DISEASE MANAGEMENT
Volume 10, Number 5, 2007

DOI: 10.1089/dis.2007.105713

On the Efficacy of Programmatic Searching of Medical Claims for the Occurrence of Hospital Admissions for Coronary Artery Disease

JAMES R. NESTOR, PH.D., AND JANET G. KNECHT, M.S., R.N.

ABSTRACT

This study was conducted in order to test the proposition that medical claim records, when searched electronically, can be reliably used to locate individual, disease-specific hospital admissions. For the study, admissions for coronary artery disease (CAD), self-reported by employer-sponsored recipients of chronic disease management (DM) services, were verified against physician-compiled medical records. Confirmed events were then subjected to electronic searching of the corresponding medical claim records using a variety of conditional requirements for included types of evidence. At maximum sensitivity (92.6%), the search algorithm positively identified 126 of 136 verified admissions while falsely identifying 1,025 others. At maximum specificity (98.7%), the algorithm positively identified 55 of 136 while falsely identifying 13. The maximum value of the true positive to false positive ratio was 4.47. The maximum Youden index value was obtained by requiring that the diagnostic intensity (proportion of event-related claims having a CAD-related diagnosis code) have a minimum value of 0.20. The study concluded that an admission search algorithm applied to typical commercial medical claims generated results that are unsatisfactory for the determination of admission incidence in the CAD population. While the methods may be sound, they fail to overcome the weaknesses of the searched data. (*Disease Management* 2007;10:293–303).

INTRODUCTION

Claim records as a research tool

COLLECTIONS OF ADMINISTRATIVE DATA such as medical insurance and pharmacy benefit claims have the potential to serve as primary information sources for clinical investigators interested in exploring—in significant patient volumes—questions of: utilization of services and drugs,[1,2,3,4] outcomes in relation to interventions,[5,6] disease incidence and prevalence,[7,8,9,10] frequency of explicit clinical events,[11] morbidity related to environmental or occupational exposure to toxic materials,[12,13] and identification of groups at elevated risk,[14] among others. Earlier investigators have properly noted the usual convenience, availability, and low cost of the claim records; and have properly described their usual limitations, including the lack of clinical detail, specificity, and completeness.

Researchers tend to be attracted to claim records because the information content extends to all persons within relatively large groups over a relatively long period, and because their format makes them easily search-

Cardium Health, Farmington, Connecticut.

able using simple computer programs. The published literature features numerous reports that (a) put forth a proposal for an appropriate investigation using claims, and (b) relate the results of a side-by-side comparison of the findings from claim records versus findings from conventional clinical records. Often in such comparisons the findings from claims have not reliably matched those from the more trusted clinical records.[9,10,15,16,17] In this work, we test the supposition that medical claim records can be reliably used to find and isolate discrete, heart-disease-related hospital admissions when searched by means of an electronic algorithm.

Medical claims and disease management

In its most common form, the population-level management of chronic disease by means of a disease management (DM) program is invariably organized on the basis of some type of health plan membership and consequently tends to be widely infiltrated with administrative data such as insurance claims. That information is used to identify potential program participants, establish some minimal baseline medical history for participants, and determine means of contact with participants and their doctors. It is also common practice to use claim data to assess DM program efficacy as it relates to utilization of services.

For some conditions, heart disease being an excellent example, the condition-specific hospital admission rate is a logical surrogate metric for aggregate, condition-specific service utilization. It is a characteristic feature of chronic heart disease that acute recurrences necessitating hospitalization with expensive procedures tend to punctuate long periods of lower level care, primarily pharmacotherapy and medical management. This pattern has the effect of isolating the cost of care to the acute episodes.

In principle, determination of hospital admission rate within a test population requires nothing more than the simple enumeration of admissions within a test time frame. A simple counting process applied to medical records, patient histories, or administrative records such as insurance claims will yield a result. As a practical matter, however, the legitimacy of this process is entirely contingent on both the quality and content of the records examined. Record sets that are improperly compiled, incomplete, or discontinuous lead immediately to erroneous results. Likewise, record sets that are subject to ambiguous interpretation or are insufficiently detailed are as problematic.

In this work, we formulate and test a method for counting hospital admissions that uses medical claims as the sole information source and an electronic search algorithm to search a database of claims for specific items of evidence that indicate admission.

BACKGROUND

The objective of the present work is to collect and rank empirical evidence of hospital admission from a generalized set of medical claim records and to test the diagnostic power of a generalized discovery algorithm based on the same empirical findings. Such a discovery algorithm has potential utility in the analysis of hospitalization incidence, utilization of services, access to care, and quality of medical services.

The generalized record set

Within this work, we define a generalized record set as one that (a) allows aggregation of multiple discrete data files without a requirement that they be identical in form and structure, (b) has a forced and invariant specification for each data field, and (c) can be populated by a process of extraction and transformation of discrete data files from any of the widely variant medical claim formats that are currently in general use by insurance claim payers in the United States. (The details of the processing are not provided in this work.) For the purposes of this work, we define a generalized discovery algorithm as one that operates upon a generalized record set.

Conventionally, claims for payment for medical services provided to insured individuals are collected into a variety of payer databases for processing and transmission. It is common practice among payers to extract and transfer to DM programs the claim records of specific groups of health plan members as a discrete electronic claim file. (This movement of electronic claim files has been the subject of federal privacy regulation since 2003.) Providers of employee benefit services such as DM are routinely supplied with discrete files containing the processed claims of all health plan members who are eligible for services.

A typical claim file is organized into rows and columns or fields. Each row is associated with a single item of medical service and each field is assigned to a specific datum associated with the service. For the most part, these discrete files have been put in place without standardization of structure or of content. This is particularly true in relation to both the number and the substance of the included fields. As a consequence, a collection of discrete files received from various sources will exhibit vacancies due to poor overlap of fields. For quantitative purposes, the term *occupancy* is used here to mean the proportion of rows in a given column that contain a legitimate, interpretable value.

Evidence of hospital admission in the generalized record set

Despite the lack of institutionalized standardization, the claim-level information content of most discrete claim record sets can be readily categorized. The typical record can be relied upon to contain information related to: (1) the identity of the payer, (2) the administrative status of the claim, (3) the identity of the service provider(s), (4) the identity and terms of coverage of the insured individual, (5) the identity of the patient, and (6) the particular details of the provided services (ie, the diagnostic and procedure codes, the place and type of service, the type of provider, the dates and associated cost of the services). For the enumeration of hospital admissions for specific cause, all but the last 2 of these categories may be ignored.

METHODS

Setting

Our study population was drawn from a group of individuals who were voluntarily enrolled in a commercial heart disease DM program for any period of time within the 4 calendar year interval of 2002 through 2005. The DM program was provided under the sponsorship of a group of medium-sized US corporations, each acting independently, and made open and freely available to each of their own benefits-eligible adult employees and adult dependents with existing heart disease. All DM program participants were continuously engaged in regularly scheduled one-on-one telephonic sessions with a cardiac nurse or other clinical specialist. The content of the counseling sessions was used to form an ongoing self-reported medical history for each participant.

Investigational plan

The process for selection of individuals, admission events, and claim sets for inclusion in the study is shown in Fig. 1. In order to quantify the diagnostic power of medical claims in relation to hospitalization events, a cohort of individuals was identified for whom (a) there existed a self-report of heart disease-related hospitalization within the study period, and (b) verification of the self-reported event was independently obtained from the individual's physician. For the purposes of this investigation, this group of self-reported, verified events was treated as wholly authentic and labeled as true—the true event set.

Subsequently, for each true event the corresponding medical claim records were inspected visually for evidence of the event occurrence, and the amount of evidence was quantified using a simple scoring scheme. These findings were then used to construct an electronic search engine designed specifically to locate and isolate the clusters of claims that are characteristically associated with hospitalization. The diagnostic power of this search engine was then evaluated by challenging it to detect the true events within the identical, genuine claim records.

A DM case was considered for inclusion in the study if the following conditions were satisfied:

(1) A history of coronary artery disease (CAD) was asserted by the participating individual upon enrollment in the DM program

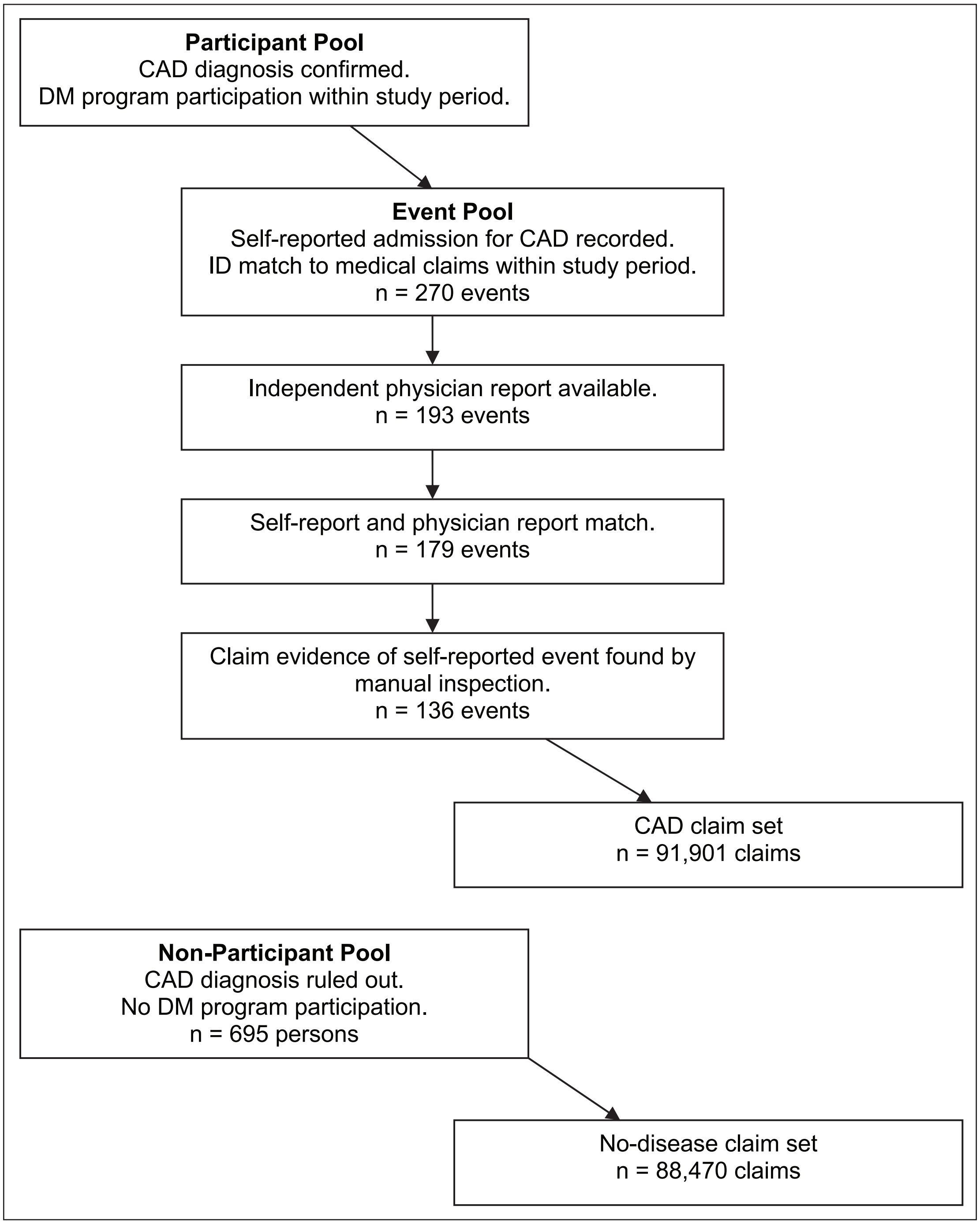

FIG. 1. Selection process.

and a firm diagnosis of existing CAD was subsequently confirmed. Confirmation was obtained either from a brief written medical history supplied to the DM program by the individual's physician or from a telephonic assessment interview with the individual conducted by a nurse with training in cardiac care.

(2) Evidence existed that the individual had experienced and self-reported 1 or more CAD-related hospital admissions within the study time frame. Such evidence was obtained as a part of the routine, ongoing telephonic communications that constituted the primary clinical activity of the DM program.

(3) It had been verified during the study period that the participating individual had been

a covered member of a participating health plan for which comprehensive electronic medical claim records were available for searching. Continuous participant-level health plan membership throughout the study period was not determinable and therefore not required for inclusion.

(4) It had been verified that the participating individual had filed at least 1 medical claim within the study period, and was thereby represented, at least for some period of time, in the claim set.

True event set

The sample group of hospitalization events investigated in this study was assembled by searching the included DM cases for instances of hospitalization for CAD as reported by participants. An event was included if the description of the event made reference to any of these terms: acute myocardial infarction; chest pain; unstable angina; acute coronary syndrome; revascularization as coronary artery bypass graft surgery (CABG), angioplasty (PTCA) or percutaneous coronary intervention (PCI); implant of a defibrillation device (AICD); or inpatient death attributed to CAD. Each included event was entered with its approximate date of occurrence, recorded as month and year. The search was continued until an acceptable volume had been identified.

For each event in the sample, an independent search was made of records that had been accumulated over the course of the study period and supplied, on behalf of the program participants, by their respective physicians. When such records were lacking or inadequate, an attempt was made to establish the physician-reported history by way of telephone request, made by a cardiac nurse, to each physician office

Each event in the sample for which a matching reference was found in the physician-reported history was then accepted as a member of the true event set. An event was considered to be matched if (a) it was described or classified as CAD-related by both parties using 1 or more of the terms listed above, and (b) the dates of occurrence reported by the 2 parties differed by no more than 2 months. It was not required of a matched pair, however, that both parties had used identical terms to describe or classify the event. Events were reported and recorded by both participants and physicians using the English language terms mentioned here; no diagnosis codes were used.

Having established their independent confirmation, the true events were thereupon considered to have proven existence.

Medical claims

All of the medical claim records used in this investigation were provided independently to the DM program by the commercial insurance carriers and health plan providers covering the study population. This group of carriers and plans consisted of organizations of various types, sizes, and locations within the United States and as such constituted a limited but somewhat representative sample of the industry that services persons who receive health care benefits through employment with a US corporation. None of the data used here were obtained from any US or state government entity.

Manual claim search for evidence of true events

A manual search of medical claims was undertaken to locate evidence of each true event. The search process involved the isolation of all claims associated with each individual in the true event set, sorting the claims according to the date of service and visually inspecting each row in the temporal vicinity (±30 days) of the reported event date for evidence of admission and CAD involvement. A qualified item of evidence was taken as any of these 8 types: (1) *multiple claims* in close temporal proximity, (2) relatively *large dollar amounts* in relatively short periods, (3) *diagnosis codes* corresponding to CAD, (4) *procedure code*(*s*) corresponding to CAD therapies, (5) *discharge code*(*s*) indicating hospitalization, (6) *place-of-service* code(s) indicating inpatient services, (7) *type-of-service* entry(ies) indicating inpatient services, and (8) provider-type entry(ies) indicating inpatient or surgical services. In order to rank the true events according to the strength of their corresponding claim evidence, a simple scoring procedure was applied in which each event was

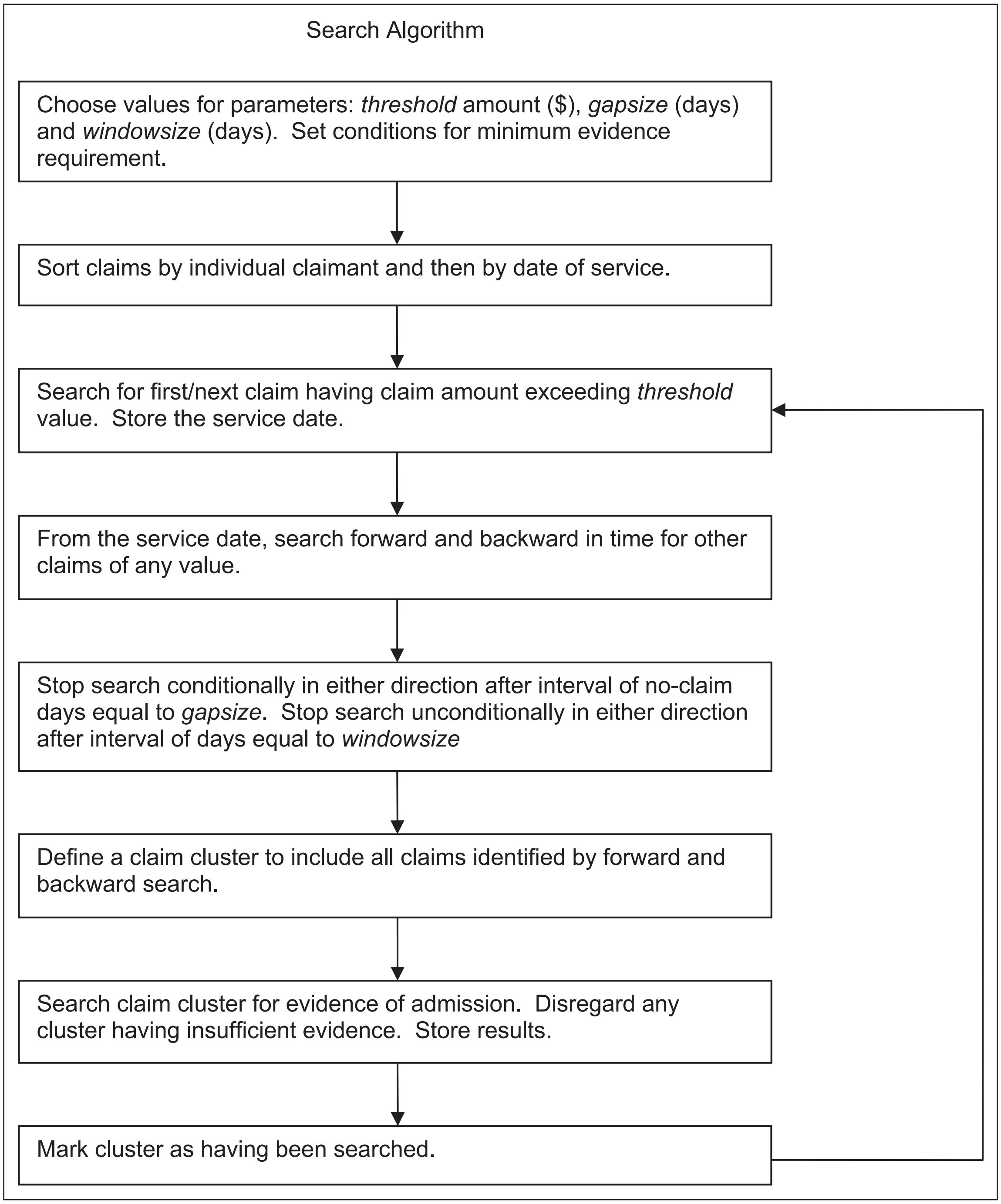

FIG. 2. Search algorithm.

assigned 1 point for each item of evidence discovered within each claim cluster.

Programmatic algorithm

Having fully characterized each true event with regard to its substantiation in the claim records, we then devised a programmatic search algorithm (Fig. 2) that was instructed to locate hospitalization events in the identical claim records without having prior knowledge of their existence.[a] The algorithm was further

[a] The program code used in this work is non-proprietary and may be obtained from the corresponding author on request.

instructed to search for any of the items of evidence developed from the above manual inspection. In order to test the ability of the algorithm to reject non-CAD events, an additional set of medical claims was drawn from DM program participants who had been determined to be free of CAD and to have no prior history of hospitalization for CAD.

RESULTS

True events

The search of participant records for self-reported admissions for CAD yielded a total of 193 events for which substantiation, either positive or negative, had been provided by the physician holding the relevant medical history. Positive agreement as to event type and event date was found for 179 of the 193 events when the test for agreement was able to be applied (Table 1). Here, substantial agreement means that the physician record contained an entry describing a hospital admission with a specific CAD cause and a date approximately equivalent to that self-reported by the participant. These 179 events were assigned to the true event set. The physician report and the self-report were in substantial conflict for 14 of the self-reported events in this category and so were not included in the true event set. On this basis, the positive predictive value of the self-reported hospitalization event is found to be equal to 93%.

Evidence of admission

For the group of 179 true events, corresponding evidence from medical claims of hospital admission for CAD was detectable by manual inspection for 136 events. In 43 instances, none of the 8 indicators of hospital admission listed earlier (ie, multiple claims, large dollar amounts, CAD diagnosis codes, CAD procedure codes, discharge codes, inpatient place of service codes, inpatient type of service codes, and inpatient provider type codes) registered in any of the claims, as a likely consequence of discontinuous insurance coverage. These cases were not considered further.

TABLE 1. SELF-REPORTED CAD ADMISSIONS WITH PHYSICIAN SUBSTANTIATION

Number of self-reported events in sample	193
Number of persons reporting 1 or more events	159
Number of denied events	14
Number of true events	179
Positive predictive value of self-report	93%

CAD, coronary artery disease

TABLE 2. DISTRIBUTION OF STRENGTH OF EVIDENCE SCORES

Score	*Events with this score*	*Percentage of Events at or above this score*
8	63	46.3%
7	18	59.6%
6	12	68.4%
5	10	75.7%
4	5	79.4%
3	21	94.9%
2	5	98.5%
1	2	100.0%

Each member of the set of true, detectable events (TD event set) was given a point score between 0 and 8, with 1 point assigned for each item of evidence discovered. The mean score was 6.2; the maximum score was obtained for 63 of 136 (46.3%) events; and 103 of 136 (75.7%) had a score of 5 or greater. The complete distribution of scores is displayed in Table 2.

The relative reliability of the 8 evidence types was ranked (Table 3) on the basis of the frequency with which each individually contributed to the scores in Table 2. The table also shows the occupancy factors for 7 of the 8 evidence fields and suggests an association between reliability and occupancy. The 7 fields for which occupancy factors are given are those that would be expected to be completely occupied in an idealized claim set. They carry data that are at least knowable for any claim but that tend to be omitted by less diligent record keepers. In contrast, discharge codes are only meaningful for inpatient hospital services and are properly unfilled otherwise.

Test of search algorithm

The programmatic algorithm described earlier was instructed to search the set of medical

Table 3. Reliability of Evidence by Type

Type	*Frequency*	*Occupancy factor*
Claim count	136	1.00
Diagnosis code	126	1.00
Claim amount	128	1.00
Place of service	96	0.76
Procedure code	95	0.56
Type of service	91	0.62
Discharge code	91	n/a
Provider type	84	0.49

claims associated with those persons having the TD events. The searched claim set comprised a total of 91,901 claim records involving 119 individuals. The algorithm was tested using a wide range for each of the available program variables. The best results, in terms of predictive power, were obtained by setting the threshold amount ($1,600), the gapsize (10 days), and the windowsize (30 days). As a further condition for inclusion, a discovered event was required to be represented by a cluster of discrete records numbering at least 4; in every case, a smaller claim cluster count resulted in an unacceptably large number of false positive results. For each individual in the TD group, any algorithmic event having a temporal correspondence (±30 days) with a TD event was classified as a discovered match. A total of 126 events were matched to 136 TD events. Ten events were classified as undiscovered on the basis of low *threshold amount* or cluster count.

In order to assess the ability of the algorithm to reject nonevents or non-CAD events, it was applied, using identical search parameters, to a separate claim set of 88,470 records of 458 persons with no history of heart disease. The 2 claim sets were otherwise matched as to their time frames and origins. Any admission events discovered by the algorithm in this no-disease claim set were classified as false positive findings.

The diagnostic power of the search method was evaluated in detail using methods conventionally applied to diagnostic tests. Here, correctly identified true events and incorrectly rejected nonevents were treated as true and false positive test results, respectively, and the test results were measured over a wide range of evidence requirement constraints. Table 4 presents the list of tested constraints.

Table 5 shows the results of the progressive, ordered, cumulative application of the full set of constraints to the baseline algorithm findings. For each step of the progression, the table provides the number of true positive test results (actual value = 136), the number of false positives (actual value = 0), the number of all positive tests, the diagnostic sensitivity, the diagnostic specificity, the Youden index,[19] the positive predictive value, and the simple ratio of true positive to false positive results.

The same constraints were alternatively applied in discrete logical groups as shown in Table 6. The groups were selected to show the influence of evidence according to its characterization as primarily diagnostic (ICD and PROC) rather than primarily locational (POS, TOS, DISC and PROV). In addition, these selections are intended to show the strong influence of the diagnostic intensity indicator (INT).

Given the test devised for this study, an ideal detection algorithm would identify 136 events, each corresponding to 1 member of the true

Table 4. Algorithmic Constraints Available for Test

Evidence type	*Requirement*	*Symbol*
Diagnostic intensity	Proportion of claims having CAD diagnosis code 0.2	INT
Diagnosis code	One or more diagnosis codes specific to CAD	ICD
Place of service	One or more codes indicating inpatient services	POS
Procedure code	One or more procedure codes specific to CAD	PROC
Type of service	One or more codes indicating inpatient or surgical services	TOS
Discharge code	One or more codes indicating discharge status	DIS
Provider type	One or more codes indicating hospital as provider	PROV

CAD, coronary artery disease

TABLE 5. PROGRESSIVE APPLICATION OF EVIDENCE CONSTRAINTS

Constraints	*TP*	*FP*	*All positive tests*	*Sensitivity*	*Specificity*	*Youden index**	*PPV*	*TP/FP*
Baseline events ($16,000, >3 claims)	126	1025	1151	0.926	0.000	−0.074	0.109	0.123
INT	111	63	174	0.816	0.939	0.755	0.638	1.762
INT & POS	78	27	105	0.574	0.974	0.548	0.743	2.889
INT & POS & PROC	76	21	97	0.559	0.980	0.539	0.784	3.619
INT & POS & PROC & TOS	67	15	82	0.493	0.985	0.478	0.817	4.467
INT & POS & PROC & TOS & DISC	63	15	78	0.463	0.985	0.448	0.808	4.200
INT & POS & PROC & TOS & DISC & PROV	55	13	68	0.404	0.987	0.391	0.809	4.231

TP, true positive results; FP, false positive results; PPV, positive predictive value, TP/(TP + FP).
*sensitivity + specificity-1

event set. Over the range of conditions tested, the actual algorithm detected true events numbering between 55 and 126, while failing to reject non-true events numbering between 13 and 1,025. At the point of greatest sensitivity, the false positive count is an order of magnitude greater than the true positive count. At the minimum sensitivity, a significant false positive proportion persists. The ratio of true results to false results reached a maximum value of 4.47.

The receiver operator characteristic curve (not shown) constructed for the combined results of Tables 5 and 6 was found to be discontinuous and unsuitable for ROC area analysis. Instead, the Youden index (sensitivity + specificity −1) was calculated for each conditional requirement in order to locate a maximum and identify the optimal point at which to dichotomize the algorithm.[18] By this measure, maximum performance is obtained using only the 2 baseline conditions for claim count and threshold dollar value together with a minimum value of 0.20 for diagnostic intensity. This third condition is fulfilled when at least 20% of the claims associated with a detected admission event contain a diagnostic code for CAD.

DISCUSSION

The methods tested here have marginal diagnostic power and produce results that have minimal practical value. If used at high sensi-

TABLE 6. LOGICALLY GROUPED APPLICATION OF EVIDENCE CONSTRAINTS

Constraints	*TP*	*FP*	*All positive tests*	*Sensitivity*	*Specificity*	*Youden index**	*PPV*	*TP/FP*
ICD or PROC	121	283	404	0.890	0.724	0.614	0.300	0.428
POS or TOS or DISC or PROV	105	807	912	0.772	0.213	−0.015	0.115	0.130
(ICD or PROC) & (POS or TOS or DISC or PROV)	101	263	364	0.743	0.743	0.486	0.277	0.384
INT & (ICD or PROC)	111	63	174	0.816	0.939	0.755	0.638	1.762
INT & (POS or TOS or DISC or PROV)	93	53	146	0.684	0.948	0.632	0.637	1.755

TP, true positive results; FP, false positive results; PPV, positive predictive value, TP/(TP + FP).
*sensitivity + specificity-1

tivity, the tested algorithm would be useful only for the intermediate purpose of isolating possible events, mostly false positives, into a subset of records suitable for further study and analysis. Conversely, under conditions of maximal specificity, the algorithm can reduce the false positive rate substantially, but it also then rejects all true results except those for which the evidence is both broad and powerful. At no point of operation is there an acceptable correspondence between the count of all positive tests and the count of all true events, thus minimizing any potential utility for use of this algorithm as a tool suitable for the simple enumeration of admission rates.

Much of the lack of correspondence between actual and claims-determined admission counts can be attributed to: (a) the structural omissions of the claim records that allow the noninclusion of potentially valuable evidence and (b) common errors in data insertion and management that tend to accumulate and propagate without detection. An illustration of the former is included in the earlier discussion related to claim occupancy. The best illustration of the latter is that related to identity error, as a consequence of which the algorithm is directed to search the wrong records entirely. An error of this type can account for the circumstance in which all, or nearly all, claim evidence of a known event seems absent.

In the present work, claims provided by 10 independent payers were aggregated and reformatted to make them uniformly searchable. Findings more favorable than those shown are potentially achievable by selective elimination of lower quality claim sets, or by treating each as uniquely searchable and aggregating all of the individual search results.

It is worth noting that the methods of this work have been applied to administrative records created for the purpose of facilitating, regulating, and recording payments made for completed medical services in the employer-sponsored health care setting in the United States. These particular records, used primarily by commercial claim payers, have been found to serve unconvincingly, here and elsewhere, as complementary documentation to formal medical records compiled by physicians and other clinicians on the basis of actual encounter experience. These commercial records lack much of the valuable detail that is often contained in the records of noncommercial payers. Researchers using data from the US Medicare MEDPAR database, for example, would have access to additional data elements that are reliable and specific markers for disease-specific admissions (DRG codes), and other helpful items such as admission and discharge dates. Therefore, we speculate that the type of algorithm put forth here is likely to function more acceptably in that type of data environment.

Nevertheless, in relation to CAD at least, the use of MEDPAR claims has not consistently tested favorably when utilized for purposes that demand finer clinical detail than can be provided from an administrative record alone. For example, claim-based methods for following outcomes of patients undergoing invasive coronary artery procedures have shown only marginal correlation with more conventional clinical record methods.[15,16,17] On the other hand, given a less challenging test, Medicare claims showed good positive predictive value in the detection of diagnosis of acute myocardial infarction.[11]

The methods developed and tested here are not adequately reliable as a means of identifying discrete hospital admissions for CAD among a group of employees and adult dependents. By inference, we conclude that application of these methods to the purposes of analyzing admission incidence, utilization of services, or cost of care for the same group of individuals would likewise be unreliable due to the unacceptable combination of sensitivity and specificity. Such inference would extend to application of such methods to the evaluation of DM efficacy.

While the methods described here have not been shown to be generally useful, at least 1 of the conceptual elements of the study may have valuable implications. The significance of the quantity referred to here as diagnostic intensity was not anticipated. We are not aware of its use or introduction prior to this work. Nonetheless, the value of its inclusion as an indicator of CAD admission is evident from the results shown in Tables 5 and 6. As evidence of admission, it surpasses all the available administrative codes.

There is reason to anticipate that the utility of this indicator, used in the general disease-specific sense, should extend beyond the scope of this work. In such a realm of investigation, a researcher attempting to locate admissions of one type or another would require only that the claims include those few data elements that contain identifier, service date, diagnosis codes, and paid amount. No additional information would lead to a better result.

REFERENCES

1. Tyree PT, Lind BK, Lafferty WE. Challenges of using medical insurance claims data for utilization analysis. *Am J Med Qual*. 2006;21:269-275.
2. Schloss EP, Lewis V, Martin N, Pitts M. Estimating pediatric primary care provider visits in a capitated environment: encounter vs. claims databases. *Manage Care*. 2002;11:43-46.
3. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323-337.
4. Peele PB, Xu Y, Kupfer DJ. Insurance expenditures on bipolar disorder: clinical and parity implications. *Am J Psychiatry*. 2003;160:1286-1290.
5. Coleman AL, Morgenstern H. Use of insurance claims databases to evaluate the outcomes of ophthalmic surgery. *Surv Ophthalmol*. 1997;42:271-278.
6. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes. Mortality and reoperation following prostatectomy. *JAMA*. 1987;257:933-936.
7. Andrade SE, Graham DJ, Staffa JA, et al. Health plan administrative databases can efficiently identify serious myopathy and rhabdomyolysis. *J Clin Epidemiol*. 2005;58:171-174.
8. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Med Care*. 2002;40(suppl):IV-26-35.
9. Gedeborg R, Furebring M, Michaelsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. *J Clin Epidemiol*. 2007;60:155-162.
10. Kolodner K, Lipton RB, Lafata JE, et al. Pharmacy and medical claims data identified migraine sufferers with high specificity but modest sensitivity. *J Clin Epidemiol*. 2004;57:962-972.
11. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004;148:99-104.
12. Reeve GR, Stout AW, Hands D, Curry E. Inpatient hospital admission rates for nonmalignant respiratory disease among workers exposed to metal removal fluids at a U.S. automobile manufacturer. *Appl Occup Environ Hyg*. 2003;18:930-938.
13. Dominici F, Peng RD, Bell ML, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA.* 2006;295:11271134.
14. Ahmed F, Janes GR, Baron R, Latts LM, Kauffman L, Hothem M. Preferred provider organization claims showed high predictive value but missed substantial proportion of adults with high-risk conditions. *J Clin Epidemiol*. 2005;58:624-628.
15. Weintraub WS, Deaton C, Shaw L, et al. Can cardiovascular clinical characteristics be identified and outcome models be developed from an in-patient claims database? *Am J Cardiol*. 1999;84:166-169.
16. Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg*. 2005;129:1309-1317.
17. Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. *Ann Intern Med*. 1993;119:844-850.
18. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163:670-675.

Address reprint requests to:
James R. Nestor, Ph.D.
Chief Scientific Officer
Cardium Health
20 Batterson Park Road
Farmington, CT 06030

E-mail: jnestor@cardiumhealth.com