Overweight	1.023	(0.797–1.313)	1.072	(0.674–1.704)
Obese	1.141	(0.826–1.576)	0.879	(0.459–1.680)
Drinking				
<1 drink/d	0.839	(0.651–1.082)	0.964	(0.591–1.573)
1–2 drinks/d	0.550*	(0.332–0.910)	0.911	(0.349–2.379)
3+ drinks/d	1.309	(0.635–2.699)	1.206	(0.177–8.203)
Never driven	1.132	(0.760–1.685)	1.621	(0.841–3.124)
Diseases				
Arthritis	1.081	(0.838–1.395)	1.718*	(1.091–2.705)
Cancer	1.653[†]	(1.243–2.197)	0.986	(0.516–1.886)
Diabetes	1.205	(0.887–1.638)	1.650	(0.956–2.849)
Hypertension	1.016	(0.809–1.276)	0.841	(0.546–1.296)
Lung disease	1.190	(0.849–1.667)	1.165	(0.596–2.278)
Heart condition	1.925[‡]	(1.526–2.429)	3.140[‡]	(2.023–4.876)
Hip fracture	1.363	(0.850–2.185)	0.758	(0.281–2.046)
Psych problems	1.462*	(1.019–2.098)	0.861	(0.374–1.982)
Functional limitations				
No. ADLs w/difficulty	0.995	(0.895–1.153)	0.890	(0.682–1.160)
No. IADLs w/difficulty	1.070	(0.929–1.232)	1.179	(0.930–1.495)
No. lower body limits	1.188	(1.083–1.303)	1.144	(0.961–1.361)
Hearing: poor or fair	1.031	(0.798–1.334)	0.643	(0.386–1.071)
Vision: poor or fair	1.195	(0.924–1.545)	1.193	(0.743–1.918)
Memory: poor or fair	0.880	(0.678–1.143)	0.940	(0.582–1.520)
Health: poor or fair	1.396*	(1.064–1.832)	0.925	(0.562–1.523)
Able to drive	0.912	(0.659–1.262)	1.233	(0.670–2.267)
CESD8 = 0	0.839	(0.631–1.114)	0.546*	(0.301–0.992)
CESD8 = 3+	0.954	(0.723–1.258)	0.936	(0.571–1.532)
TICS7 = 0–10	0.939	(0.698–1.263)	1.598	(0.946–2.700)
TICS7 = 14–15	0.800	(0.611–1.048)	0.709	(0.391–1.285)
Number in category		398		101
Pseudo R ²				
Nagelkerke			0.1344	
Cox and Snell			0.0756	
χ ² P value			<0.0001	

* $P < 0.05$;

[†] $P < 0.01$;

[‡] $P < 0.001$.

For convenience, cells with statistically significant AORs are shown in bold.

TABLE 3

Weighted Comparison of Self-Report vs. Claim-Based Physician Visits*

Self-Report	Claims-Based													Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13+
0	250	72	39	29	17	4	6	7	5	3	8	5	0	10
1	136	87	101	57	45	33	17	20	9	8	7	5	4	13
2	127	58	83	66	65	66	35	46	26	11	11	14	6	20
3	107	19	41	60	51	63	42	41	27	24	23	7	14	27
4	102	36	24	49	51	55	48	51	31	26	23	25	22	66
5	31	15	7	9	12	15	25	16	13	14	17	11	10	52
6	61	13	13	12	23	19	28	22	25	24	21	10	10	57
7	6	1	1	3	4	5	6	1	6	7	10	1	3	18
8	10	6	1	1	1	2	1	10	4	12	7	4	9	22
9	4	1	1	0	0	3	0	1	1	4	3	3	2	13
10	12	2	3	1	6	6	6	6	8	10	13	6	10	32
11	2	0	1	0	0	1	0	2	0	0	1	1	2	2
12	45	15	9	8	5	9	12	17	7	5	14	11	9	117
13+	32	3	4	5	2	2	12	3	11	3	7	9	5	105
Total	927	328	326	300	282	282	239	243	172	150	163	112	105	552
Agree%	27.0%	26.4%	25.4%	19.9%	18.1%	5.2%	11.6%	0.5%	2.6%	2.7%	8.0%	0.5%	8.8%	19.0%

* Numbers may not total because of rounding.

TABLE 4

Frequency Distribution for AHEAD Respondents Who Over-, Concordant-, and Underreport Their Number of Physician Visits Under 3 Criteria

	Criterion 3	Criterion 2	Criterion 1
Overreporting	765 (18.3%)	1003 (24.0%)	1362 (32.6%)
Concordant-reporting	2013 (48.1%)	1423 (34.0%)	614 (14.7%)
Underreporting	1404 (33.6%)	1756 (42.0%)	2206 (52.7%)

TABLE 5
AORs from Weighted Multinomial Regressions of Concordance Categories on Physician Visits

Variables	≥3 Overreport	≤-3 Underreport	≥2 Overreport	≤-2 Underreport	≥1 Overreport	≤-1 Underreport
Sociodemographic						
Age	0.999	1.032[‡] 1.198	0.997	1.023[‡] 1.116	0.993	1.021[*] 1.004
Men	0.527[‡]		0.485[‡]		0.606[‡]	
Race						
Black	1.615[‡]	0.894	1.374[*]	0.829	1.410	0.855
Hispanic	2.304[‡]	2.215[‡]	2.055[‡]	2.001[‡]	1.825[*]	1.896[*]
Living Alone	0.807[*]	0.980	0.937	0.985	0.937	0.957
Socioeconomic						
Education						
Grade school	0.866	0.831	0.838	0.878	0.967	1.007
Some college	1.171	1.041	1.067	1.005	1.070	0.992
Income						
≤\$7000	1.063	0.875	0.938	0.883	0.842	0.790
≥\$50,000	0.863 [‡]	1.183	0.871	1.106	1.009	1.205
Veteran	1.563[‡]	1.083	1.531[‡]	0.952	1.071	0.877
Private Insurance	1.125	1.280[*]	1.065	1.317^{**}	1.357[*]	1.720[‡]
Lifestyle						
Smoker (ever)	1.087	1.123	1.210[*]	1.138	1.033	1.105
Overweight	1.168	0.978	1.046	0.989	1.029	1.001
Obese	1.273	0.926	1.092	0.877	0.916	0.822
Drinking						
<1 drink/d	1.072	0.870	0.979	0.863	1.098	0.985
1–2 drinks/d	0.993	0.843	1.038	0.852	1.312	1.098
3+ drinks/d	0.802	0.647	0.833	0.727	0.978	0.797
Never driven	0.855	1.058	0.814	0.946	0.864	0.871
Diseases						
Arthritis	1.346[‡]	1.432[‡]	1.511[‡]	1.618[‡]	1.919[‡]	2.107[‡]
Cancer	1.458[‡]	1.444[‡]	1.537[‡]	1.529[‡]	1.399[*]	1.469[*]
Diabetes	1.462[‡]	1.481[‡]	1.255	1.265	2.134[‡]	2.151[‡]
Hypertension	1.544[‡]	1.166[*]	1.862[‡]	1.401[‡]	1.773[‡]	1.511[‡]
Lung disease	1.360	1.203	1.428[*]	1.151	1.369	1.198
Heart condition	1.380[‡]	1.393[‡]	1.223	1.315[‡]	1.168	1.252
Hip fracture	1.310	1.140	1.063	0.932	0.819	0.847
Psych problems	1.163	1.322	1.150	1.289	1.074	1.231
Functional limitations						
No. ADLs w/difficulty	1.000	0.979	0.983	0.953	0.909	0.858
No. IADLs w/difficulty	0.925	0.896	0.874[*]	0.908	0.997	1.055
No. Lower body limits	1.161[‡]	1.115[‡]	1.151[‡]	1.111[‡]	1.132[‡]	1.089
Hearing: poor or fair	0.978	0.925	1.037	0.935	0.927	0.917
Vision: poor or fair	0.986	0.905	1.056	0.971	0.819	0.821
Memory: poor or fair	1.341[‡]	1.345[‡]	1.263[*]	1.276[‡]	1.273	1.239
Health: poor or fair	1.062	1.043	1.092	1.075	1.162	1.101
Able to drive	1.000	1.016	1.016	1.086	1.079	1.231
CESD8 = 0	0.946	1.007	0.946	1.005	1.000	1.037
CESD8 = 3+	1.185	0.989	1.161	0.933	1.120	0.952
TICS7 3 0–10	1.162	1.149	1.251	1.107	1.231	1.206
TICS7 = 14–15	1.051	1.178	1.089	1.116	1.151	1.223
Pseudo R²						
Nagelkerke		0.1079		0.1040		0.0839
Cox and Snell		0.0944		0.0920		0.0724

Variables	≥ 3 Overreport	≤ -3 Underreport	≥ 2 Overreport	≤ -2 Underreport	≥ 1 Overreport	≤ -1 Underreport
-----------	---------------------	-----------------------	---------------------	-----------------------	---------------------	-----------------------

* $P < 0.05$;
† $P < 0.01$;
‡ $P < 0.001$.

For convenience, cells with statistically significant AORs are shown in bold.