Evaluation of Disease Status based on Patient Self-report in the Medicare Health Outcomes Survey:

Using Linked Data from Surveys and Computerized Medical Data from the Veterans Health Administration

Donald R. Miller, Sc.D., William H. Rogers, Ph.D., Avron Spiro III, Ph.D., and Lewis E. Kazis, Sc.D.

December 19, 2003

Technical Report prepared by:
The Health Outcomes Technologies Program, Health Services Department, Boston University School of Public Health, Boston, MA. & The Institute for Health Outcomes and Policy, Center for Health Quality, Outcomes and Economic Research, Department of Veterans Affairs, Bedford, VAMC, MA.

Supported by the Centers for Medicare & Medicaid Services through the National Committee for Quality Assurance, the Health Outcomes Technologies Program, Health Services Department, Boston University School of Public Health, and the Center for Health Quality, Outcomes and Economic Research, Department of Veterans Affairs.

Questions concerning this work can be e-mailed to Drs. Miller, Rogers, or Kazis at:
drmiller@bu.edu, whrogers@comcast.net, lek@bu.edu

This report does not reflect the viewpoints of any of the organizations supporting this work

The Centers for Medicare & Medicaid Services' Office of Research, Development, and Information (ORDI) strives to make information available to all. Nevertheless, portions of our files including charts, tables, and graphics may be difficult to read using assistive technology.

Persons with disabilities experiencing problems accessing portions of any file should contact ORDI through e-mail at ORDI_508_Compliance@cms.hhs.gov.

Executive Summary

In this study, we evaluated the disease self-report questions from the Medicare Health Outcomes Survey (HOS) using linked Veterans Health Administration (VA) data for patients who are dually eligible. Responses to questions from the survey were used to classify respondents for 12 diseases and these were compared with classifications based on either similar questions from a VA survey or diagnostic codes from VA medical records. Agreement between classifications based on the two surveys was good with over 75% of those patients affirming the disease in the HOS also affirming it in the VA survey for most of the diseases; lower confirmation for other diseases was likely attributable to differences in the wording of the questions. HOS disease status also agreed reasonably well with VA based disease status using diagnostic codes. For most diseases, those patients who did not have the disease in the medical record were unlikely to report it in the survey (good specificity – 70-94%) and the probability that patients would report the disease given its indication in the medical records was generally good (sensitivity – 65-85%). There was considerable variability by medical condition although there were reasonable explanations for the relatively poor measures of agreement for some of the conditions. These findings varied only slightly by education, age, and race. As a further evaluation, independent decrements in patient reported health status derived from the SF-36 associated with each disease based on the survey questions were compared in the two surveys and found to be similar. These results suggest that patients can provide reasonably good reports of their morbidity in survey questions. This is critical information to consider in evaluating and improving health care.

Introduction

Information on medically diagnosed diseases in patient populations has become a critical element of case-mix measures used in health services research and quality of care monitoring systems [1]. In the past, disease information usually had to be collected through medical chart abstractions, which are relatively labor intensive and expensive [2-4]. The increasing computerization of medical data has made measures of disease status through diagnostic codes readily available, although such measures may result in some misclassification. As an alternative, patients' reporting of their diagnosed diseases through surveys has become a frequent source of information on morbidities [5], particularly with the increasing emphasis placed on patient perceptions in the assessment of health care quality [6].

Rigorous and systematic evaluation of disease measures is essential to understand their utility, yet such evaluation of patient self-report measures is often lacking. While a number of studies have compared various sources of data on morbidities [7-21], there remains uncertainty about their accuracy and value. As a foundation for future quality of care monitoring and research, patient reported morbidity data should be subjected to rigorous evaluation for both accuracy and completeness.

Since 1998, the Centers for Medicare and Medicaid Services (CMS, previously called Health Care Finance Administration) has administered surveys to samples of patients enrolled in Medicare + Choice health plans in order to measure changes in health status as part of a health care quality monitoring system. The Medicare Health Outcomes Survey (HOS, previously called the Health of Seniors Survey) includes a patient-based measure of health status (MOS SF-36 version 1.0) along with a series of questions assessing the patients' recall of doctor reported diagnoses of a number of medical conditions. While these questions on diagnosed diseases have

been used and evaluated in other studies (22-24), it would be very useful to know how accurate and informative they are in the Medicare population surveyed.

Unfortunately, it was not possible to evaluate these questions using patient medical data from medical encounters covered by Medicare, since these were not available. There were, however, a number of Medicare patients who completed the HOS who were also dually eligible for health care from the Veterans Health Administration (VA). For these patients, diagnosed disease status reported in the HOS could be evaluated in a number of ways using data from the VA. First, as a measure of reliability, they could be compared directly with responses to similar questions from a comparable survey completed by many of the dually eligible patients (22). Second, as a measure of concurrent validity, HOS survey responses could be compared with disease status based on diagnostic codes recorded at medical encounters in the VA (4,25). Third, the disease burden as measured by the SF-36 (26-28) associated with each diagnosed disease could be compared in the two patient populations using data from the two surveys, as a measure of predictive validity. This evaluation provides a means to gauge the value of the self-reported disease questions from the HOS, which are important elements in the case-mix adjustment of health status change among health plans and in the stratification of patients by diseases.

Methods

Samples

The Medicare HOS survey has been conducted in the spring of each year since 1998. Approximately 300,000 beneficiaries enrolled in Medicare + Choice plans are approached each year for a baseline assessment and the responders form a cohort. Two years after the initial survey, approximately 90,000 beneficiaries from the cohort are approached of whom about

75,000 respond to a follow-up survey. As of this writing, survey data are available from four cohorts of patients for the years 1998 through 2001, with follow-up assessments from the first three cohorts in years 2000 through 2002.

Many Medicare beneficiaries are veterans of military service and also may be eligible for health care from the Veterans Health Administration (VA) (29). Those veterans who applied for VA eligibility based on poverty, the presence of a medical condition that was related to military service, or any of a number of other conditions would be included in the VA enrollment file. In addition, for all veterans who obtained health care at a VA facility, there is a record on file of their use of VA services. The VA maintains a centralized automated record system of all VA medical encounters including outpatient visits and stays at hospitals and long term care facilities (25,30,31). These records include ICD-9-CM diagnostic codes based on provider notes of patient evaluation and treatments at each medical encounter. Over 4 million patients are treated each year in this medical care system.

The VA has conducted a number of surveys of their patient populations, the largest of which was the "1999 Large Health Survey of Veteran Enrollees" (VA Survey or VAS) (22). This survey included an instrument on health status (Veterans SF-36) and questions on diagnosed diseases similar to those in the HOS. For the VAS, a stratified random sample of 1,406,049 veterans eligible for VA care were identified and contacted, of whom 887,775 or 63% responded.

In conducting this evaluation, we linked patient records from both CMS and the VA in order to compare diagnostic classifications of patients from the various sources. Assessment of health systems operations is a permitted use under HIPAA. Patient records were matched using Social Security Numbers and Health Insurance Claim Numbers (HICnums) (32) using a method

previously shown to be at least 99.8% accurate. After matching and before analysis, all personal identifying information was removed.

Our linking efforts yielded 87,636 match able HOS surveys from 64,599 unique respondents with potential VA eligibility across the 4 HOS cohorts. For the comparison of HOS disease assignments with those from the VA survey, we identified 7,953 patients with VA use who completed both surveys. In further analysis, we found 4,252 of them completed the HOS prior to the VAS. For the comparison of HOS disease assignments with those using VA diagnostic codes, we limited the sample to 17,089 patients who had VA medical encounters in the two years prior to the survey administration. In a further analysis, we limited it to the 8,987 patients with VA use in all four of the prior years. For the comparison of decrements in health status (SF-36) associated with self-reported diseases between the two surveys, we used survey data from 435,911 respondents in the first two cohorts of the HOS and 731,304 respondents of the VAS.

Disease Classification

Questions on disease status based on recall of a medical diagnosis have been used in a number of studies including several large national surveys (9,22,23). Both the HOS and the VAS used questions that were identical or very similar to those from these other studies. The HOS had questions assessing 13 patient conditions including four types of cancer. A subset of these questions was used in this evaluation and these are listed in Table1 showing the exact wording of the questions. For comparison, the exact wording of comparable questions from the VAS are also shown in Table 1

Diagnostic (ICD-9-CM) codes (33) were obtained from all inpatient and outpatient visits

in the VA for a period of time prior to the date the survey was completed. The lists of specific codes used for identifying each condition are shown in Table 1. These lists were abstracted from the ICD9-CM coding manual and/or derived from published reports (34-36) with modifications made after extensive review by VA clinicians. Patients were classified as having the disease if one or more of the codes for that condition were present in the medical records in the specified time periods preceding the survey. Use of a two year time period is based on prior work in identifying VA patients with specific diseases from the medical records (37). The additional analysis using a four year period was done to give more opportunity for detecting conditions in the medical record that may have been diagnosed further in the past or with less frequent diagnostic coding.

Health Status and Demographics

For the analysis of health status decrements associated with each medical condition, we used the MOS SF-36 (version 1.0) and the Veterans SF-36 instruments that were administered in the HOS and VAS, respectively. The MOS SF-36 is a well established survey of health status (26-27); the Veterans SF-36 is a modified and improved version of the MOS SF-36 (28). Both have been widely used and evaluated with demonstrated reliability and validity in a variety of patient populations (22,28). Both instruments measures eight concepts of health: physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, energy/vitality, social functioning, role limitations due to emotional problems, and mental health. It also is used to generate physical (PCS) and mental health (MCS) component summary scores, which are weighted summaries of the eight scales. PCS and MCS are standardized with a norm of 50 and a standard deviation of 10 to the general United States population. Lower scores

denote worse health.

The Veterans SF-36 includes modifications to the MOS SF-36 in the response choices of the role physical and role emotional items. The dichotomized two point yes/no choices were changed to five-point Likert scales in order to reduce floor and ceiling effects. With the exception of these role scales and the change items for physical and emotional health, scoring of the Veterans SF-36 scales is the same as that for the MOS SF-36.[1] This process includes a linear transformation from a raw score so that scores range from 0 to 100, where 100 denotes the best health. Scoring of the Veterans SF-36 RP and RE scales and physical (PCS) and mental (MCS) summaries uses an algorithm previously developed and validated to ensure comparability with the MOS SF-36 (22,28).

Information on patient sex, age, race, and education were obtained from responses to questions in the HOS. Each patient's baseline HOS was used preferentially. When data were missing from the baseline HOS, we attempted to reduce missing values using additional information from the follow-up HOS and the VAS, if they were available.

Statistical Methods

For the assessment of reliability, self-reported morbidity status based on questions from the two surveys were cross-tabulated and compared. The proportion of those reporting the disease in each survey was used to estimate prevalence. Since our goal was to evaluate the HOS questions, our primary measure in the comparison of the two surveys was confirmation of the HOS disease status with responses from the VAS. We present the proportion of those who affirm the disease in the HOS who also affirm it in the VAS. This is equivalent to a predictive value positive for the HOS if we considered the VAS response as the standard (39). As

additional measures of concordance, we also present the percent agreement and the kappa statistic (40, 41). The percent of overall agreement is computed as the percent of all individuals in the sample whose measures of morbidity from the two surveys agree (either both negative or both positive). The kappa statistic is the percent agreement corrected for chance (42). While these two statistics do serve as summary measures of agreement, they are used cautiously since they are difficult to interpret and may obscure critical reasons for lack of agreement (42,43). For purposes of discussion, we use descriptions of kappa values suggested by Landis and Koch [44]. This analysis was conducted initially in all patients in the sample who completed both surveys. To eliminate misclassification from new diseases diagnosed soon after completing the HOS and to minimize any effects of prompting from a first survey, the analysis was repeated limiting the sample to those who completed the HOS prior to the VAS.

For the assessment of concurrent validity, morbidity status based on self-report from the HOS and based on diagnostic codes from VA medical records were cross-tabulated and compared. Disease prevalence from the two sources was computed as the proportion of those affirming the disease in the survey or having diagnostic codes for the disease in the VA records. In this analysis we considered the disease status based on VA records as the standard, even though there may be a number of reasons why these records are incomplete or inaccurate. Indication of diseases from the HOS was evaluated as a screening test against the VA medical record status (39,40). *Sensitivity* was computed as the proportion of those with the disease based on diagnostic codes from the medical record in the VA who also affirm the disease in the HOS survey. *Specificity* was computed as the proportion of those without the disease based on diagnostic codes who also do not affirm the disease in the HOS survey. *Predictive value positive* was computed as the proportion of those who affirm the disease in the HOS who also have

indication of the disease in the diagnostic codes from the medical record in the VA. This analysis was conducted in similar fashion using two and four year periods for assessing disease status based on diagnostic codes.

For the analysis of predictive validity, we compared the associations of each patient reported medical condition with health status (PCS and MCS) obtained from analyses of the two surveys. With PCS and MCS as the dependent variables, we performed multivariate analysis using ordinary least-square regression. Models included indicator terms for each of the medical conditions that were measured in both surveys (see Table 1) as well as terms for gender (female), age (55-64, 65-74, 75-84, 85+ years), and race/ethnicity (African-American, Hispanic, other race, unknown race). Exploratory analyses were done to evaluate interactions among the medical conditions and between them and with the demographic variables. Few significant interactions were observed in analyses from either sample and they had virtually no effect on the main effects coefficients. For this reason, we present only the parsimonious models containing main effects terms. To allow for direct comparison of coefficients for each medical condition, identical models were used with data from the two survey samples. These coefficients may be interpreted as decrements in PCS or MCS associated with the presence of the condition, independent of other factors in the models.

Results

The demographics and health status measures of the various samples are presented in Table 2. Among the 64,599 HOS respondents who were eligible for VA care, the average age was 72 years and 95% were 65 years of age or older. Also, 95% of them were men, 88% (of those with known race) were Caucasian, and 27% did not complete high school. Limiting the

sample to those with VA use in the two years prior to the survey or those who also completed the VAS made only small changes in the demographic profile, slightly increasing the proportion of respondents who were male, non-white, and did not complete high school. Health status measures were similar in the three groups with a mean PCS of 34 to 36 and a mean MCS of 47 to 48, and both were comparable to the national VA mean norms for those 65 years of age and older of 34.5 for PCS and 46.9 for MCS.

Reliability

The comparison of responses from the two surveys for those who completed both surveys is presented in Table 3A. Disease prevalence is relatively high in these dually eligible patients. For most of the conditions analyzed prevalence rates from the two surveys were similar, within one or two percentage points. The two exceptions are chronic low back pain and arthritis, but these differences are almost certainly due to differences in wording of the questions from the two surveys (Table 1). The higher prevalences from the VAS are consistent with the broader scope of the questions in that survey (i.e. chronic low back pain, not just sciatica; and arthritis, not just limited to hip or knee and hand or wrist). Confirmation rates were also good, with over 75% of those patients who affirmed the disease in the HOS also affirming it in the VAS for each of the medical conditions except for chronic low back pain (68%). The other statistics showed comparable results with over 85% overall agreement and kappa statistics that are considered to be substantial (0.6-0.8) (44) for all diseases except chronic low back pain and arthritis.

Restricting the sample to those who completed the HOS before the VAS (Table 3B) improved these statistics, but only slightly. Confirmation rates increased between 1% and 7% but there was very little change in the overall agreements and the kappa statistics.

Concurrent validity

Table 4A presents the comparison of morbidity status based on HOS responses with those based on VA diagnostic codes for those HOS respondents who used VA health care in the two years before the survey. Prevalence estimates from the two sources were within two percentage points for diabetes, hypertension, chronic lung diseases, and cancer, but the estimates from the HOS self-reports were higher than those from VA diagnostic codes for the other conditions.

This discrepancy is likely related to several factors. The two year window may not be a sufficiently long enough time to pick up conditions that present as discrete episodes, such as myocardial infarction and stroke, which may have occurred in the more distant past. Some symptomatic conditions such as chronic low back pain, arthritis, and angina, may be reported by patients but under diagnosed in the medical records (4,24). In addition, these patients are splitting their care between VA health services and those covered by Medicare and it is possible that some diseases may be treated exclusively in one health care setting and not the other, so that diagnostic codes would not appear in the VA record.

There was also variability by medical condition in measures of concordance between the survey and the VA medical record. Questions on diabetes, chronic lung disease, congestive heart failure, stroke, and the cancers had high specificity of 85% or higher. This means that patients who do not have the disease in the medical record are unlikely to report it in the survey. Specificity was moderately good (70%-85%) for hypertension, angina, myocardial infarction, and chronic low back pain, and worse (55%) for arthritis, but these less favorable statistics may be related to the limitations of these codes that were discussed previously. Sensitivity, or the

probability that the patient would report the disease given its indication in the medical records, was moderate to good (65-85%) for most of the conditions except for chronic low back pain, congestive heart failure, and lung and colon cancer which had lower sensitivity. Predictive value positive, or the probability for self-reported disease to be confirmed in the medical record, was more variable and generally modest (19% to 65%), except for diabetes and hypertension which both had values of 84%. Predictive value positive tends to be lower with decreasing prevalence and may be influenced more by the potential problems of diagnostic codes discussed previously.

This analysis was repeated using a four year window and restricted to those patients who used VA services in the four years prior to the survey (Table 4B). This was done to give more opportunity for detecting conditions in the medical record that may have been diagnosed further in the past or with less frequent diagnostic coding. This did result in higher disease prevalences based on diagnostic codes and, for several conditions (chronic low back pain, angina, myocardial infarction, stroke, and lung and colon cancers), the increases were substantial. This was accompanied by sizable increases in predictive value positive and modest increases in specificity. It seems likely that even longer periods of surveillance for diagnostic codes for these conditions would result in further improvements in concordance. This suggests that self-reported disease classification using HOS self-reported questions may be more accurate than is evident from the evaluation using diagnostic codes from the medical record.

Variation by demographics

Our evaluation of how the performance of self-reported disease questions varies by patient demographics is summarized in Table 5. To illustrate our findings, results are presented for four of the medical conditions (diabetes, chronic lung disease, arthritis, and cancer); results

for the other medical conditions were similar. Variation was examined by age (above and below the median of 72 years, race (white, non-white), and education (less than high school, high school graduate but no college, college). There were insufficient numbers of women to evaluate variation by gender.

Prevalence of cancer and arthritis was higher in older patients, but measures of reliability (comparison of HOS with VAS) and concurrent validity (comparison of HOS with VA diagnostic codes) varied little between age groups. The only notable differences were that older patients had higher confirmation of self reported cancer in the VAS, higher sensitivity to identify cancers indicated by diagnostic codes in the VA records, but lower sensitivity for identifying diabetes and chronic lung disease.

There was also only slight variation in performance measures by race. Except for higher rates of diabetes and lower rates of self-reported cancer in non-whites, prevalence rates showed little variation with race. Chronic lung disease and cancer were confirmed in the VAS at a higher rate in whites, and the HOS question on chronic lung disease had higher sensitivity in whites, but the cancer question had higher predictive value positive for cancer based on diagnostic codes in non-whites.

The greatest variation in the performance of the morbidity questions was found with education but, even here, the variation was modest. The more educated patients had lower prevalence of diabetes, chronic lung disease, and arthritis but, since education is strongly inversely related to age in these populations, this could be a function of age differences. HOS responses to disease questions were confirmed in the VAS at a higher frequency among the more educated patients for all diseases except arthritis, and kappa statistics tended to increase with education for all diseases except cancer. In terms of comparisons with the VA diagnostic codes,

specificity varied little with education, but sensitivity increased for all conditions except cancer.

Predictive validity

To further evaluate these questions, we present multivariable linear regression models to estimate the independent decrement in health status associated with each self-reported medical condition, and we compare results from the HOS and VAS (Table 6). With identical terms in the models, overall model performance was quite similar between the two samples with R square values about 32.0%-32.5% for the PCS models and 12.7%-14.5 % for the MCS models. The intercepts were slightly higher in the VAS sample and there were some inconsistencies in the associations with the demographic variables. For example, relative to those less than 55 years old, physical function (PCS) was lower and declined with increasing age above age 65 years in the VA sample but was higher in the older Medicare patients. This is not surprising given the level of physical disability of younger Medicare beneficiaries in comparison with those over age 65 years. Being female also had opposite effects in the two samples but associations with race were largely consistent.

In spite of these differences, the associations between the self-reported medical conditions and the measures of health status were reasonably consistent between the two samples. They are also consistent with published reports of similar analyses from other populations (38). The largest decrements in health (4.4 points or more in PCS) were found in both samples for arthritis, chronic low back pain, and chronic lung disease, followed by congestive heart failure and stroke (3.5 to 4.3 PCS points). The other conditions (diabetes, angina, cancer, hypertension, and myocardial infarction) were associated with PCS decrements of 1.0 to 2.7. Although disease associated decrements for PCS were higher in the model from

the HOS for all conditions except for chronic low back pain and MI, the differences were small and the relative ranking of decrements by disease were much the same in the two models.

Summary and Discussion

These results indicate that patients are fairly reliable reporters of their medical conditions. Most diseases reported in the HOS were confirmed in the VAS and were evident in VA medical records as indicated by the presence of appropriate diagnostic codes. Our findings on agreement in morbidity status comparing self report with medical records are generally consistent with those found in a number of other studies (7-21). This evaluation provides some evidence that these patient reported measures are likely to be useful for their intended purposes of case-mix adjustment of health status change among health plans and stratification of patients by diseases.

There are a number reasons why the levels of agreement in our study were not better. First, there is probably some real error in self reported morbidity because patients may not have understood the questions, may have recalled what their doctors told them inaccurately, or may have misinterpreted information received as part of health care (8,9). For example, patients may have interpreted diagnostic testing, such as blood pressure or blood glucose measurements, as indication of disease even though the test results were negative. One would expect this type of error to be related to the patient's level of education. Although measures of agreement did improve with education in our study, the differences were very modest. Patients may have also reported the disease when the doctor did not actually give them a diagnosis but simply discussed the possibility with them. This may be particularly the case for conditions with diagnosis based on symptoms such as chronic low back pain, arthritis, and angina. In other cases, patients may actually have the condition but if they were not prescribed medication or did not otherwise

receive specific treatment, appropriate diagnostic codes may not have been placed in the medical record. Nevertheless, these patients would experience the disease and this would have an impact on their reported health status measures.

There are other reasons for lack of agreement that do not necessarily reflect the validity of the survey questions but more the limitations of the evaluation. Patients have a more comprehensive view of their previous care than may be captured in a single system assessment such as we did in this evaluation. When they are asked if a doctor ever told them if they have a condition, they can refer to care experiences over many years in the past and across a variety of providers and multiple systems of care. In our evaluation, we reviewed only two and four years of diagnostic records from VA care prior to the survey. Most conditions with ongoing treatment in the VA should be captured in these records, but we might have missed those diagnoses that are acute events, such as myocardial infarction, stroke, and cancer, which occurred in the more distant past and might not have required recent treatment. We might have also missed certain conditions, such as arthritis and chronic low back pain that are episodic in nature, since an episode might not have occurred during the period of record reviews. We have presented some evidence for these phenomena in that these are the very conditions with significant increases in prevalence when the review period was increased from two to four years.

Diagnostic codes in VA medical records might also be missing because coding may be too conservative for some conditions. There were some indication of this in Kashner's examination of VA administrative data in comparison with abstractions from medical charts (4), and such a finding has also been corroborated for diagnostic codes in other health care systems (17, 40).

Another reason for lack of confirmatory codes in VA medical records is that some

patients seek care for certain chronic conditions from providers outside of the VA and thus the code does not appear for visits to the VA. This includes acute episodes of disease, such as myocardial infarctions, that might be treated on an emergency basis and resolved at a hospital closer to the patient's residence. It also might include conditions for which patients often exercise preferences in terms of type and place of care, such as cardiac surgery or cancer treatment (45). These conditions may be treated at hospitals or clinics outside of the VA, even at higher cost. This also covers care through alternative or complementary health-care providers, who frequently treat chronic conditions such as arthritis or chronic low back pain.

In addition to unconfirmed self reported disease there were a smaller number of patients with diagnostic codes for conditions that they did not affirm in the survey. Patients may have forgotten that they were told they had a diagnosis, or they may not have been told even though the indications were there in the medical records. Studies from taped patient encounters indicate that over 20% of recorded patient information is not discussed with the patient (40). This may be particularly true for sicker patients requiring lengthy, complicated care, or for patients with cognitive dysfunction or psychiatric problems. It has also been reported that certain conditions may be over-coded, resulting in diagnoses that were not actually made and discussed with the patient (4,46-49).

Together these may explain some of the lack of agreement between self reported disease and disease status based on diagnostic codes. There is good reason to believe that the value of the self reported disease questions is even better than what is indicated in this evaluation. This is supported by our analysis of predictive validity. Patient report of medical conditions is associated with substantial decrements in physical and mental health scores that are predictable for each disease across patient populations. The consistent associations of self reported diseases

with measures of health status has implications for their use as a case-mix measure. While it would be preferable to have additional measures of disease from medical records, patients can provide reasonably good reports of their morbidity status that carry importance about their perceived illness burden. This is critical information to consider in evaluating and improving health care.

References

1. Iezzoni LI. **Risk Adjustment for Measuring Health Care Outcomes**. Chicago, IL: Health Administration Press; 1997.
2. Lloyd SS, Rissing R. Physician and coding errors in patient records. **Journal of American Medical Association** 1985; 254:1330-1336.
3. Romm FJ, Putnam SM. The validity of the medical record. **Medical Care** 1981; 19:310-315.
4. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. **Medical Care** 1998; 36:1324-1336.
5. Kravitz RL, Greenfield S, Rogers W, Manning WG, Zubkoff M, Nelson E, Tarlov AR, Ware JE. Differences in the mix of patients among medical specialists and systems of care: Results from the Medical Outcomes Study. **Journal of American Medical Association** 1992; 267:1617-1623.
6. Greenfield EC, Nelson S. Recent developments and future issues in the use of health status assessment measures in clinical settings. **Medical Care** (supple.) 1992; 30:MS23-MS41.
7. Commission on Chronic Illness. **Chronic Illness in the United States, Vol. IV.:Chronic illness in a large city: the Baltimore study**. Cambridge, MA: Harvard University Press. 1957; 297-328.
8. Madow WG. **Interview Data on Chronic Conditions Compared with Information Derived from Medical Records**. Vital and Health Statistics. Series 2, No.23. Washington, DC: National Center for Health Statistics; 1967.
9. National Center for Health Statistics**. Health Interview Responses Compared with Medical Records.** Vital and Health Statistics. Series 2, No. 7. Rockville, MD: National Center for Health Statistics; 1965.

10. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, Hennekens CH, Speizer FE. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. **American Journal of Epidemiology**1986; 123:894-900.
11. Tretii S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. **Journal of Epidemiology and Community Health** 1982; 36:269-273.
12. Linet MS, Harlow SD, McLaughlin JK, McCaf- frey LD. A comparison of interview data and medical records for previous medical conditions and surgery. **Journal of Clinical Epidemiology** 1989; 42:1207-1213.
13. Bush TL, Miller SR, Golden AL, Hale WE. Self- report and medical record report agreement of selected medical conditions in the elderly. **American Journal of Public Health** 1989: 79:1554-1556.
14. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. **Medical Care** 1997; 35:9324-9327.
15. Bowlin SJ, Morrill BD, Nafziger AN, Lewis C, Pearson TA. Reliability and changes in validity of self-reported cardiovascular disease risk factors using dual response: the behavioral risk factor survey. **Journal of Clinical Epidemiology** 1996; 49:511-517.
16. Gambassi G, Landi F, Peng L, BrostrupJensen C, Calore K, Hiris J, Lipsitz L, Mor V, Bernabei R. Validity of diagnostic and drug data in standardized nursing home resident assessments: potential for geriatric pharmacoepidemiology. SAGE Study Group. Systematic Assessment of Geriatric drug use via Epidemiology. **Medical Care** 1998; 36:1677-1679.
17. Newschaffer CJ, Bush TL, Penberthy LT. Comorbidity measurement in elderly female breast

cancer patients with administrative and medical records data. **Journal of Clinical Epidemiology** 1997; 50:725-733.

18. Muhajarine N, Mustard C, Roos LL, Young TK, Gelskey DE. Comparison of survey and physician claims data for detecting hypertension. **Journal of Clinical Epidemiology** 1997; 50:711-718.
19. Bergmann MM, Byers T, Freedman DS, Mokdad A. Validity of self-reported diagnoses leading to hospitalization: comparison of self-reports with hospital records in a prospective study of American adults. **American Journal of Epidemiology** 1998; 147:969-977.
20. Lampe FC, Walker M, Lennon LT, Whincup PH, Ebrahim S. Validity of a self-reported history of doctor-diagnosed angina. **Journal of Clinical Epidemiology** 1999; 52:73-81.
21. Martin LM, Leff M, Calonge N, Garrett C, Nelson DE. Validation of self-reported chronic conditions and health services in a managed care population. **American Journal of Preventive Medicine** 2000; 18:215-218.
22. Kazis L, Perlin J, Skinner K, Ren XS, Lee A, Rogers W, Spiro A, Selim A, Miller DR. **1999 Large Health Survey of Veteran Enrollees: Health status and outcomes of veterans**: Veterans SF-36 physical and mental component summary scores, Executive Report, Washington, D.C.: Office of Quality and Performance, Veterans Health Administration, May, 2000.
23. Kazis LE, Miller DR, Skinner KM, Lee A, Clark JA, Rogers WH, Spiro A III, Selim AJ, Linzer M, Payne S, Mansell D, Fincke G. Patient reported measures of health: The Veterans Health Study. Journal of Ambulatory Care Management 2003, in press.
24. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self reports and medical records for chronic conditions: Experience from the Veterans Health

Study. Journal of Ambulatory Care Management, 2004, in press .

25. Cowper DC, Hynes DM, Kubal JD, Murphy PA. Using administrative databases for outcomes research: select examples from VA Health Services Research and Development. **Journal of Medical System** 1999; 23:249-259.
26. Ware JE, Sherbourne CD. The MOS 36-item Short Form Health Survey (SF-36), I. Conceptual framework and item selection. **Medical Care** 1992; 30:473-483.
27. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. **Medical Care** 1993; 31:247-263.
28. Kazis LE, Miller DR, Clark J, Skinner K, Lee A, Ren XS, Spiro III A, Rogers W, Ware J. Improving the response choices on the SF-36 role functioning scales: Results from the Veterans Health Study (Veterans SF-36). Journal of Ambulatory Care Management, 2004, in press.
29. Shen Y, Hendricks A, Zhang S, Kazis LE. VHA enrollees' health care coverage and use of care. Med Care Res Rev 60:, 2003.
30. Boyko EJ, Koepsell TD, Gaziano JM, Horner RD, Feussner JR. US Department of Veterans Affairs medical care system as a resource to epidemiologists**. American Journal of Epidemiology** 2000; 151:307-314.
31. Ashton CM, Menke TJ, Deykin D, Camberg LC, Charns MP. A state-of-the-art conference on databases pertaining to veterans' health. A resource for research and decision making. **Medical Care** 1996; 34(Suppl.):MS1-MS8.
32. Agency for Health Care Policy and Research**. Report to Congress: The feasibility of linking research-related data bases to Federal and non-Federal medical administrative**

databases. Agency for Health Care Policy and Research Pub. No. 91-003. Rockville, MD: Agency for Health Care Policy and Research; 1991.

33. **International Classification of Diseases, 9th Revision**, Clinical Modifications, 5th Edition, Medicode, 1996.
34. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. **Medical Care** 1998; 36:8-27.
35. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. **J Clin Epidemiol**. 1992; 45: 613-619.
36. Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J. Developing a prediction rule from automated clinical databases to identify high-risk patients in a large population with diabetes. **Diabetes Care**. 2001; 24: 1547-55.
37. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Veterans Health Administration based on computerized patient data. **Diabetes Care**, 2003, in press. (is this in press????)
38. Ware JE, Kosinski M, Keller SD. **SF-36 Physical and Mental Health Summary Scales**: A User's Manual. Boston, MA: The Health Institute; 1994.
39. Sackett DL, Haynes RB, Tugwell P. Clinical Epidemiology. Boston: Little Brown and Co.; 1985.
40. West SL, Strom BL. Validity of pharmacoepidemiology drug and diagnosis data. In: Pharmacoepidemiology. 2nd Edition. 1994.
41. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; I:307-310.
42. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. American

Journal of Epidemiology 1987; 126:161-169.

43. Seigel DG, Podger MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. American Journal of Epidemiology 1992; 135:571-578.
44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174.
45. Wright SM, Daley J, Peterson ED, Thibault GE. Outcomes of acute myocardial infarction in the Department of Veterans Affairs: Does regionalization of health care work? **Medical Care** 1997;35:128
46. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM. Identifying persons with diabetes using Medicare claims data. **American Journal of Medical Quality** 1999; 14:270-277.
47. Woods CR. Impact of different definitions on estimates of accuracy of the diagnosis data in a clinical database. **Journal of Clinical Epidemiology** 2001; 54:782-788.
48. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy if the ICD-9-CM in identifying the diagnosis of ischemic cerebrovascular disease. **Neurology** 1997; 49:660-664.