IDENTIFYING RURAL ELDERLY INDIVIDUALS WITH DIABETES MELLITUS USING MEDICARE CLAIMS DATA

Paul L. Hebert B.A.
Barbara P. Yawn M.D.
A. Marshall McBean M.D.
Rural Health Research Center
Division of Health Services Research and Policy
School of Public Health
University of Minnesota

Working Paper #22

March 1998

Support for this paper was provided by the Office of Rural Health Policy, Health Resources and Services Administration, PHS Grant No. CSRUC 0002-01.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Health care utilization data have been used to estimate the prevalence of certain diseases, to study disease-specific care, and to evaluate the quality of health care provided. In urban areas, large health plans, group practices or insurers often develop computerized data bases to provide the necessary utilization data. In rural areas, large health plans or large group practices are uncommon and few smaller practices have either computerized databases or sufficiently large patient populations to allow meaningful study of quality or disease prevalence. The only large computerized databases available in many rural areas are maintained by Medicare and Medicaid. With their large elderly population, rural regions may especially benefit from the analysis of Medicare data. However, whether Medicare claims data can be used to identify persons with specific disease states such as diabetes has been questioned.

This study investigates the accuracy of claims-based case definitions for diabetes by comparing self-report diabetes to diabetes-related diagnoses appearing on Medicare claims. The goal of this research is to construct a valid cohort of rural elderly individuals with diabetes, using Medicare claims data. The construction of a valid cohort is the critical first step in assessing the health status of and quality of care for rural elderly individuals with diabetes.

Using data for individuals who participated in the Medicare Current Beneficiary Survey, we compare self-reported diabetes to diabetes-related diagnosis on all types of Medicare utilization claims, including inpatient hospital, skilled nursing facility, home health agency, outpatient hospital, and physician/supplier claims. Several case definitions for diabetes then were constructed by combining and weighting diabetes-related diagnoses from these claims, and by varying the number of years of claims used. Case definitions were evaluated for agreement with self-report. Individuals who were misclassified by various case definitions were analyzed for potential racial or other biases introduced by the case definitions.

Diabetes-related diagnoses on Medicare claims are highly specific and predictive indicators of the presence of the disease. The specificity of diabetes-related diagnoses on short stay hospital, skilled nursing facility and home health claims was greater than .99, but the sensitivity was low (sensitivity <.34). Ambulatory care claims were less specific (.97) but far more sensitive (.67). The specificity of a case definition based on ambulatory claims can be improved by requiring two or more diabetes-related diagnoses for a positive case identification (specificity>.99), or by limiting diagnoses to those associated with face-to-face encounters with physicians (specificity>.99). Extending the identification period from 1 to 2 years greatly increases the sensitivity of a case definition at a low cost to specificity. A case definition that combines all Medicare claims over two years and requires two ambulatory diagnoses is both reasonably sensitive and specific (sensitivity = .76, specificity = .98, kappa = .77). This case definition does not systematically misidentify individuals by race, socioeconomic status, or rural/urban residency, although males in good general health tend to be misclassified as not having diabetes. Prevalence rates for diabetes developed by applying these case definitions to Medicare claims are similar to other published prevalence rates. The prevalence of diabetes identified in rural areas is not clinically significantly different from the prevalence in urban areas using these methods.

We conclude that Medicare claims data can be used to identify a cohort of rural elderly individuals with diabetes. This use of existing data offers significant new opportunities for monitoring the health and medical care of individuals with diabetes in rural areas.

INTRODUCTION

Diabetes mellitus, a prevalent and costly chronic disease, is a major cause of morbidity and premature mortality in the United States. One in ten elderly individuals has a diagnosis of diabetes (Kenny, Aubert, and Geiss, 1993), and the estimated annual medical costs of treating a person with diabetes exceeded $11,000 in 1992 (American Diabetes Association, 1993). Recent research suggests that more intensive diabetes care can reduce diabetes-related morbidity and mortality (Diabetes Control and Complications Trial Research Group, 1993), and that well controlled diabetes is associated with lower long run heath care costs (Gilmer et al., 1997). Consequently, tracking and assessing the medical care received by individuals with diabetes is an important and growing public health concern.

Increasingly, researchers are turning to administrative or claims data bases to address critical issue of access, cost and quality of health care provided to individuals with diabetes (U.S. General Accounting Office, 1997). Medicare claims data comprise an especially rich resource due to Medicare's nearly comprehensive coverage of medical services for individuals 65 years or older. Since the claims record both diagnostic codes and procedure codes, they can be used to identify individuals with diabetes and to track the services these individuals receive.

Using Medicare administrative data to track diabetes care is an especially attractive prospect for rural health research. Unlike their urban counterparts, rural health care providers typically do not have patient populations large enough to justify the development of an in-house computerized data base. Collecting data through medical chart abstraction from geographically dispersed rural health providers is prohibitively expensive, if not impossible. Consequently, Medicare claims may be the only feasible data set from which to create a nationally representative cohort of individuals with diabetes in rural areas.

The crucial first step in identifying and tracking services is the development of methods for identifying individuals with diabetes mellitus. Errors in this step may invalidate further analyses, as well as conclusions drawn from those analyses. A method of detecting individuals that is not sufficiently specific would falsely identify some non-diabetic individuals as having diabetes, thereby elevating prevalence estimates and underestimating the percent of persons with diabetes who are receiving the recommended care. Alternatively, a method that is not sufficiently sensitive may fail to recognize certain groups of people with diabetes (e.g., those receiving only occasional services), and would consequently blind researchers to problems of access to or quality of care for these individuals.

While individuals with diabetes can be identified through diagnosis codes on Medicare claims, previous studies have documented imperfect agreement between a diagnosis recorded on a Medicare claim and the diagnosis appearing in an individual's medical record. While the agreement for a diagnoses of diabetes mellitus is higher than rates of agreement for other diagnoses (Quam et al., 1993), errors do occur. A single claim that lists a diagnosis of diabetes may not be a valid indicator of the presence of the disease since a clinician may record a diagnosis of diabetes mellitus as the justification for ordering a fasting blood glucose or HgA_{1C} during the work up to rule out diabetes. A case definition that requires several claims with a diagnosis of diabetes for the same person may be a more specific indicator. However, this increase in specificity may come at the price of reduced sensitivity, since individuals who have infrequent encounters with the medical community may be overlooked. A recent study that addressed this trade-off for a rural Canadian province (Robinson, et al., 1997) concluded that there was little benefit to requiring additional diagnoses for case definition. However, no research has explored this trade-off in the United States Medicare population. Given the

substantial differences in both the health care systems and rural demographics in Canada and the United States, an analysis of the Medicare population is warranted.

Selecting the duration of time over which encounters are recorded also presents inherent trade-offs. Because clinicians significantly underreport chronic diseases on the forms used to generate Medicare claims (Horner et al., 1991), a person with diabetes, cardiac disease or arthritis may have many ambulatory visits in which only new or acute problems are recorded, making it difficult to identify the presence of the chronic problems. Lengthening the time over which reports of the chronic diseases are sought may improve the capture rate. Conversely, extending the time periods may allow repeated evaluations of possible diabetes, resulting in multiple claims which lists diabetes mellitus as a diagnosis. Thus, a trade-off arises between sensitivity and specificity in the selection of the identification period. To date, this trade-off for the identification of a diabetes mellitus cohort among Medicare beneficiaries has not been addressed in the literature.

Given the substantial burden of diabetes in the elderly, and the potential benefits of improved diabetes surveillance in rural areas, it is important to investigate the best strategies for identifying individuals with diabetes through the use of Medicare claims. The purpose of this study was to develop and test various case definitions for identifying individuals with diabetes by combining and weighting Medicare claims of different types. We analyze information on Medicare claims with a diagnosis of diabetes over the period 1992-93 for individuals who participated in the Medicare Current Beneficiary Survey (MCBS). We then use these individuals' self-reported diabetes status from the MCBS to measure the sensitivity and specificity of the various claims-based case definitions for identifying individuals with diabetes. While self-report is not always

accurate, it is the best available standard against which we can compare a national sample of claims data. The study design attempted to address several important issues:

- Does the validity of diagnoses of diabetes differ according to the type of claim (e.g., in-patient institutional claim, physician services claim, and home health services)? Can invalid diagnoses be distinguished based on the source or content of the claim?
- How does the choice of the number of claims with a diagnoses of diabetes affect the sensitivity and specificity of the case definition? Does this number depend on the type of claim on which the diagnosis of diabetes is recorded?
- How many years of data should be collected to most reliably identify a cohort of individuals with diabetes?
- What are the characteristics of the individuals who are misclassified (both as not having or as having diabetes) when we apply various cohort case definitions, e.g., different numbers of encounters with a diagnosis of diabetes, or different durations of the time over which encounters are sought?
- How do estimates of the prevalence of diabetes generated using these claims-based case definitions compare to other published estimates?

DATA AND METHODS

The sources of data for the analysis are the Medicare Current Beneficiary Survey (MCBS) (Adler, 1994) and the accompanying Medicare utilization files. As described below, the MCBS provides survey-based, self-reported information of the diabetes status of participating individuals. The utilization files contain all of the Medicare claims that have been filed for the care of the MCBS respondents. Linking these two data sets provides a means of correlating self-reported diabetes status with the existence of a diagnosis of diabetes on claims submitted to Medicare.

The Medicare Current Beneficiary Survey

The Medicare Current Beneficiary Survey (MCBS) was administered to approximately 13,000 individual yearly from 1992 to 1995. Survey respondents are followed over time, with additional members drawn to replace those lost to follow-up or death. Each year the survey

respondents are asked a series of questions regarding their health status. One of these questions is: "Have you ever been told you have diabetes?" We used the response to this question as the standard against which we measure the validity of the diagnosis fields in the Medicare claims data. Studies have shown that self-reported diabetes is accurate when compared against medical records (Kehoe et al., 1994) (Quam et al., 1993) (Kriegsman et al., 1996).

The MCBS Cost and Use files for 1992 and 1993 were used to identify individuals eligible for the analysis. Demographic information (age, race, gender, and rural/urban residency) were extracted from the analytic files. A beneficiary is defined as a rural resident in the MCBS if he/she resides in a county that is not in a Metropolitan Statistical Area (Office of Management and Budget, 1993). Diabetes status was determined using the Health Survey files. Only individuals 65 of age as of 1/1/92 who participated in both the 1992 and 1993 survey (n = 8,052) were selected. Of these, we eliminated individuals who:

- Lived outside the 50 US states in either 1992 or 1993 (n = 108)
- Were not continuously eligible for Medicare Part A and Medicare Part B during both years (n = 265)
- Were ever in a managed care organization in 1992 or 1993 (n = 600)
- Did not respond identically ("yes" or "no") to the diabetes question in both 1992 and 1993 (n = 121)

Individuals who were ever in a managed care organization were eliminated because managed care organizations are not required to submit encounter-specific claims to Medicare. The last selection requirement listed above has the possible effect of eliminating incident cases of diabetes in 1993 (1.5% of the otherwise eligible sample) as well as a relatively small number of individuals who responded that they did have diabetes in 1992, but did not in 1993 (13 individuals; 0.2%).

The net effect of these requirements is that of the 8,052 individuals over the age of 65 who participated in both the 1992 or 1993 survey, we eliminated 1,094, to obtain a working sample of 6,958. Table 1 shows summary statistics for the MCBS data used in this analysis.

Medicare Claims Data

The Medicare claims data for the years 1992-93 come from six utilization files: Short stay hospital (SS), skilled nursing facility (SNF), hospital outpatient (OP), Part B physician/ supplier (PB), durable medical equipment (DME),[1] and home health (HHA).

The first step was to identify all claims that had a diagnosis of diabetes in any diagnosis position. Medicare claims indicate between four (Part B physician/supplier and durable medical equipment) and ten (short-stay hospital, skilled nursing facility, home health agency, and hospital outpatient) diagnoses associated with each claim. The first listed diagnosis is called the principal diagnosis. We identified the following ICD-9-CM codes (Parente et al., 1995) as diabetes-related diagnoses:

Diabetes and complications	250.0 - 250.9
Diabetic retinopothy	362.0-362.1
Polyneuropathy in Diabetes	357.2
Diabetic Cataract	366.41

The second step was to examine other information on the claim that may be useful in distinguishing a more valid diagnosis of diabetes from a less valid diagnosis. For short-stay hospital, skilled nursing facility claims, home health agency, and hospital outpatient claims, we used the position in which the diagnosis appeared on the record as a potential indicator of the quality of the diagnosis. Diagnoses of diabetes were coded as principal diagnoses if they

[1] Durable medical equipment is covered under Medicare Part B; this database generated a very small number of claims with a diagnosis of diabetes. Therefore, the information from the durable medical equipment database is included in the figures for the other Part B database, the physician/supplier database.

Table 1

Mean Age and Distribution of Other Demographic Characteristics of MCBS Participants Used to Validate Medicare Claims Data (Standard Errors in Parentheses)

Characteristics	Number	Percent
Age		
65 - 74	2,947	42.4
75 - 85	2,767	39.8
85+	1,244	17.9
Mean	76.9 (7.66)	
Caucasian	5,699	81.9
Female	4,342	62.4
Rural resident	2,046	29.4
Self-reported diabetes	1,099	15.8
Total Respondents	6,958	

Because the MCBS uses a stratified random sampling design, inverse probability weights were used to estimate the statistics presented below.

appeared in the first diagnosis field on the claim, or is secondary diagnoses if they appeared in any other position.

In addition to the diagnoses that may be listed for each visit, Part B claims contain additional information that may be useful in distinguishing valid from invalid diagnoses of diabetes. Part B claims have two sections in which a diagnosis of diabetes may be found. The first section, known as the "header," contains four fields in which a physician or other provider can enter an ICD-9 diagnosis code that is relevant to the patient. The second section, called the "trailer" or line-item, contains HCFA Common Procedure Codes (HCPC) and Unique Physician Identification Numbers (UPINs) as well as the ICD-9 diagnosis code at which the treatment was directed. Any diagnosis that appears in the trailer must also be in the header. However, diagnoses in the header portion are not required to appear in the trailer.

Using this information, we divided Part B claims with a diagnosis of diabetes into five categories. The first category contains claims with a diagnosis of diabetes in any position on the claim, and is thus the least restrictive category. The second category contains claims in which a diagnosis of diabetes in the trailer section is associated with a HCPC code that represents a face-to-face procedure. The HCPC codes on a claim are based on Current Procedural Terminology (CPT) (American Medical Association, 1994) codes and thus refer to a wide array of procedures. Some of these procedures require direct physician contact while others do not. Previous research (Quam et al., 1993) has suggested that diagnoses associated with a procedure that requires direct, face-to-face contact between patient and physician (for example, a dilated eye exam) may be more valid than diagnoses associated with procedures not requiring direct contact (e.g., HgA1c lab test).

To test this hypothesis, every procedure that was associated with a diagnosis of diabetes was coded as face-to-face or not, based on an algorithm developed at the University of Washington (Sugarman, 1997). However, not all of these claims with face-to-face diagnoses of diabetes identified a physician as the individual performing the procedure and therefore responsible for the diagnosis. The third and fourth categories of Part B claims were created to investigate whether a claim that clearly identified a physician as the party making the diagnosis was a more valid indicator of diabetes than a claim in which the training (MD, PA, RNP) of the individual documenting the diagnosis was ambiguous. The third category contains Part B claims with face-to-face diagnoses of diabetes in which the person performing the procedure is identified as a physician either by a UPIN or an appropriate specialty code. The fourth category uses a more restrictive definition of a physician contact. It includes only those face-to-face diagnoses of diabetes in which the physician was identified by a UPIN.

The fifth Part B claim category contains claims in which the diagnosis of diabetes appears in the header but not in the trailer or line-item portion of the records. In these records, the provider acknowledged the presence of diabetes but did not administer any procedures that were specific to diabetes. It was believed that this type of record would be a more valid indicator of diabetes because the physician associated the disease with that patient so strongly that he or she recorded the diagnosis even though doing so was not required for reimbursement purposes. Table 2 summarizes the criteria we developed for each type of claim.

Developing Aggregate Criteria

For each person in the data base we counted the number of claims over the 1992-3 period in each category listed in Table 2. We then developed six case definitions which identified a

Table 2

Information Available on Medicare Claims that May Distinguish Valid from Invalid Diagnoses of Diabetes (by type of Medicare claim)

Short Stay Hospital, Skilled Nursing Facility, Home Health Agency, or Hospital Outpatient

a. Principal diagnosis of diabetes
b. Any diagnosis of diabetes (i.e., a diagnosis of diabetes which occurs in any of the ten diagnosis fields).

Part B Physician/Supplier and Durable Medical Equipment

a. Any diagnosis: Any diagnosis of diabetes (either in the header portion of the record, or the procedure specific trailer portion of the record).
b. Face-to-face: A diagnosis of diabetes that is associated with a procedure code that represents a face-to-face encounter with physician (using the University of Washington definition of face-to-face).
c. Face-to-face + (UPIN or appropriate physician specialty code): A diagnosis of diabetes that is associated with 1) a procedure code that represents a face-to-face encounter with a physician, and 2) either a physician identifying number (UPIN) or a specialty code that rules out non-physician suppliers (e.g., labs).
d. Face-to-face + UPIN: A diagnosis of diabetes that is associated with 1) a procedure code that represents a face-to-face encounter with a physician, and 2) a unique physician identification number.
e. Header only: a diagnosis of diabetes which appears in the header portion of the record, but not the trailer.

person as having diabetes according to the number and mix of claims a person had in each of these categories. These case definitions are shown in Table 3.

everal principles guided the construction of the various case definitions:

- Case Definition 1 was developed by first comparing the specificity and sensitivity of individual claim types shown in Table 6. The category combined the claim types with the highest specificity (short-stay hospital, skilled nursing facility, and home health agency), with the sequential addition of claim types with lower specificity but higher sensitivity. The goal was to increase the sensitivity at a minimal cost to specificity.

- Case Definitions 2 and 3 were developed to assess the consequences of limiting the data and analytic requirements to select cases. Case Definition 2 deletes hospital outpatient records from the data sources of Case Definition 1. Case Definition 3 eliminates hospital outpatient and home health claims. If the hospital outpatient or home health file contains little information that is not already present in the physician supplier file, then researchers could forego the considerable time and effort of compiling Medicare hospital outpatient or home health records.

- Similarly, Case Definition 4 was developed to explore whether the computationally expensive task of recoding a HCPC procedure code as face-to-face or not contributed significantly to the sensitivity and specificity of the criterion. The major difference between Case Definition 4 and 1 is the coding of the HCPC procedure as face-to-face.

- Case Definition 5 was used by the General Accounting Office in its study of the services received by elderly Medicare beneficiaries (U.S. General Accounting Office, 1997). The Health Care Financing Administration's Region X Office of Peer Review has used a similar criterion to identify Medicare beneficiaries with diabetes in its effort to monitor quality of care.

- Case Definition 6 is both the most sensitive measure possible and the easiest to determine.

Because sensitivity and specificity estimates and the computational work required to derive them vary according to the amount of time over which data are collected, we tested each of the six case definitions above using both a single year and two years of claims data.

Analysis

Three types of analysis were conducted. The first analysis simply categorized individuals as "has diabetes" or "does not have diabetes" based on whether the claims data for the individual met the case definition under investigation. This estimated indicator of diabetes was then

Table 3

Case Definitions for Diabetes
Developed from Medicare Claims Data

Case Definition 1: One hospitalization, one skilled nursing stay, or home health claim in any position, or 2 outpatient hospital, or 2 Part B claims
Case Definition 2: Case Definition 1 without outpatient hospital
Case Definition 3: Case Definition 1 without outpatient hospital or home health
Case Definition 4: One hospitalization, one skilled nursing stay, or one home health claim or one Part B record that is either only in the header or a face-to-face diagnoses in the trailer (using University of Washington definition if face-to-face)
Case Definition 5: One hospitalization or two Part B records that are face-to-face diagnoses (University of Washington definition of face-to-face)
Case Definition 6: One diabetes diagnosis on any claim

compared with the individual's self-reported diabetes status. Each individual was classified as true positive, true negative, false positive or false negative based on the agreement with self-reported diabetes status. These data were then used to calculate sensitivity, specificity and kappa measures for each criterion.

The second analysis used a logit regression to model the likelihood that an individual has diabetes based on the number and type of claims with a diagnosis of diabetes. The dependent variable in this analysis was self-reported diabetes status. Independent variables include demographic information (age, race, and gender) and the number and percentage of claims which met various criteria. The result of this analysis is an estimate of the probability (P_i^* = (0 to 1)) that individual i has diabetes, given the number and types of claims he/she has with a diagnosis of diabetes. All else being equal, the greater the number of claims an individual has with a diagnosis of diabetes, the greater the value of P_i^* and the more likely the individual in fact has diabetes. Individuals are predicted to have diabetes if the P_i^* for that individual exceeds a specified cut-off value. The cut-off value can be set according to the needs of the researcher. For this analysis, we set the cut-off to achieve a specificity of .975. That is, we chose a P* cut-off so that 97.5% of the individuals who did not have diabetes according to self-report also had a P_i^* less than P*.

In the third analysis, we investigated characteristics of self-designated persons with diabetes whose claims data identified them as not having diabetes. A logistic regression was performed on the subset of the data base containing individuals who responded affirmatively to the diabetes status question on the MCBS. These individuals were coded 1 (true positive) or 0 (false negative) if they did or did not meet a particular criterion listed in Table 3 above. Separate logistic regressions were performed to test Case Definitions 1 and 5, respectively. These case

definitions were selected because Case Definition 1 has the highest percent agreement with self-report as measured by the kappa statistic, and Case Definition 5 has been used in other diabetes research (U.S. General Accounting Office, 1997). Independent variables for the logistic regression include:

demographic information (age, race, gender),

information on health status (e.g., self-reported general health status and the existence of several important diabetes-related complications),

and measures of access to care, including rural/urban residence, income, and insurance status (Medicaid and supplemental Medicare insurance).

The maximum likelihood HLOGIT procedure in Stata 4.0 (Stata Corporation, 1995) was used. The *pweights* option was used and set equal to the inverse probability weights provided in the MCBS. The MCBS uses a stratified sampling strategy in which certain demographic groups (e.g., the disabled and the very old) are systematically over-sampled. As a result, the demographic composition of MCBS respondents is dissimilar to the general Medicare population. To account for this, MCBS provides inverse probability weights that are the inverse of the probability that an individual was selected from his primary sampling unit. Using inverse probability weights assigns a lower relative weight or importance to data for individuals who were over-sampled in the MCBS, and thus allows one to generalize the analysis to the general Medicare population.

Estimation of the Prevalence of Diabetes in Rural Areas

Case Definition 1 was used to define a cohort of rural and urban individuals with diabetes. Medicare claims data were drawn from a five percent sample of Medicare beneficiaries who were over the age of 65 as of 1/1/92, resided in the 50 U.S. states in 1992 and 1993, were

not in an HMO in 1992 or 1993, were continuously eligible for both Medicare Part A and Part B coverage through 1993, and were alive on 1/1/94.

RESULTS

Validity of Diagnoses of Diabetes by Type of Claim

A diagnosis of diabetes on any of the types of claims was highly specific. The sensitivity and kappa statistics varied significantly by the type of claim considered. Table 4 summarizes the analysis of the sensitivity and specificity of individual claims data submitted for the year 1992 alone, as well as claims submitted over the two-year period 1992-3. The sensitivity, specificity and kappa statistics in the tables have been adjusted to reflect the non-random sampling strategy of the MCBS.

Considering only hospitalizations, 414 persons in the sample had at least one hospital claim that contains a diagnosis of diabetes in one of the 10 diagnosis fields over the 1992-3 period. Of these individuals, 375 responded that they did have diabetes on the MCBS (true positive). There were 5,821 individuals who did not have a hospitalization claim with a diagnosis of diabetes, and responded that they did not have diabetes on the MCBS (true negative). There were also 723 individuals who responded that they did have diabetes, but had no hospitalization with a diagnosis of diabetes. After adjusting for sampling strategy used in the MCBS, these figures yield a sensitivity estimate of .336, a specificity of .993, and a kappa of .441.

The specificity of principal diagnosis is essentially equal to 1.0 for short stay hospital, skilled nursing facilities, home health agency and outpatient hospital records. The sensitivity of these measures, however, are very low: from <0.1 to .336. Part B records are the most sensitive and least specific. A single Part B claim in one year with a diagnosis of diabetes has a specificity

Table 4

Sensitivity and Specificity by Type of Claim Over a One Year (1992) and Two Year (1992-1993) Identification Period (Adjusted Self-Reported Prevalence 15.8%; n=6,958)

Type of Claim with Diagnosis of Diabetes	One Year Identification Period				Two Year Identification Period			
	Number of Individuals	Sensitivity	Specificity	Kappa	Number of Individuals	Sensitivity	Specificity	Kappa
One Diagnosis								
Short Stay Hospital								
Principal diagnosis of diabetes	25	0.027	1.000	0.044	48	0.048	1.000	0.077
Any diagnosis of diabetes	222	0.188	0.997	0.273	414	0.336	0.993	0.441
Skilled Nursing Facility								
Principal diagnosis of diabetes	4	0.002	1.000	0.003	14	0.009	1.000	0.014
Any diagnosis of diabetes	19	0.013	1.000	0.022	18	0.032	0.999	0.051
Home Health Agency								
Principal diagnosis of diabetes	70	0.059	1.000	0.095	109	0.086	0.997	0.232
Any diagnosis of diabetes	130	0.105	0.999	0.161	203	0.158	0.999	0.135
Outpatient Hospital								
Principal Diagnosis of diabetes	195	0.157	0.996	0.229	282	0.227	0.993	0.314
Any diagnosis of diabetes	275	0.220	0.994	0.306	423	0.329	0.989	0.422
Part B: Physician/Supplier								
Line-item diagnosis of diabetes	882	0.667	0.974	0.697	1,070	0.041	0.956	0.705
"Header-only" diagnosis of diabetes	437	0.334	0.990	0.441	625	0.482	0.983	0.562
Face-to-face diagnosis (UW)	731	0.601	0.988	0.682	902	0.696	0.977	0.726
Two Diagnoses								
Part B. Physician/Supplier								
Line-item diagnosis of diabetes	655	0.562	0.992	0.661	842	0.686	0.983	0.737
"Header-only" diagnosis of diabetes	219	0.181	0.997	0.264	367	0.309	0.994	0.412
Face-to-face diagnosis (UW)	534	0.466	0.995	0.581	708	0.602	0.991	0.691
Outpatient Hospital								
Principal diagnosis of diabetes	95	0.084	1.000	0.132	163	0.147	0.999	0.221
Any diagnosis of diabetes	131	0.115	0.999	0.178	234	0.207	0.998	0.300

of .974 and a sensitivity of .667. Adding the face-to-face requirement does increase the specificity of a single Part B claim, from .974 to .988, but also reduces sensitivity from .667 to .601. Contrary to expectations, adding the UPIN and/or physician specialty requirement does not substantially change either the specificity or sensitivity of Part B claims. As a result, these items do not appear in Table 4. The "header-only" diagnosis of diabetes, however, is highly specific (specificity = .99 over 1 year), as hypothesized.

Increasing the number of claims required to define a case also increases specificity. Since one institutional or home health claim is already highly specific, we focus on ambulatory claims. Requiring two Part B claims with a line-item diagnosis of diabetes increases the specificity from .974 to .992, although at a substantial cost to sensitivity which drops from .667 to .562. The cost to sensitivity is even more pronounced when the face-to-face restrictions are imposed. Requiring two Part B claims with a University of Washington-defined face-to-face diagnosis of diabetes yields a sensitivity of .466. That is, over half of the people who report that they have diabetes will not be identified by this case definition using claims from one year.

Increasing the duration of the identification period also has important consequences for the case definition. Using two years of data greatly increases the sensitivity of hospitalization claims with a diagnosis of diabetes. The increase in sensitivity for ambulatory services such as physician services and outpatient hospital is less pronounced but still significant. Interestingly, of the case definitions based on Part B claims alone, two Part B claims over two years with no other restrictions has the highest percent agreement as measured by the kappa statistic (.732).

Case Definitions Developed by Combining Claims of Various Types

The results of the analysis of case definitions defined by combining claims of various types are shown in Table 5. Case definitions that combined data from different types of claims

Table 5

Sensitivity and Specificity of Case Definitions
Combination of Files Over One Year and Two Years

Case Definition	One Year Identification Period			Two Year Identification Period		
	Sensitivity	Specificity	Kappa	Sensitivity	Specificity	Kappa
1. Case Definition 1: One hospitalization, skilled nursing, or home health claim, or 2 outpatient hospital, or 2 Part B claims	.655	.987	.724	.760	.975	.768
2. Case Definition 2: Case Definition 1 without outpatient hospital	.648	.988	.719	.752	.975	.765
3. Case Definition 3: Case Definition 1 without outpatient hospital or home health	.625	.989	.712	.748	.976	.765
4. Case Definition 4: One hospitalization, skilled nursing, or home health claim or one Part B record that is either header only for face-to-face diagnoses	.712	.987	.741	.786	.961	.751
5. Case Definition 5: One hospitalization or two Part B records that are face-to-face diagnoses (UW definition)	.530	.993	.634	.662	.985	.723
6. Case Definition 6: One diabetes diagnosis on any claim	.752	.962	.730	.808	.938	.709

had less variability in sensitivity, specificity and kappa statistics than did case definitions based on a single type of claim. Comparisons of the one-year and two-year identification periods in Table 5 demonstrate again the importance of defining a diabetes cohort over two years to attain the greatest sensitivity. The combination of records that achieves the highest kappa statistic is Case Definition 1. Case Definitions 2 and 3 show that very little is lost in either sensitivity or specificity when outpatient and home health files are neglected. Most of the individuals identified in these files have Part B or hospitalization records. The additional work of searching home health and hospital outpatient files for diagnoses of diabetes provides limited value.

The analysis also suggests that coding a procedure as face-to-face can increase sensitivity. By decreasing the number of required Part B records from two to one and imposing the University of Washington face-to-face criterion, one can increase the sensitivity of the Case Definition 1 from .760 to .786, although the specificity drops from .975 to .961.

Case Definition 5 has high specificity, but low sensitivity. Over 33 percent of the individuals who say that they have diabetes are not identified as such when Case Definition 5 is employed. The case definition with the highest sensitivity is the criterion that accepts a single diagnosis of diabetes on any claim over a two-year period (Case Definition 6). The sensitivity is .808, although the specificity is a rather low .938. This suggests that even when the most liberal criterion for diabetes is employed over a two-year period, 20% of the individuals who said they have diabetes are not identified.

Logistic Regression Analysis

An increase in the number of diabetes-related claims is positively associated with self-reported diabetes (Table 6). To account for the fact that a single claim with a diagnosis of diabetes may be significant if the individual has infrequent physician encounters and thus

Table 6

Identifying Individuals with Diabetes Using Logistic Regression

	Odds Ratio	Coefficient	p
Male	1.4	0.371329	<0.001
African American	2.0	0.682142	<0.001
Age	1.0	-0.021960	0.003
SS claims	2.3	0.848853	<0.001
SNF claims	3.5	1.254644	0.015
HH claims	1.2	0.172680	0.007
Part B: header only	1.8	0.561719	0.001
Part B: any diagnosis	1.7	0.507101	<0.001
Part B: face-to-face	1.3	0.285184	0.025
% of Part B claims with a diagnosis of diabetes	2.4	0.865139	0.009
Outpatient claims	3.1	1.134313	<0.001
Constant	0.1	-2.265530	<0.001
Sensitivity .748 Specificity .975 Kappa .759			

generates few Part B claims, we included a variable that represent the percent of all Part B claims that have a diagnosis of diabetes. As expected, this variable was significantly correlated with self-reported diabetes.

The logistic regression analysis was used to account for the effect that differences in health care utilization between various demographic groups may have on the sensitivity of the case definitions. Men and members of minority groups tend to have fewer physician encounters and thus tend to generate fewer Medicare claims. A case definition that requires two or more ambulatory claims may fail to identify these infrequent users. The logit analysis partially corrects for this by estimating a probability of diabetes (P_i*) that is specific to the demographic characteristics of each individual. Since the coefficients on the variables "male" and "African American" in Table 6 are significant and greater than 0, an African American male with a single Part B claim with a diagnosis of diabetes yields a higher estimated probability of diabetes than does a Caucasian female with a single Part B claim. Using the methodology described in section 2.4, we used the results of the logit analysis to specify a cut-off for the estimated probability of diabetes that would yield a specificity of .975. An individual with a Pi* higher than the cut-off is defined as a positive case identification. For a non-Caucasian male aged 65 years, a single diabetes-related claim of any kind yields a Pi* above the cut-off, thus reducing the bias against infrequent health care users.

This methodology improves case identification only if individuals in traditionally low utilization demographic groups are generating some claims with a diagnosis of diabetes but too few claims to meet a case definition based on counts. This does not appear to be the case for many individuals in the MCBS. The sensitivity estimate for Case Definition #6 (Table 5) suggests that nearly 20 percent of the individuals with self-reported diabetes generated no claims

with a diagnosis of diabetes, and thus would not be identified by this or any other claims-based methodology. Moreover, the logit methodology would tend to under-identify low utilization individuals who are members of high utilization demographic groups. As a result, when this methodology is applied to all individuals in the sample, the resulting sensitivity is .748 and kappa is .759, which is slightly worse than Case Definition, as 1 shown in Table 5.

Analysis of the Mis-classified Individuals

Table 7 displays the results of the analysis of those self-reported diabetic individuals that we failed to identify as such by our various criteria. An odds ratio greater than 1.0 means that a factor is positively associated with the likelihood that individual with diabetes would be falsely identified as not having diabetes. Only two factors available from the MCBS C gender and general health C were significantly related to being mis-classified as not having diabetes. Males and individuals who respond that they are in good general health are under-identified by these case definitions. No other characteristics of the patient were significantly correlated with mis-classification across all case definitions analyzed. In particular, there was no significant difference in mis-classification between rural and non-rural residents, where rural is defined as non-MSA.

To correct for possible autocorrelation distorting the standard errors, groups of related explanatory variables were also tested for joint significance, but none were found to be significant. For example, the explanatory variables for coronary artery disease, high blood pressure, and myocardial infarction were not jointly or individually significant.

Table 8 shows summary statistics for individuals mis-classified by Case Definition 1. The figures in Table 8 clearly show that individuals who did not meet the case definition (one hospitalization, skilled nursing, home health, or two ambulatory claims with a diagnosis of

Table 7

Logistic Regression Analysis of False-Negative Classifications By Selected Classification Criteria

	Case Definition 1		Case Definition 5	
	Odds Ratio	**p**	**Odds Ratio**	**p**
Male	**1.40**	**0.042**	**1.51**	**0.007**
African American	1.23	0.379	1.38	0.127
Age	1.01	0.476	1.01	0.579
General health	**0.83**	**0.012**	**0.79**	**0.000**
Coronary artery disease	1.08	0.698	1.04	0.813
High blood pressure	1.02	0.914	1.22	0.206
Myocardial infarction	1.00	0.992	1.11	0.571
Congestive heart	0.71	0.092	0.79	0.214
Other heart disease	1.13	0.450	0.97	0.860
Stroke	0.93	0.729	0.73	0.086
Amputation	0.80	0.669	0.91	0.848
Income	0.94	0.695	1.00	0.990
Rural	1.04	0.329	1.14	0.366
Medicaid eligible	0.71	0.175	0.75	0.188
Private supplemental	0.73	0.123	0.85	0.393
Constant	0.32	0.241	0.62	0.582

Table 8

Number of Persons and Average Number and Type Claims over Two Years for People in the Study Organized According to Self-Reported Diabetes Status: Results of Applying Case Definition 1 to Medicare Claims for These Individuals

	Affirmative Self-Report of Diabetes		Negative Self-Report of Diabetes	
	True Positive	**False Positive**	**False Positive**	**True Positive**
Number of persons	n=826	n=272	n=157	n=5,703
Mean number of claims over two years				
Short Stay Hospital	1.2	0.5	1.2	0.6
Skilled Nursing home	0.2	0.1	0.4	0.1
Home Health	2.0	0.8	1.7	0.7
Outpatient Hospital	7.2	0.4	7.3	3.9
Mean number of diagnoses appearing on Part B claims over a two-year period	105.0	53.5	96.0	60.5

diabetes) still had many encounters that could have generated a claim with a diagnosis of diabetes, although far fewer than individuals who did meet the case definition.

Using Medicare Claims for Diabetes Surveillance in Rural Areas

The estimated prevalence rate is consistently lower in rural areas than in urban areas for Caucasian individuals of all ages (Table 9). Because of the large sample size, the urban-rural differences are statistically significant, although the magnitude of the difference is not large. The prevalence rate drops in the population over 80 years of age in both rural and urban areas. Estimated prevalence rates in the African American community are generally higher in rural areas than in urban areas, although the differences are not great (6.3 percent all age groups).

DISCUSSION

Medicare claims can be used to identify a cohort of rural elderly individuals with diabetes. By combining claims from several Medicare data bases researchers can construct a case definition for diabetes that is both adequately sensitive and specific. For many areas of rural health research, the analysis of Medicare records can take the place of a more expensive and time consuming review of patients' medical records, and perhaps lessen the pressure to acquire expensive electronic data bases in individual practices in rural areas. The types of information available on Medicare claims also are essentially uniform across states. This greatly facilitates the creation of a combined cohort drawn from rural areas throughout the country, enabling rural health researchers to analyze the care provided in a wide range of rural settings.

The creation of a diabetes cohort is made possible by the high specificity of diabetes-related diagnoses on Medicare claims. A diabetes-related diagnosis on a Medicare claim is a specific and predictive indicator of the presence of the disease. As anticipated, short stay hospital, skilled nursing, and home health agency claims are very highly specific. Physicians

Table 9

Claims-based Estimates of the Prevalence of Diabetes Mellitus in Rural and Urban Areas in 1994 Based on a 5% Sample of Medicare Beneficiaries: Case Definition 1 (By Race and Age)

	Urban			Rural			
	Estimated Cases of Diabetes	Total Population	Estimated Prevalence Rate/100	Estimated Cases of Diabetes	Total Population	Estimated Prevalence Rate/100	Percentage Difference From Urban
All Races							
Age 67+	118,255	888,353	13.3	40,612	311,784	13.0	-2.3
67-69	21,369	176,258	12.1	7,164	61,373	11.7	-3.3
70-74	36,290	268,928	13.5	12,303	92,345	13.3	-1.5
75-79	28,757	203,647	14.1	9,963	72,038	13.8	-2.1
80-84	18,510	136,648	13.5	6,704	49,687	13.5	0.0
85+	13,329	102,872	13.0	4,479	36,341	12.3	-5.4
Caucasian							
Age 67+	99,464	789,300	12.6	35,661	286,819	12.4	-1.6
67-69	17,347	154,828	11.2	6,193	56,674	10.9	-2.7
70-74	30,255	238,452	12.7	10,778	85,375	12.6	-0.8
75-79	24,564	182,817	13.4	8,821	66,695	13.2	-1.5
80-84	15,818	121,395	13.0	5,933	45,144	13.1	0.8
85+	11,480	91,808	12.5	3,936	32,931	12.0	-4.0
African American							
Age 67+	14,811	71,973	20.6	4,137	18,869	21.9	6.3
67-69	3,150	15,595	20.2	807	3,600	22.4	10.9
70-74	4,745	22,563	21.0	1,271	5,474	23.2	10.5
75-79	3,394	15,712	21.6	983	4,229	23.2	7.4
80-84	2,087	10,479	19.9	610	3,109	19.6	-1.5
85+	1,435	7,624	18.8	466	2,457	19.0	1.1
Other Races (including unknown)							
Age 67+	3,980	27,080	14.7	814	6,096	13.4	-8.8
67-69	872	5,835	14.9	164	1,099	14.9	0.0
70-74	1,290	7,913	16.3	254	1,496	17.0	4.3
75-79	799	5,118	15.6	158	1,114	14.2	-9.0
80-84	605	4,774	12.7	161	1,434	11.2	-11.8
85+	414	3,440	12.0	77	953	8.1	-32.5

and other providers in these settings have ample time and interaction with the patient, and are thus more likely to make and record a diagnosis of diabetes. Quality assurance and discharge planning coordinators in hospital settings provide another check on the validity of diagnoses in these settings. Finally, because a complication or comorbidity of diabetes elevates a patient to a more lucrative Diagnostic Related Group, hospitals have a financial incentive to code a diagnosis of diabetes when it is present.

While highly specific, such claims occur too infrequently to be used to define a cohort of individuals with diabetes. In order to create a case definition for diabetes that is both specific and sensitive, researchers must analyze claims from ambulatory care settings such as physician claims and claims for laboratory services. While these records are fairly specific, our analysis suggests that a single diagnosis is not a reliable indicator of the presence of the disease. A single diagnosis of diabetes on a Medicare Part B claim (a claim for ambulatory services) in this study had a specificity over 0.95. This finding is consistent with other studies that have found a relatively high degree of agreement between diagnoses of diabetes on physician office claims and medical charts (Fowles et al., 1995) and between physician-reported and patient self-reported (Kehoe et al., 1994) diabetes. However, if a single ambulatory claim with diabetes-related diagnosis were used as a case definition, the resulting cohort would likely include many individuals who do not in fact have diabetes. Of the 5,860 people in our sample who said they did not have diabetes, 266 (4.5%) had one or more ambulatory claim with a diabetes-related diagnosis, and no similar diagnoses on any other type of claim.

Our research suggests two proposed methods to increase the specificity of a case definition based on ambulatory claims. One method is to restrict ambulatory claims to those that contain face-to-face diagnoses of diabetes. This eliminates diagnoses of diabetes that are

associated with laboratory tests where a diagnosis of diabetes may have been intended as a "rule out" diagnosis; it also eliminates and other procedures where the physician-patient interaction is distant or non-existent.

Our analysis shows that the face-to-face restriction does increase the specificity of the case definition, but there are still quite a few individuals in this study who claimed that they did not have diabetes and yet generated a face-to-face diagnosis of diabetes (n = 146; 2.5% of 5,860 negative self-reports). On average, a cohort of 1,000 persons with diabetes as defined by a single face-to-face diabetes diagnosis would contain 84 persons who report that they do not have diabetes (positive predicitive value = .906). This would complicate the interpretation of any analysis of the quality of care provided to individuals with diabetes. Moreover, 18 individuals in this study reported that they had diabetes, but only generated diabetes-related diagnoses on Part B claims that were not face-to-face. A healthy individual with diabetes may have encounters that only result in laboratory tests to monitor the progression of diabetes. These people are not identified by a case definition based on the face-to-face restriction.

The second method to improve the specificity of ambulatory records is to require two or more diagnoses of diabetes for a positive case definition. The analysis suggests that a case definition that requires two diabetes-related diagnoses is more specific than a case definition which requires a single face-to-face claim. Of the individuals who claimed they did not have diabetes (n=5,860), only 100 (1.7%) individuals generated two or more Part B claims with a diabetes-related diagnosis, as compared to 266 (4.5%) who generated one or more such claims, and 146 (2.5%) who generated one or more face-to-face diabetes diagnosis. Since requiring two ambulatory claims is easier to implement than is coding each encounter as face-to-face or not

face-to-face, the two-claim restriction may be a more attractive solution to maximizing specificity, particularly for those researchers with limited computer resources.

However, maximizing specificity can compromise the sensitivity of the case definition, especially for groups of healthy persons who have infrequent encounters with the medical community. The problem could be ameliorated by extending the identification period to more than one year. This provides individuals more opportunities for a physician encounter, and therefore more opportunities to generate a claim with a diagnosis of diabetes. While this may increase the chances of generating a false claim of diabetes, in our analysis extending the identification period from one to two years greatly increased the sensitivity of a case definition, and decreased specificity only slightly. Applying Case Definition 1 (one hospitalization, one home health, or one skilled nursing claim or 2 ambulatory claims) to a one-year identification period yields a diabetes cohort that contains only 65.5 percent of the individuals who claim they have diabetes. Adding a second year to the identification period increases this percentage to 76.0 percent, yet specificity decreases only slightly, from .987 over one year, to .975 over two years.

Extending the identification period to multiple years could complicate the analysis by increasing the data requirements. For example, a research project which seeks to analyze the care provided to individuals with diabetes in a given year requires a total of three years of data: two years for the identification period and one for the analysis. In addition, this requirement limits any analysis to those individuals who survive at least two years. Similarly, a study which seeks to analyze individuals in a particular health plan or geographic area will be limited to analyzing those individuals who did not disenroll or did not relocate over a three-year study period. While these restrictions may be a concern for some areas of research, they are not likely to adversely impact studies of the rural elderly since the relocation of rural elderly is rare and, at

present there is very little penetration of managed care or other alternatives to traditional fee for service Medicare in rural areas.

By combining Part B, inpatient, and other Medicare claims over two years we can develop a case definition that is both reasonably sensitive and specific. Case Definition 1 has the highest relative percentage agreement as measured by the kappa statistic. This definition relies in part on home health agency and hospital outpatient (i.e., claims submitted by the outpatient clinic, not the physician). Both these Medicare data files are large, and the hospital outpatient file is especially difficult to use. Our analysis suggests researchers can create a valid case definition and a reliable cohort of individuals with diabetes without the considerable effort involved in compiling these records. A case definition based on hospitalization, skilled nursing and physician claims is nearly as specific and sensitive as Case Definition 1.

Case Definition 1, however, may not be appropriate for all research. The selection of an appropriate case definition should be governed by the goals of the research. For example, Case Definition 5, which requires two face-to-face diagnosis diagnoses of diabetes for positive case identification, creates fewer false positives than Case Definition 1 and may be the more desirable case definition to use in research where specificity is a priority.

Our analysis also suggests that case definitions developed from Medicare claims do not systematically mis-classify individuals by race or socioeconomic status. This again is good news for rural health researchers. The case definitions applied equally well to individuals living in urban areas (i.e., counties in an MSA) as to individuals living in rural (non-MSA) counties. However, men and individuals of good general health tend to be mis-classified as not having diabetes. These "false negative" individuals tended to have few encounters with the medical community, and consequently generated fewer claims that could contain a diagnosis of diabetes.

Further research is required on the implications of these biases for epidemiological and health services research.

One potential limitation of this study is that it relies on patient self-reported diabetes status as the gold standard against which we judge the validity of Medicare claims records. However, evidence from previous studies suggest that self-reported presence of chronic diseases is a fairly specific indicator of the disease. Self-reported diabetes has been validated against medical records (Kehoe et al., 1994) (Quam et al., 1993) (Kriegsman et al., 1996) and more recently against administrative claims data bases (Robinson et al., 1997). In addition, the self-report of diabetes in the MCBS has some external validity in that it matches published rates of incidence and prevalence. In the MCBS, 15.8 percent of the respondents responded that they had been told they have diabetes. This overall prevalence rate and the race specific prevelence rates (15 percent Caucasian, 24 percent non-Caucasian) are similar to, although slightly higher than other estimates (NHANES III)[2]. In addition, only 13 individuals (0.2%) in the MCBS sample responded that they did have diabetes in 1992, but not in 1993. This high reliability also suggests that self-reported diabetes is an adequate standard against which we can judge the validity of claims data.

CONCLUSION

Relatively accurate case definitions can be developed from the information on Medicare claims to identify a cohort of patients with diabetes. This cohort can be used to monitor the quality and cost of care provided to individuals with diabetes. The ability to use Medicare claims data to identify and monitor the care of individuals with chronic diease is likely to be

[2] These unpublished estimates were provided by Maureen I. Harris, Ph.D., MPH, of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health.

particularly important in rural areas where few large health plan or managed care data bases are available for assessment of rural citizens' health access or quality of care.

REFERENCES

Adler, G. "A Profile of the Medicare Current Beneficiary Survey." *Health Care Financing Review* 15:153-163, 1994.

American Diabetes Association. *Direct and Indirect Costs of Diabetes in the United States in 1992*. Alexandria, VA, 1993.

American Medical Association. *Physicians' Current Procedural Terminology (CPT)*. Chicago, IL: American Medical Association, 1994.

Diabetes Control and Complications Trial Research Group. "Intensive Diabetes Treatment and Complications in IDDM." *New England Journal of Medicine* 329:977-986, 1993.

Fowles, J., Lawthers, A., Weiner, J., Garnick, D., Petrie, D., and Palmer, R. "Agreement Between Physicians' Office Records and Medicare Part B Claims Data." *Health Care Financing Review*, 16:189-198, 1995 .

Gilmer, T., Manning, W., O'Connor, P., and Rush, W. "The Cost to Health Plans of Poor Glycemic Control." *Diabetes Care* 20:1847-1853, 1997.

Horner, R., Paris, J., Purvis, J., and Lawler, F. "Accuracy of Patient Encounter and Billing Information in Ambulatory Care." *Journal of Family Practice* 33:593-598, 1991.

Kehoe, R., Wu, S., Leske, C., and Chylack, L. "Comparing Self-reported and Physician-reported Medical History." *American Journal of Epidemiology* 139:813-818, 1994.

Kenny, S., Aubert, R., and Geiss, L. "Prevalence and Incidence of Non-Insulin-Dependent Diabetes. National Diabetes Data Group." *Diabetes in America*, 2nd ed., NIH Publication No. 95-1468. Bethesda, MD: National Institutes of Health, 1995.

Kriegsman, D., Penninx, B., van Eijk, J., Boeke, A., and Deeg, D. J. "Self-reports and General Practitioner Information on the Presence of Chronic Diseases in Community Dwelling Elderly. A Study on the Accuracy of Patients' Self-reports and on Determinants of Inaccuracy." *Journal of Clinical Epidemiology* 49:1407-1417, 1996.

NHANES III. Unpublished data from the National Health and Nutrition Examination Survey, III 1988-1994. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Centers for Health Statistics, Hyattsville, MD, 1997.

Office of Management and Budget. *OMB Bulletin, 93-17*. Washington, DC: U.S. Government Printing Office, 1993.

Parente, S., Weiner, J., Garnick, D., Richards, T., Fowles, J., Lawthers, A., Chandler, P., and Palmer, R. "Developing a Quality Improvement Database Using Health Insurance Data: a

Guided Tour with Application to Medicare's National Claims History File." *American Journal of Medical Quality* 10:162-176, 1995.

Peer Review Organization (PRO) for the State of Washington. *Operation of Utilization and Quality Control*. (Report No. 500-96-P618). Seattle, WA: PRO-West, 1997.

Quam, L., Ellis, L., Venus, P., Clouse, J., Taylor, C., and Leatherman, S. "Using Claims Data for Epidemiologic Research." *Medical Care* 31:498-507, 1993.

Robinson, J., Young, T., Roos, L., and Gelskey, D. "Estimating the Burden of Disease." *Medical Care* 35:932-947, 1997.

Stata Corporation. Version 4.0, Stata Statistical Software. College Station, TX: Stata Press, 1995.

U.S. General Accounting Office. *Beneficiaries with Diabetes Do Not Receive Recommended Monitoring Services*, (Report No. HEHS-97-48). Washington D.C.: US G