Specificity and Sensitivity of Claims-Based Algorithms for Identifying Members of Medicare+Choice Health Plans That Have Chronic Medical Conditions

Thomas S. Rector, Steven L. Wickstrom, Mona Shah, N. Thomas Greenlee, Paula Rheault, Jeannette Rogowski, Vicki Freedman, John Adams, and José J. Escarce

Objective. To examine the effects of varying diagnostic and pharmaceutical criteria on the performance of claims-based algorithms for identifying beneficiaries with hypertension, heart failure, chronic lung disease, arthritis, glaucoma, and diabetes.

Study Setting. Secondary 1999–2000 data from two Medicare+Choice health plans.

Study Design. Retrospective analysis of algorithm specificity and sensitivity.

Data Collection. Physician, facility, and pharmacy claims data were extracted from electronic records for a sample of 3,633 continuously enrolled beneficiaries who responded to an independent survey that included questions about chronic diseases.

Principal Findings. Compared to an algorithm that required a single medical claim in a one-year period that listed the diagnosis, either requiring that the diagnosis be listed on two separate claims or that the diagnosis to be listed on one claim for a face-to-face encounter with a health care provider significantly increased specificity for the conditions studied by 0.03 to 0.11. Specificity of algorithms was significantly improved by 0.03 to 0.17 when both a medical claim with a diagnosis and a pharmacy claim for a medication commonly used to treat the condition were required. Sensitivity improved significantly by 0.01 to 0.20 when the algorithm relied on a medical claim with a diagnosis or a pharmacy claim, and by 0.05 to 0.17 when two years rather than one year of claims data were analyzed. Algorithms that had specificity more than 0.95 were found for all six conditions. Sensitivity above 0.90 was not achieved all conditions.

Conclusions. Varying claims criteria improved the performance of case-finding algorithms for six chronic conditions. Highly specific, and sometimes sensitive, algorithms for identifying members of health plans with several chronic conditions can be developed using claims data.

Key Words. Medical informatics, claim review, algorithms, specificity and sensitivity

Health services researchers, health plan administrators, and health policy-makers often need to identify people with specific chronic medical conditions,

such as diabetes or heart failure. For example, researchers may be interested in assessing the outcomes of alternative treatments. Health plan administrators may wish to identify members for quality improvement or disease management programs. Policymakers may be interested in tracking access to or quality of care. These efforts typically rely on diagnoses listed on physician and hospital claims submitted by providers to health insurers, and pharmacy claims when available, to identify patients who have the conditions of interest.

Information about chronic conditions in administrative databases may be incomplete or inaccurate for a variety of reasons (Virmig and McBean 2001). For example, health care that is not covered or billed is not reflected in the data. Information that is unnecessary for processing payments may not be collected or recorded accurately. Even when care is sought for a chronic condition, the diagnosis might not appear on provider claims (Horner et al. 1991; Fowles et al. 1995; Fowles, Fowler, and Craft 1998). Conversely, diagnoses listed on claims may be related to testing for disease rather than confirmed disease. Despite these caveats, administrative data are used extensively by researchers and by the National Commission for Quality Assurance, the Centers for Medicare and Medicaid Services, and managed care organizations to identify people that have specific chronic conditions (National Committee for Quality Assurance 1999; Centers for Medicare and Medicaid Services 2002a; Centers for Medicare and Medicaid Services 2002b).

Few studies have described efforts to maximize the performance of claims-based algorithms for identifying health plan enrollees with chronic medical conditions. We previously found that the positive predictive value of claims collected over a two-year period could be improved by requiring two or more claims rather than one claim listing a diagnosis of hypertension, or by

This work was supported, in part, by Grant No. RO1-HS/AG09630 from the Agency for Healthcare Research and Quality.

Address correspondence to Thomas S. Rector, Ph.D., Research Specialist, Center for Chronic Disease Outcomes Research, VA Medical Center, 152/2E, One Veterans Dr., Minneapolis, MN 55417. Steven L. Wickstrom, M.S., is a Senior Researcher, Center for Health Care Policy & Evaluation, Eden Prairie, MN. Mona Shah, M.S., is a Senior Network Information Analyst, Blue Cross and Blue Shield of Minnesota, Eagan, MN. N. Thomas Greenlee, M.S., is a Major, USAF, MSC, Bolling AFB, Washington, DC. Paula Rheault, B.S., is a Programmer, Center for Health Care Policy & Evaluation, Eden Prairie, MN. Jeannette Rogowski, Ph.D., is a Senior Economist with RAND, Arlington, VA. Vicki Freedman, Ph.D., is Director, Polisher Research Institute Abramson Center for Jewish Life, North Wales, PA. John Adams, Ph.D., is a Senior Statistician with RAND, Santa Monica, CA. José J. Escarce, M.D., Ph.D., is Professor of Medicine, UCLA, David Geffen School of Medicine, Los Angeles, and Senior Natural Scientist, RAND, Santa Monica, CA.

combining the diagnostic criterion with the requirement for a pharmacy claim for a medication commonly used to treat hypertension (Quam et al. 1993). In a comparison of self-reports of chronic conditions and administrative data, Robinson et al. (1997) found that more cases of self-reported diabetes or hypertension were confirmed by administrative data as the number of years of claims data increased. Others studied factors that affect the sensitivity and specificity of case-finding algorithms using Medicare claims to identify people who reported they had diabetes on the Medicare Current Beneficiary Survey (Hebert et al. 1999). The sensitivity of their algorithms increased by approximately 10 percent with a small drop in specificity when two rather than one year of claims data were used. For different types of claims (hospital, physician, etc.) requiring the diabetes diagnosis to be the first one listed on the claim or listed on two claims rather than one or listed on claims associated with direct physician contact improved the already near-perfect specificity slightly and lowered the sensitivity.

Further investigation of the sensitivity and specificity of case-finding algorithms for chronic conditions using different administrative data is warranted given the extensive use of claims data for this purpose. This study extends previous research by systematically evaluating the performance of a wide range of alternative claims-based algorithms for identifying members of Medicare+Choice plans with six chronic medical conditions that are prevalent in the elderly population and are frequently the focus of research studies and managed care initiatives. We used survey data collected from members of two Medicare+Choice health plans to assess the sensitivity and specificity of algorithms that required different claims criteria for the diagnosis, more than one claim with the same diagnosis, one or two years of claims history, and incorporation of pharmacy claims to identify patients with hypertension, heart failure, diabetes, arthritis, glaucoma, or chronic lung disease.

METHODS

Setting

Both study plans had Medicare risk contracts under the Medicare+Choice program and contracted with health care providers in their communities to create provider networks. One plan was located in a Midwestern metropolitan area, the other in the Northeast. Members of both plans were required to select a primary care physician. Only members of the Northeastern plan were required to obtain a referral before they saw a specialist. During the period of the study (1999 and 2000), the copayment for an office visit was \$10 in the

Midwestern plan and \$15 in the Northeastern plan. The annual limit on drug benefits in the Midwestern plan was \$900 in 1999 and \$500 in 2000; the limit in the Northeastern plan was \$300 for both years.

Survey Data

A study of the influence of socioeconomic status on the utilization of medical care by elderly members of the study plans was the source of information on chronic conditions. Briefly, members who were enrolled in the plans in January 2000 were randomly sampled within socioeconomic strata defined by dual (Medicare and Medicaid) eligibility and household income at the zip code level. Members who were disabled and less than 65 years old, members in institutions or hospice programs, and members with end-stage renal disease were excluded from the study population. All 942 dually eligible members of the health plans were included in the sample, as were 700 members living in zip codes where approximately 50 percent of the households had incomes less than 200 percent of the poverty level, and 5,354 other members. The total sample consisted of 6,996 members equally distributed between the two plans.

Telephone interviews were completed for 4,613 members in the sample between April and October 2000 (response rate, 72 percent, after adjusting for ineligible subjects). Six of the interview questions asked, "Has a doctor ever told you that you had (1) high blood pressure or hypertension, (2) congestive heart failure, (3) chronic lung disease such as chronic bronchitis, emphysema, or asthma, (4) arthritis, (5) glaucoma, or (6) diabetes or high blood sugar?" Respondents answered yes, no, or don't know or refused to answer these questions. Our analysis excluded the 0.1 percent to 0.6 percent of respondents who did not give a definite yes or no answer to each question about the presence of a chronic condition. An additional 548 surveys completed with proxy respondents were excluded as well.

Claims Data

We used physician, facility, and pharmacy claims to develop case-finding algorithms for each chronic condition. Physician claims contained information on each service provided to health plan members by physicians, including the Physicians' Current Procedural Terminology (CPT) code for the service, the date of the service, and up to four diagnoses identified using *International Classification of Diseases*, 9th Edition, Clinical Modification (ICD-9-CM) diagnosis codes. We used the CPT codes to identify claims for patient visits to a physician (so-called face-to-face encounters), which may have more

valid diagnostic codes than other types of claims such as those for laboratory tests or diagnostic procedures that are often used to rule out a condition even though they might have been done directly by a physician. The CPT codes representing a visit to the emergency room were not included as face-to-face encounters because of the potential for more “rule-out” diagnoses during emergency care. (See appendix for codes used to define face-to-face encounters. Online-only appendix is available at <http://www.blackwell-synergy.com>.)

Facilities (for example, hospitals) submit claims for inpatient, outpatient, and ancillary services. Each facility claim listed a date of service, up to nine diagnosis codes, and up to six procedures using either CPT or ICD-9-CM procedure codes. Facility claims are itemized by “revenue codes” that represent units of service such as use of an inpatient bed day, an emergency room visit, or a radiologic test. Similar to physician claims, facility claims were classified as face-to-face encounters using procedure codes for physician visits or revenue codes. The revenue codes typically represent room and board charges. (See appendix for codes used.)

Pharmacy claims list the date the medication was dispensed and the National Drug Code (NDC) that was linked to proprietary codes for generic ingredients. Pharmacy claims were further grouped and coded by therapeutic class such as medications used to treat diabetes or hypertension or pharmacological class such as loop diuretics, angiotensin-converting enzyme inhibitors, or nonsteroidal anti-inflammatory agents to simplify incorporation into algorithms.

Claims-Based Algorithms

We evaluated 38 different algorithms for each of the six chronic conditions using physician, facility, and pharmacy claims for 3,633 survey respondents who were continuously enrolled during the two-year period from 1999 to 2000 and met the other inclusion criteria. The ICD-9-CM diagnosis codes and types of medications used to identify members with the six conditions are listed in Table 1.

Case-finding algorithms for each condition were developed and tested by systematically varying several factors. Three algorithms searched claims from 1999 for (1) an identifying diagnosis on any physician or facility claim, (2) an identifying diagnosis on a face-to-face claim as previously defined, and (3) an identifying diagnosis listed in the primary position on a face-to-face claim. Three additional algorithms searched diagnostic information in claims as

Table 1: Diagnostic Codes and Types of Medications Used to Identify Members with Chronic Conditions

<i>Condition</i>	<i>ICD-9* Code Numbers</i>	<i>Medications</i>
Diabetes mellitus	250.xx (diabetes mellitus)	insulin
	357.2x (neuropathy in diabetes)	oral agents including sulfonylureas,
	362.0x (diabetic retinopathy)	thiazolidinediones, biguanides, and
	366.41 (diabetic cataract)	alpha-glucosidase inhibitors
Hypertension	401.0, 401.1, 401.9 (essential hypertension)	diuretics
	402.xx (hypertensive heart disease)	angiotensin converting enzyme inhibitors
	403.xx (hypertensive renal disease)	angiotensin-II receptor antagonists
	404.xx (hypertensive heart and renal disease)	beta-adrenergic receptor blockers
		calcium channel blockers
		central and alpha-adrenergic antagonists
		vasodilators
Heart failure	398.91 (rheumatic heart failure)	combinations of above
	402.01, 402.11, 402.91, 404.01, 404.11, 404.91, 404.03, 404.13, 404.93 (hypertensive heart disease with heart failure)	loop diuretics
	428.xx (heart failure)	angiotensin converting enzyme inhibitors
		angiotensin-II receptor antagonists
		digoxin
Arthritis	714.xx (rheumatoid arthritis and other inflammatory polyarthropathies)	select beta-adrenergic receptor blockers—carvedilol, metoprolol and bisoprolol
	715.xx (osteoarthritis and allied disorders)	non-steroidal anti-inflammatory drugs
	716.xx (other and unspecified arthropathies)	salicylates
		corticosteroids
		disease modifying agents including gold salts, pencillamine, leflunomide, etanercept, sulfasalazine, methotrexate, hydroxychloroquine, azathioprine, and cyclophosphamide
	720.xx (except 720.1: ankylosing spondylitis and other inflammatory spondylopathies)	
	721.0 to 721.4x, 721.9x (spondylosis)	
Chronic lung disease	491.xx (chronic bronchitis)	oral and inhaled beta-adrenergic agonists
	492.xx (emphysema)	inhaled corticosteroids
	493.xx (asthma)	leukotriene inhibitors
	496.xx (chronic airway obstruction)	xanthine derivatives
Glaucoma		ipratropium
		mast cell stabilizers
	365.1x to 365.9x (types of glaucoma)	ocular beta-adrenergic blockers
	365.0x (borderline glaucoma or glaucoma suspect)	ocular cholinesterase inhibitors
		ocular cholinergic miotics
		ocular carbonic anhydrase inhibitors
		ocular sympathomimetics
		ocular prostaglandin analogues

**International Classification of Diseases, 9th Clinical Modification where “x” indicates any digit.*

above except they required an identifying diagnosis to be listed on two claims with different dates of service.

One algorithm used only pharmacy claims from 1999 to identify individuals. Six algorithms identified cases using an identifying pharmacy claim *or* one of the six diagnostic criteria, while an additional six algorithms required the presence of both an identifying pharmacy claim *and* one of six diagnostic criteria. Finally, we applied each of the above 19 algorithms to claims from both 1999 and 2000.

In preliminary analyses, we evaluated claims-based algorithms that incorporated therapeutic procedures such as administration of home oxygen as an indication of chronic lung disease, administration of intravenous inotropes for heart failure, and intravenous infusion of infliximab or joint replacement surgery for arthritis. In general, it was rare to have a claim for a therapeutic procedure without a claim listing the diagnosis of interest. Consequently, information about therapeutic procedures did not enhance the claim-based algorithms, and was not used in the algorithms presented in this paper.

Data Analysis

We adopted members' self-reports on the survey as the gold standard for determining who had a chronic condition, and calculated the sensitivity and specificity of the claims-based algorithms relative to this gold standard. We defined sensitivity as the proportion of members reporting that a doctor told them they had a condition who were identified by an algorithm as having the condition. We defined specificity as the proportion of members reporting that they did not have a condition who were not identified by the algorithm. The sensitivity and specificity calculated for the 19 algorithms based on one year of claims were summarized for each condition by plotting the sensitivity on the vertical axis and one minus the specificity on the horizontal axis, similar to a receiver operating characteristic (ROC) curve (McNeil, Keeler, and Adelstein 1975). One minus the specificity represents the proportion of people not reporting a condition who were apparently falsely identified by an algorithm. McNemar's test was used to determine when sensitivity or specificity of algorithms differed significantly when applied to the same sample (Bennett 1972).

We also calculated the ratio of the sensitivity to one minus the specificity, known as the likelihood ratio positive (LR+). The LR+ is a measure of the ability of an algorithm to "rule in" a condition, and represents how much the odds that a plan member has a condition increase when identified by the algorithm as having it. In addition, we calculated the ratio of one minus the

sensitivity to the specificity, known as the likelihood ratio negative (LR⁻). The LR⁻ is a measure of the ability of an algorithm to “rule out” a condition, and represents how much the odds that a member has a condition decrease when identified by the algorithm as not having it.

RESULTS

Self-reported prevalence varied substantially across the six chronic conditions in the study. Hypertension (58 percent) and arthritis (56 percent) were the most common conditions, while diabetes (18 percent), chronic lung disease (14 percent), glaucoma (10 percent), and heart failure (6 percent) were reported less frequently. The prevalence of these conditions among the 3,633 individuals used to study the sensitivity and specificity of the claims-based algorithms was similar to the prevalence in the entire group of 4,613 survey respondents (data not shown).

The ROC curves in Figures 1 and 2 summarize the results from the 19 algorithms applied to the 1999 claims data for each condition. A supplemental appendix showing the sensitivity and specificity of all 38 algorithms for all six conditions is available in the electronic version of this report (see <http://www.blackwell-synergy.com>). The sensitivity and specificity of the algorithms varied considerably for all the conditions. For hypertension, the specificity ranged from 0.60 to 0.96, while the sensitivity ranged from 0.32 to 0.90. For heart failure, the specificity ranged from 0.75 to 0.99 and the sensitivity from 0.23 to 0.78. For chronic lung disease, the specificity ranged from 0.87 to 0.99 and the sensitivity from 0.22 to 0.62. For arthritis, the specificity ranged from 0.77 to 0.99 and the sensitivity from 0.07 to 0.55. For glaucoma, the specificity ranged from 0.95 to 0.99 and the sensitivity from 0.32 to 0.73. Lastly, for diabetes, the specificity ranged from 0.93 to 0.99 and the sensitivity from 0.44 to 0.91.

Interestingly, the ROC curves for heart failure and for arthritis exhibited a “discontinuity” in which seven of the algorithms do not appear to follow the contour of the curve established by the other 12 algorithms. The seven algorithms are displaced toward lower specificities, although they generally have higher sensitivities. These seven algorithms used only pharmacy claims, or an identifying pharmacy claim *or* one of the six diagnostic criteria. Use of pharmacy claims alone to identify cases of arthritis or heart failure increased sensitivity at the price of a substantial decline in specificity, especially for heart failure.

Figure 1: Variation in Sensitivity and Specificity of 19 Claims-Based Algorithms Developed to Identify Medicare+Choice Members Who Self-Reported They Had Hypertension, Heart Failure, or Chronic Lung Disease. (See Methods for a description of the 19 algorithms.) Points above the 45° line where the sensitivity equals one minus the specificity represent algorithms that discriminated to some degree members who did or did not report they had the condition.

Figure 2: Variation in Sensitivity and Specificity of 19 Claims-Based Algorithms Developed to Identify Medicare+Choice Members Who Self-Reported They Had Arthritis, Glaucoma, or Diabetes. (See Methods for a description of the 19 algorithms.) Points above the 45° line where the sensitivity equals one minus the specificity represent algorithms that discriminated to some degree members who did or did not report they had the condition.

The algorithm that required at least two face-to-face claims with a first-listed diagnosis and at least one prescription for a medication commonly used to treat the condition produced the highest specificity for all six conditions. The specificity of this algorithm was 0.99 for all the conditions except hypertension, where it was 0.96. However, the sensitivity of this algorithm was uniformly low: 0.32, 0.23, 0.22, 0.07, 0.32, and 0.44, respectively, for hypertension, heart failure, chronic lung disease, arthritis, glaucoma, and diabetes. This algorithm had the highest LR+ for all six conditions, with values of 8, 24, 24, 21, 50, and 23. Therefore, it is useful for identifying cases that almost surely have the condition of interest—that is, for “ruling in” the condition—although it yields many false negatives. The algorithm that required at least two face-to-face claims with a diagnosis in any position and at least one prescription for a medication commonly used to treat the condition had lower specificity by only 0.01 or less for all the conditions compared to the algorithm that required the diagnosis to be in the first-listed position, and it had higher sensitivity by 0.03 to 0.08. Thus requiring that the diagnosis be the first one listed did not improve specificity very much while reducing sensitivity appreciably.

The algorithm that required only one diagnosis in any position on any type of claim or a pharmacy claim produced the highest sensitivity for all six conditions. The sensitivity of this algorithm for hypertension, heart failure, chronic lung disease, arthritis, glaucoma, and diabetes was 0.90, 0.78, 0.62, 0.55, 0.73, and 0.91, respectively. Conversely, the specificity of this algorithm was lower than for most other algorithms: 0.60, 0.75, 0.87, 0.77, 0.95, and 0.93. This algorithm had the lowest LR – for all six conditions, with values of 0.17, 0.30, 0.43, 0.59, 0.28, and 0.10. Thus it is useful for “ruling out” the condition of interest.

In practice, researchers and health plan managers may often wish to use algorithms that have high specificity, to avoid falsely identifying as cases people who do not actually have the conditions of interest, but that simultaneously possess good sensitivity, to identify as many true cases as possible. As shown on Figures 1 and 2, several other algorithms had nearly as high specificity and better sensitivity compared with the algorithms that required at least two face-to-face claims with the diagnosis and at least one prescription for a medication used to treat the condition. Table 2 lists algorithms for each condition that had the highest sensitivity while maintaining the specificity above 0.90. Whenever the most sensitive algorithm required pharmacy claims, the most sensitive without pharmacy claims is tabulated as well. Also, when two algorithms that did not incorporate pharmacy claims had similar sensitivity, both are listed.

Table 2: Case-Finding Algorithms with the Highest Sensitivity and Specificity of at Least 0.90 Using One Year of Claims

<i>Condition</i>	<i>Algorithm</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>LR+</i>	<i>LR-</i>
Hypertension	Diagnosis on 1 face-to-face claim and a pharmacy claim	0.52	0.91	5.6	0.53
Hypertension	Diagnosis on 2 face-to-face claims	0.48	0.90	4.8	0.58
Heart failure	Diagnosis on 1 claim	0.58	0.93	7.0	0.45
Chronic lung disease	Diagnosis on 1 face-to-face claim or a pharmacy claim	0.56	0.91	6.6	0.48
Chronic lung disease	Diagnosis on 1 face-to-face claim	0.50	0.93	7.2	0.54
Arthritis	Diagnosis on 1 face-to-face claim	0.29	0.93	4.3	0.76
Arthritis	Diagnosis on 2 or more claims	0.28	0.94	4.7	0.77
Glaucoma	Diagnosis on 1 claim or a pharmacy claim	0.73	0.95	14	0.28
Glaucoma	Diagnosis on 1 claim	0.68	0.95	14	0.34
Diabetes	Diagnosis on 1 claim or a pharmacy claim	0.91	0.93	13	0.10
Diabetes	Diagnosis on 1 claim	0.90	0.93	13	0.11

The most sensitive algorithms that still had specificity greater than 0.90 relied on pharmacy claims to help detect hypertension, chronic lung disease, glaucoma, and diabetes, although for each of these conditions there was also an algorithm that was nearly as sensitive that did not require pharmacy claims. Requiring only one claim with diagnosis appeared to be a useful approach for identifying members with diabetes, glaucoma, heart failure, chronic lung disease, and arthritis even though the diagnosis had to be on a face-to-face claim for the latter two conditions. It is important to note that the sensitivity of the algorithms listed in Table 2 is modest for all the conditions except glaucoma and diabetes.

Tests of the effects of using different diagnostic criteria on the specificity and sensitivity of case-finding algorithms are shown in Table 3. Compared with requiring any one claim with a diagnosis, requiring that the diagnosis be listed on a claim for a face-to-face encounter or requiring at least two claims that list the diagnosis increased the specificity for all conditions, but also reduced sensitivity. The effects on specificity were smallest when the specificity was already greater than 0.90. Requiring that the diagnosis on the claim for a face-to-face encounter be the first one listed increased the specificity only slightly, but decreased the sensitivity appreciably for most of the conditions.

Table 3: Effect of Different Diagnoses Criteria Applied to One Year of Claims on Specificity and Sensitivity of Case-Finding Based Algorithms for Six Chronic Conditions

<i>Chronic Condition</i>	<i>(A) Diagnosis on 1 Claim</i>		<i>(B) Diagnosis on ≥ 2 Claims</i>		<i>(C) Diagnosis On 1 Face-to-Face[§] Claim</i>		<i>(D) 1st Diagnosis on 1 Face-to-Face Claim</i>	
	<i>Sens.</i>	<i>Spec.</i>	<i>Sens.</i>	<i>Spec.</i>	<i>Sens.</i>	<i>Spec.</i>	<i>Sens.</i>	<i>Spec.</i>
Diabetes	0.90	0.93	0.85*	0.96*	0.82*	0.96*	0.72 [#]	0.98 [#]
Hypertension	0.83	0.69	0.72*	0.80*	0.67*	0.80*	0.62 [#]	0.83 [#]
Heart failure	0.58	0.93	0.47*	0.95*	0.45*	0.96*	0.42 [#]	0.97 [#]
Arthritis	0.43	0.87	0.28*	0.94*	0.29*	0.93*	0.23 [#]	0.95 [#]
Lung disease	0.59	0.89	0.48*	0.94*	0.50*	0.93*	0.44 [#]	0.95 [#]
Glaucoma	0.68	0.95	0.57*	0.98*	0.61*	0.96*	0.58 [#]	0.96 [#]

[§]Claim with a procedure code for a physician visit or a revenue code for an inpatient bed day to indicate member had a face-to-face encounter with a physician. Sens. = sensitivity, Spec. = specificity.

* $p \leq 0.01$, comparing column B or C to column A.

[#] $p \leq 0.01$, comparing column D to column C.

Sensitivities and specificities of algorithms that incorporated pharmacy claims are shown in Table 4. Except for heart failure, algorithms based only on pharmacy claims were as or more specific but less sensitive than the algorithms based on a diagnosis in any position on a single claim. Algorithms that used diagnosis *or* pharmacy criteria to identify members had better sensitivity, however they had lower specificity with the exception of algorithms requiring one diagnosis for diabetes and glaucoma. Use of diagnosis *and* pharmacy criteria significantly increased the specificity of case-finding algorithms for all conditions except pharmacy-based algorithms for diabetes and glaucoma that already had a specificity of 0.99. Requiring both a diagnosis and pharmacy claim reduced sensitivity.

When the algorithms that required a single diagnosis on any claim were applied to two years of claims data (1999 and 2000) rather than one year (1999) the sensitivity of the algorithms increased significantly for all conditions: diabetes (0.90 to 0.95), hypertension (0.83 to 0.92), heart failure (0.58 to 0.74), arthritis (0.43 to 0.60), chronic lung disease (0.59 to 0.70), and glaucoma (0.68 to 0.75). With the exception of hypertension, reductions in specificity were similar to or less than the gains in sensitivity: diabetes (0.93 to 0.88), hypertension (0.69 to 0.56), heart failure (0.93 to 0.88), arthritis (0.87 to 0.77), chronic lung disease (0.89 to 0.82), and glaucoma (0.95 to 0.92). Table 5 lists

Table 4: Effect of Pharmacy Claims on Sensitivity and Specificity of Claims-Finding Algorithms for Six Chronic Conditions

Chronic Condition	(A) Diagnosis on 1 Claim		(B) Pharmacy§ Claim		(A or B)		(A and B)	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
Diabetes	0.90	0.93	0.64	0.99	0.91* [#]	0.93 [#]	0.62 ^{‡,⊖}	0.99 [‡]
Hypertension	0.83	0.69	0.71	0.77	0.90* [#]	0.60* [#]	0.64 ^{‡,⊖}	0.86 ^{‡,⊖}
Heart failure	0.58	0.93	0.58	0.78	0.78* [#]	0.75* [#]	0.39 ^{‡,⊖}	0.96 ^{‡,⊖}
Arthritis	0.43	0.87	0.32	0.87	0.55* [#]	0.77* [#]	0.20 ^{‡,⊖}	0.96 ^{‡,⊖}
Lung disease	0.59	0.89	0.38	0.96	0.62* [#]	0.87* [#]	0.35 ^{‡,⊖}	0.98 ^{‡,⊖}
Glaucoma	0.68	0.95	0.50	0.99	0.73* [#]	0.95 [#]	0.45 ^{‡,⊖}	0.99 [‡]

§Claim for a medication commonly, but not exclusively, used to treat the condition (see Table 1).

Sens. = sensitivity, Spec. = specificity.

* $p \leq 0.01$, comparing column (A or B) to column A.

[#] $p \leq 0.01$, comparing column (A or B) to column B.

[‡] $p \leq 0.01$, comparing column (A and B) to column A.

[⊖] $p \leq 0.01$, comparing column (A and B) to column B.

algorithms using two years of data that had the highest sensitivity while maintaining specificity above 0.90, except for hypertension where algorithms with the highest specificity are presented. Four conditions (hypertension, heart failure, glaucoma, and diabetes) had algorithms with a sensitivity ≥ 0.70 when two years of claims data were analyzed.

DISCUSSION

Case identification algorithms based on diagnoses listed on physician and facility claims and pharmacy claims exhibited varying performance in identifying members of Medicare+Choice plans who reported they had one of six chronic conditions. Algorithms with a specificity of at least 95 percent were found for all conditions. However, highly specific algorithms may not identify substantial proportions of members who have the conditions of interest because these algorithms typically have low sensitivity. Diabetes was the only condition where an algorithm had a specificity and sensitivity greater than 0.90.

The positive and negative predictive values of these algorithms will depend on the prevalence of the condition in the population being screened as

Table 5: Case-Finding Algorithms with the Highest Sensitivity and Specificity of at Least 0.90 (or Highest Observed) Using Two Years of Claims

<i>Condition</i>	<i>Algorithm</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>LR+</i>	<i>LR–</i>
Hypertension	Diagnosis on 2 primary face-to-face claims	0.61	0.85	4.1	0.46
Hypertension	Diagnosis on 2 face-to-face claims and a pharmacy claim	0.70	0.83	4.1	0.36
Hypertension	Diagnosis on 2 claims and a pharmacy claim	0.73	0.81	3.8	0.34
Heart failure	Diagnosis on 1 face-to-face claim	0.70	0.91	7.8	0.33
Heart failure	Diagnosis on 2 claims	0.67	0.92	8.0	0.36
Heart failure	Diagnosis on 1 claim and a pharmacy claim	0.53	0.92	7.0	0.50
Heart failure	Diagnosis on 1 face-to-face claim and a pharmacy claim	0.51	0.94	8.3	0.53
Heart failure	Diagnosis on 2 claims and a pharmacy claim	0.49	0.94	8.4	0.54
Lung disease	Diagnosis on 2 face-to-face claims	0.56	0.91	6.4	0.48
Lung disease	Primary diagnosis on 1 claim	0.55	0.92	6.7	0.49
Lung disease	Diagnosis on 2 primary face-to-face claims or a pharmacy claim	0.58	0.92	7.5	0.45
Arthritis	Diagnosis on 2 face-to-face claims	0.37	0.90	3.6	0.70
Arthritis	Diagnosis on 1 or more claim and a pharmacy claim	0.32	0.92	3.9	0.74
Glaucoma	Diagnosis on 1 claim	0.75	0.92	9.5	0.27
Glaucoma	Diagnosis on 1 face-to-face claim	0.74	0.93	9.9	0.40
Glaucoma	Diagnosis on 1 claim or a pharmacy claim	0.78	0.92	9.4	0.24
Glaucoma	Diagnosis on 1 face-to-face claim or a pharmacy claim	0.76	0.92	9.8	0.26
Diabetes	Diagnosis on 2 face-to-face claims	0.91	0.95	18.9	0.10
Diabetes	Diagnosis on 2 claims	0.93	0.93	14.4	0.07
Diabetes	Diagnosis on 1 face-to-face claim	0.94	0.92	11.3	0.07
Diabetes	Diagnosis on 2 face-to-face claims or a pharmacy claim	0.93	0.95	18.1	0.08
Diabetes	Diagnosis on 2 claims or a pharmacy claim	0.94	0.93	14.1	0.07
Diabetes	Diagnosis on 1 face-to-face claim or a pharmacy claim	0.94	0.92	11.2	0.06

well as the sensitivity and specificity (McNeil, Keeler, and Adelstein 1975). For example, assuming 58 percent of the study population truly did have hypertension, the positive predictive value for finding members with hypertension using the algorithm requiring one face-to-face claim and a pharmacy claim that had an estimated specificity of 0.91 and sensitivity of 0.52 would be 89

percent. In other words, 89 percent of the members identified by the algorithm would be expected to self-report they had hypertension. However, the positive predictive value would be only 34 percent for a less prevalent condition such as heart failure (estimated prevalence = 6 percent) even though the estimated sensitivity and specificity of the algorithm requiring one claim with the diagnosis were 0.93 and 0.58, respectively.

Selection of an appropriate algorithm depends on whether it is more important to find as many cases of a condition as possible using an algorithm that has high sensitivity but that also identifies many false positives, or to find fewer but nearly certain cases using an algorithm that has high specificity but misses many members who have the condition. When high sensitivity is of primary interest, our findings suggest that the sensitivity of claims-based algorithms might be improved by (1) using either diagnostic or pharmacy criteria to identify cases, or (2) using a longer period of time to capture claims that list a diagnosis for a chronic condition or a medication. Although these techniques improved the sensitivity of algorithms for all six conditions, the magnitude of the effect varied considerably. We can only speculate about what caused this variation. Sensitivity of the diagnosis-based algorithm for diabetes was already 0.90 leaving little room for improvement by examining pharmacy claims or a longer period of time. Pharmacy claims would not identify people with conditions that are not being treated, such as hypertension or diabetes being controlled by diet alone and some types of chronic lung disease that are generally not responsive to medications. Adherence to medication prescriptions might vary across conditions as well. People with arthritis might be using nonprescription medications that are generally not covered by drug benefits. Unlike medications for diabetes and glaucoma, medications for heart failure are not used solely for this condition. Therefore, use of pharmacy claims may increase sensitivity of algorithms for heart failure while substantially decreasing specificity.

We did not determine when use of pharmacy claims, or longer periods of time, or both, would be the best way to improve the sensitivity of case finding algorithms. Information in the supplemental electronic appendix could be used to explore these questions (see <http://www.blackwell-synergy.com>). Additional investigations are also needed to determine whether the sensitivity of algorithms might be improved by including claims for diagnostic procedures that are typically done by a physician. We excluded these types of "face-to-face" claims because of concerns that the diagnosis code might be the indication for the test rather than the conclusion from test results, thus any gain in sensitivity from using claims for diagnostic procedures

might be accompanied by a decrease in specificity. Furthermore, if physicians usually submit claims for both the diagnostic procedure and a patient visit, addition of the claims for diagnostic tests might not add any information. We observed this phenomenon when we tried to improve the sensitivity of algorithms by using claims for therapeutic procedures.

When specificity is most important, algorithms might be improved by requiring that the diagnosis be listed on a claim for face-to-face physician encounter, requiring both a provider claim that lists the diagnosis *and* a pharmacy claim for a medication commonly used to treat the condition of interest, or requiring at least two claims with different dates of service that list the diagnosis of interest. Algorithms that required two face-to-face claims and at least one corresponding prescription had good specificity for all conditions studied.

In many practical applications, researchers and health plan administrators may wish to use algorithms that have high specificity and as high of a sensitivity as possible, both to identify as many cases for study as is feasible and to minimize the selection bias that can occur when cases are identified using algorithms with very low sensitivity. Tables 2 and 5 can be used by researchers and plan administrators as guides for choosing appropriate algorithms depending on whether they have one or two years of claims data and on whether they have pharmacy claims.

Our study has several limitations. Use of self-reports as a gold standard for conditions may have introduced error into our estimates of sensitivity and specificity. Self-reports are subject to lack of knowledge or misperceptions about the presence of a condition, and the imprecise lay terminology used in surveys may be inconsistent with medical definitions of disease. Validation of the self-reports by review of medical records was not feasible for this secondary analysis of data. Previous studies have indicated that self-reports of the absence of chronic conditions including diabetes, hypertension, heart failure, chronic lung disease, and joint problems agree with medical records in a high percentage of cases (Fowles, Fowler, and Craft 1998; Martin et al. 2000). These and other studies have found that people who do have chronic conditions tend to under report their presence, and the extent of underreporting varies by condition (Turner et al. 1997). Whenever a claims-based algorithm identified members who falsely reported they did not have the condition, the estimated specificity would be less than the true specificity. Nevertheless, our observations on how one can improve the specificity of claims-based algorithms by requiring more than one claim with the diagnosis or that the diagnosis be listed on a claim for a face-to-face encounter was consistent with previous research

(Quam et al. 1993; Hebert et al. 1999). However, we did not program and compare exactly the same algorithms presented in previous reports. Further evaluation is needed to determine which algorithms are optimal or equivalent.

Our study was limited to two Medicare+Choice health plans that had limited outpatient drug benefits. Thus, our results may not be representative of algorithm performance when applied to the overall Medicare population or to younger insured populations. Furthermore, the performance of pharmacy claims may differ in health plans where outpatient drug benefits are less limited.

Although several nuances of claims data make them less than perfect for the identification of health plan members who have a particular condition, our results suggest that claims-based algorithms can help identify members who have the six chronic conditions studied. Researchers and health plan administrators must choose algorithms that yield the performance they desire depending on the objective of identifying members with specific conditions. A note of caution, however, is that the limits of performance vary considerably across conditions and may be different in other populations.

REFERENCES

- Bennett, B. M. 1972. "On Comparisons of Sensitivity, Specificity and Predictive Value of a Number of Diagnostic Procedures." *Biometrics* 28 (September): 793–800.
- Centers for Medicare and Medicaid Services. 2002a. "Medicare Managed Care Manual: Chapter 5—Quality Assessment" [accessed June 4, 2003]. Available at http://cms.hhs.gov/manuals/116_mmc/mc86c05.asp.
- Centers for Medicare and Medicaid Services. 2002b. "The New Risk Adjustment Model" [accessed June 4, 2003]. Available at <http://www.cms.gov/healthplans/riskadj/FactSheet.asp>.
- Fowles, J. B., A. G. Lawthers, J. P. Weiner, D. W. Garnick, D. S. Petrie, and R. H. Palmer. 1995. "Agreement between Physicians' Office Records and Medicare Part B Claims Data." *Health Care Financing Review* 16 (4): 189–99.
- Fowles, J., E. Fowler, and C. Craft. 1998. "Validation of Claims Diagnoses and Self-Reported Conditions Compared with Medical Records for Selected Chronic Diseases." *Journal of Ambulatory Care Management* 21 (1): 24–34.
- Hebert, P. L., L. S. Geiss, E. F. Tierney, M. M. Engelgau, B. P. Yawn, and A. M. McBean. 1999. "Identifying Persons with Diabetes Using Medicare Claims Data." *American Journal of Medical Quality* 14 (6): 270–7.
- Horner, R. D., J. A. Paris, J. R. Purvis, and F. H. Lawler. 1991. "Accuracy of Patient Encounter and Billing Information in Ambulatory Care." *Journal of Family Practice* 33 (6): 593–8.
- Martin, L. M., M. Leff, N. Calonge, C. Garrett, and D. E. Nelson. 2000. "Validation of Self-Reported Chronic Conditions and Health Services in a Managed Care Population." *American Journal of Preventative Medicine* 18 (3): 215–8.

- McNeil, B. J., E. Keeler, and S. J. Adelstein. 1975. "Primer on Certain Elements on Medical Decision Making." *New England Journal of Medicine* 293 (5): 211–5.
- National Committee for Quality Assurance. 1999. *HEDIS® 2000 Technical Specifications*. Washington, DC: National Committee for Quality Assurance.
- Quam, L., L. Ellis, P. Venus, J. Clouse, C. Taylor, and S. Leatherman. 1993. "Using Claims Data for Epidemiologic Research: The Concordance of Claims-Based Criteria with the Medical Record and Patient Survey for Identifying a Hypertensive Population." *Medical Care* 31 (6): 498–507.
- Robinson, J. R., T. K. Young, L. L. Roos, and D. E. Gelskey. 1997. "Estimating the Burden of Disease Comparing Administrative Data and Self-Reports." *Medical Care* 35 (9): 932–47.
- Turner, C. F., T. K. Smith, L. K. Fitterman, T. Reilly, K. Pate, M. B. Witt, A. M. McBean, J. T. Lessler, and B. H. Forsyth. 1997. "The Quality of Health Data Obtained in a New Survey of Elderly Americans: A Validation Study of the Proposed Medicare Beneficiary Health Status Registry (MBHSR)." *Journal of Gerontology: Social Sciences* 52B (1): S49–58.
- Virnig, B. A., and M. McBean. 2001. "Administrative Data for Public Health Surveillance and Planning." *Annual Review Public Health* 22: 213–30.

Supplemental Appendix. Specificity and Sensitivity of Case-Finding Algorithms for Six Chronic Conditions.

Algorithm Criteria	Hypertension		Heart Failure		Chronic Lung Disease		Arthritis		Glaucoma		Diabetes	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
One 1999 claim with dx	0.83	0.69	0.58	0.93	0.59	0.89	0.43	0.87	0.68	0.95	0.90	0.93
One 1999 face-to-face encounter claim with dx	0.67	0.80	0.45	0.96	0.50	0.93	0.29	0.93	0.61	0.96	0.82	0.96
One 1999 face-to-face encounter claim with primary dx	0.62	0.83	0.42	0.97	0.44	0.95	0.23	0.95	0.58	0.96	0.72	0.98
Two 1999 claims with dx	0.72	0.80	0.47	0.95	0.48	0.94	0.28	0.94	0.57	0.98	0.85	0.96
Two 1999 face-to-face encounter claims with dx	0.48	0.90	0.36	0.98	0.38	0.97	0.16	0.97	0.50	0.98	0.70	0.98
Two 1999 face-to-face encounter claims with primary dx	0.39	0.93	0.29	0.99	0.30	0.98	0.12	0.98	0.43	0.99	0.57	0.99
One 1999 Rx claim	0.71	0.77	0.58	0.78	0.38	0.96	0.32	0.87	0.50	0.99	0.64	0.99
One 1999 claim with dx or a Rx claim	0.90	0.60	0.78	0.75	0.62	0.87	0.55	0.77	0.73	0.95	0.91	0.93
One 1999 face-to-face encounter claim with dx or Rx claim	0.86	0.67	0.71	0.77	0.56	0.91	0.46	0.82	0.69	0.96	0.87	0.96
One 1999 face-to-face encounter claim with primary dx or a Rx claim	0.84	0.68	0.69	0.77	0.54	0.93	0.43	0.83	0.68	0.96	0.82	0.97
Two 1999 claims with dx or a Rx claim	0.86	0.68	0.72	0.76	0.54	0.93	0.45	0.82	0.67	0.97	0.88	0.96
Two 1999 face-to-face encounter claims with dx or a Rx claim	0.80	0.73	0.67	0.77	0.50	0.95	0.38	0.85	0.65	0.98	0.82	0.98
Two 1999 face-to-face encounter claims with primary dx or a Rx claim	0.78	0.74	0.64	0.78	0.46	0.95	0.36	0.85	0.62	0.98	0.77	0.99
One 1999 claim with dx and a Rx claim	0.64	0.86	0.39	0.96	0.35	0.98	0.20	0.96	0.45	0.99	0.62	0.99
One 1999 face-to-face encounter claim with dx and Rx claim	0.52	0.91	0.32	0.97	0.32	0.98	0.15	0.98	0.42	0.99	0.58	1.00
One 1999 face-to-face encounter claim with primary dx and a Rx claim	0.48	0.92	0.30	0.98	0.28	0.98	0.12	0.98	0.40	0.99	0.53	1.00
Two 1999 claims with dx and a Rx claim	0.57	0.89	0.33	0.97	0.32	0.98	0.15	0.98	0.40	0.99	0.61	1.00

Two 1999 face-to-face encounter claims with dx and a Rx claim	0.39	0.95	0.26	0.98	0.27	0.99	0.10	0.99	0.36	0.99	0.52	1.00
Two 1999 face-to-face encounter claims with primary dx and a Rx claim	0.32	0.96	0.23	0.99	0.22	0.99	0.07	1.00	0.32	0.99	0.44	1.00
One 1999 or 2000 claim with dx	0.92	0.56	0.74	0.88	0.70	0.82	0.60	0.77	0.75	0.92	0.95	0.88
One 1999 or 2000 face-to-face encounter claim with dx	0.90	0.61	0.70	0.91	0.67	0.86	0.53	0.81	0.74	0.93	0.94	0.92
One 1999 or 2000 face-to-face encounter claim with primary dx	0.77	0.74	0.59	0.95	0.55	0.92	0.36	0.90	0.68	0.94	0.87	0.96
Two 1999 or 2000 claims with dx	0.87	0.67	0.67	0.92	0.61	0.89	0.45	0.87	0.69	0.96	0.93	0.93
Two 1999 or 2000 face-to-face encounter claims with dx	0.83	0.73	0.62	0.94	0.56	0.91	0.37	0.91	0.64	0.96	0.91	0.95
Two 1999 or 2000 face-to-face encounter claims with primary dx	0.61	0.85	0.49	0.97	0.45	0.96	0.20	0.96	0.54	0.97	0.77	0.98
One 1999 or 2000 Rx claim	0.79	0.70	0.67	0.72	0.46	0.94	0.42	0.79	0.55	0.98	0.72	0.99
One 1999 or 2000 claim with dx or a Rx claim	0.95	0.48	0.88	0.67	0.74	0.80	0.70	0.64	0.78	0.92	0.95	0.88
One 1999 or 2000 face-to-face encounter claim with dx or Rx claim	0.94	0.52	0.87	0.69	0.72	0.84	0.66	0.67	0.76	0.92	0.94	0.92
One 1999 or 2000 face-to-face encounter claim with primary dx or a Rx claim	0.92	0.58	0.82	0.70	0.65	0.89	0.57	0.73	0.74	0.94	0.91	0.95
Two 1999 or 2000 claims with dx or a Rx claim	0.93	0.56	0.85	0.69	0.66	0.86	0.62	0.71	0.75	0.95	0.94	0.93
Two 1999 or 2000 face-to-face encounter claims with dx or a Rx claim	0.92	0.60	0.82	0.70	0.63	0.88	0.57	0.74	0.72	0.96	0.93	0.95
Two 1999 or 2000 face-to-face encounter claims with primary dx or a Rx claim	0.88	0.64	0.78	0.71	0.58	0.92	0.49	0.77	0.68	0.96	0.87	0.97
One 1999 or 2000 claim with dx and a Rx claim	0.76	0.77	0.53	0.92	0.43	0.96	0.32	0.92	0.52	0.98	0.72	0.99
One 1999 or 2000 face-to-face encounter claim with dx and Rx claim	0.75	0.78	0.51	0.94	0.41	0.96	0.29	0.93	0.52	0.98	0.71	0.99
One 1999 or 2000 face-to-face encounter claim with primary dx and a Rx claim	0.64	0.85	0.44	0.96	0.36	0.97	0.21	0.96	0.49	0.98	0.67	0.99
Two 1999 or 2000 claims with dx and a Rx claim	0.73	0.81	0.49	0.94	0.42	0.97	0.26	0.95	0.49	0.98	0.71	0.99

Two 1999 or 2000 face-to-face encounter claims with dx and a Rx claim	0.70	0.83	0.47	0.96	0.39	0.97	0.22	0.96	0.47	0.99	0.70	0.99
Two 1999 or 2000 face-to-face encounter claims with primary dx and a Rx claim	0.69	0.83	0.46	0.96	0.39	0.97	0.21	0.96	0.46	0.99	0.70	0.99

Dx = diagnosis, Rx = prescription claim, Sens = sensitivity, Spec = specificity