

## Accuracy of Computerized Outpatient Diagnoses in a Veterans Affairs General Medicine Clinic

Herbert C. Szeto, MD, MPH; Robert K. Coleman, PharmD; Parisa Gholami, MPH; Brian B. Hoffman, MD; and Mary K. Goldstein, MD, MS

**Background:** Electronically available data, both *administrative*, such as outpatient encounter diagnostic data, and *clinical*, such as problem lists, are being used increasingly for outcome and quality assessment, risk adjustment, and clinical reminder systems.

**Objective:** To determine the accuracy of outpatient primary care diagnostic information recorded in administrative and clinical files in a Veterans Affairs VISTA (Veterans Health Information Systems and Technology Architecture) database compared with medical chart notes.

**Study Design:** Cross-sectional medical chart review of 148 patients attending a general medicine clinic at a university-affiliated Veterans Affairs hospital for 9 diagnoses relevant to the choice of drug therapy for hypertension.

**Patients and Methods:** An administrative file of encounter diagnoses, for a 2-year period, and a clinical file of the problem list maintained by the clinician were the sources of electronic diagnoses. We compared these sources with diagnoses abstracted by medical chart review. We estimated the sensitivity and specificity of each electronic data source for detecting medical chart note diagnoses.

**Results:** The sensitivity for 8 of the 9 study diagnoses was greater than 80% in the administrative file and 49% in the clinical problem list. The specificity was good for the administrative file (91% to 100%) and even better for the clinical file (98% to 100%).

**Conclusions:** Outpatient encounter diagnoses relevant to hypertension recorded as electronic data had high specificity, and some codes had high sensitivity when collected over multiple visits. The administrative file was more sensitive but less specific than the clinical file. Administrative vs clinical files can be selected to minimize either the false-negative or the false-positive designations, respectively, as dictated by the needs of the quality assessment review.

(*Am J Manag Care* 2002;8:37-43)

researchers have incorporated outpatient data into predictive models for the ambulatory setting.<sup>5-8</sup> Outpatient data can be used as an adjunct to inpatient data to improve its quality and predictive power; more important, outpatient data apply to the large group of patients who have not been hospitalized. Such data are increasingly being incorporated into clinical support systems in the form of automated clinical reminders or treatment recommendations. Although the accuracy of diagnostic codes has been studied in hospitalized patients, in both non-veterans<sup>9</sup> and veterans,<sup>10-12</sup> less is known about outpatient administrative data. One large veterans study<sup>11</sup> examined outpatient clinic codes but was conducted before outpatient diagnostic data were available.

Quality assessment programs typically use abstraction of clinic medical charts as the standard for evaluation. We posed the following question: How well do computerized diagnoses, from administrative data files or from clinical data files, predict medical chart note diagnoses?

To address this question, we evaluated the sensitivity, specificity, positive predictive value (PPV),

---

From the Department of Veterans Affairs, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA (HCS, RKC, PG, BBH, MKG); and the Department of Medicine, Stanford University School of Medicine, Stanford, CA (HCS, BBH, MKG). Dr Szeto is now with Kaiser Permanente, Redwood City, CA.

Presented in abstract form at the Department of Veterans Affairs Health Services Research and Development Annual Meeting, Washington, DC, February 25, 1999.

This project was supported by Veterans Affairs Ambulatory Care Fellowship (HCS), Veterans Affairs HSR&D Career Development Award RCD-96-301 (MKG), and Veterans Affairs HSR&D grant CPG 97-006.

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Address correspondence to: Mary K. Goldstein, MD, MS, GRECC (182B), VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304.

Administrative data are being used increasingly to profile providers, to adjust risk, and to assess outcomes in large healthcare systems, including the Department of Veterans Affairs (VA),<sup>1-5</sup> particularly in the inpatient setting. More recently,

and negative predictive value (NPV) of computerized diagnostic data. The application of sensitivity/specificity in the evaluation of the accuracy of computerized data files provides information not found in traditional  $\kappa$  statistics and has been used in 2 major recent validation studies of inpatient diagnoses.<sup>13,14</sup> Sensitivity is a measure of the likelihood that the database will contain the particular diagnosis if it is present in the text of the medical chart note, and specificity is the measure of the likelihood that a diagnosis will not be listed in the database if it is absent from the medical chart note. A data file with greater sensitivity and less specificity, or vice versa, may be preferred depending on the intended use of the data. For instance, angiotensin-converting enzyme inhibitors are now recommended for patients who have both hypertension and diabetes. If one were trying to identify all patients who may potentially benefit from this intervention, one may want to choose the data file that was most sensitive for these diagnoses. One could then send reminders to their providers. On the other hand, if one were using this information to evaluate providers on their compliance with this recommendation, one may want to choose the file that was most specific, thereby giving providers the benefit of any doubt.

We designed our study to assess 2 forms of computerized diagnostic data: data from administrative files (visit encounter diagnoses and hospital discharge diagnoses lists) and data from a clinical file (the problem list). We expected clinical files to be more specific and less sensitive than administrative files because diagnoses in the clinical file are screened and updated by the primary care provider. For example, if a patient presents to the emergency department with chest pain and is admitted to the hospital to exclude myocardial infarction, the emergency department physician may code coronary artery disease as the reason for the visit. Even if the patient is later found to have noncardiac chest pain, the administrative data files would maintain the *International Classification of Diseases, Ninth Revision (ICD-9)*, code for coronary artery disease. The problem list maintained by the primary care provider, in contrast, would be less likely to reflect this error (and therefore, be more specific) because the primary care provider would not enter the diagnosis on the problem list until it was more certain, or would remove it from the list if it were later excluded. But because the problem list must be actively managed by the primary care provider, it is also more likely to have omissions, especially for conditions that may be managed by other providers,

thereby making it less sensitive compared with administrative files. We selected diagnoses related to drug choice for treatment of hypertension as the focus of study because of the high prevalence of hypertension and its importance as a target of guideline implementation and quality assessment.

---

... METHODS ...

This study was a cross-sectional medical chart review of patients attending a general medicine clinic at the Palo Alto, CA, campus of the VA Palo Alto Health Care System. The clinic provided care to approximately 7000 patients during the study period and had approximately 23,000 patient encounters annually. The clinic was staffed by 5 nurse practitioners, 5 attending physicians, and 16 internal medicine residents.

A half-day clinic module was chosen by convenience for each physician or nurse practitioner with a primary care clinic the week of May 18 to 22, 1998. The study group was composed of patients who were scheduled to attend those half-day clinic modules (index visit). New patients were excluded. For the 148 eligible patients, all clinic medical charts were located and reviewed for 9 diagnoses that potentially affect the drug choice for treatment of hypertension: atrial fibrillation, benign prostatic hyperplasia, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, diabetes, gastroesophageal reflux disease, gout, and hyperlipidemia.

Electronic medical record systems typically include both *administrative* data, such as appointment schedules and billing information, and *clinical* data, such as laboratory, pharmacy, and diagnostic data. There is not a clear consensus on terminology for this rapidly evolving technology, but general usage would regard records related to billing information as *administrative* and records related to clinical management as *clinical*. In accord with this usage, we regard diagnoses listed in encounter form billing records for each visit as *administrative* and diagnoses listed by the clinician on the problem list as *clinical*. VISTA (Veterans Health Information Systems and Technology Architecture) is the information system used by the VA.<sup>15</sup> There are multiple files contained within VISTA, including those with patient demographics, diagnostic codes, pharmacy data, and laboratory data. Within VISTA, we extracted diagnostic information from 3 separate files, 2 administrative and 1 clinical, for comparison: (1)

the Veterans Purpose of Visit (VPOV) file, (2) the patient treatment (PTF) file, and (3) the problem file.\* The VPOV and PTF files are considered administrative files because information contained in them is generally not used by clinicians. These files contain ICD-9 codes similar to those collected by Medicare for billing and other administrative purposes. The VPOV file contains ICD-9 codes for outpatient encounters, whereas the PTF file contains ICD-9 codes for all hospital discharges. The search of the VPOV file was limited to diagnostic codes contained in VISTA from October 1, 1996 (the start date for collection of outpatient diagnoses), through May 31, 1998 (see Table 1 for a list of ICD-9 codes). We reviewed the PTF file for hospital admissions from January 1, 1990, through May 31, 1998. Forty (27%) of 148 patients had previous hospital admissions. The problem list file is a clinical file, with problem lists generated by clinicians either by instructions to clinic clerks or by the clinician directly managing the diagnoses on the problem list. Computerized problem lists have replaced paper problem lists for all medical charts and constitute clinical data in computerized form.

One reviewer (HCS) abstracted the general medicine clinic medical chart notes inclusive of the index visit, recording the presence or absence of each of the 9 diagnoses. A second reviewer (PG) then reabstracted all 148 medical charts. The reviewers' results were compared with each other, and discrepancies were reviewed by a third reviewer

(MKG), and then again with the first and second reviewers. All 3 reviewers reached consensus on all medical charts. The VISTA diagnostic codes were then compared with those from the medical chart review.

The presence or absence of each diagnosis was considered for the entire period rather than at each individual encounter. Hence, a patient was considered to have hypertension as a medical chart diagnosis if he was noted to have hypertension at any visit covered by the medical chart review. Similarly, the VPOV file was considered to contain the diagnosis if it contained the ICD-9 code at any encounter during the 20-month review period. We then estimated the sensitivity and specificity of the VISTA diagnoses, using the medical chart as the standard for comparison. True positives were defined as diagnoses found in both VISTA and the medical chart. False positives were diagnoses found in VISTA but not in the medical chart. True negatives were diagnoses found in neither VISTA nor the medical chart,

**Table 1.** International Classification of Diseases, Ninth Revision (ICD-9), Codes

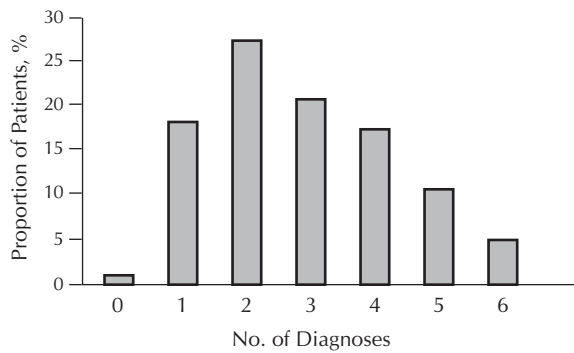
Diagnosis	ICD-9 Codes
Atrial fibrillation	427.31
Benign prostatic hyperplasia	600
Coronary artery disease	414.0, 414.05, 414.04, 414.03, 414.01, 414.8, 414.02
Congestive heart failure	428.1, 428.0, 428.9
Chronic obstructive pulmonary disease	491.0, 492.8, 492.0, 493.0, 493.1, 493.2, 493.9, 491.8, 491.1, 491.2, 491.9
Diabetes	250.1, 250.0, 250.61, 250.91, 250.81, 250.71, 250.51, 250.41, 250.31, 250.11, 250.10, 250.40, 250.21, 250.42, 250.93, 250.92, 250.83, 250.82, 250.73, 250.72, 250.63, 250.62, 250.53, 250.43, 250.20, 250.33, 250.32, 250.70, 250.30, 250.52, 250.60, 250.23, 250.80, 250.90, 250.22, 250.02, 250.50, 250.03, 250.12, 250.13
Gastroesophageal reflux disease	530.8, 530.1
Hypertension	401.9, 401.1
Hyperlipidemia	272.0, 272.1, 272.2, 272.3, 272.4

\*The VPOV file, one of the "V" (visit) files, is the first site of entry for outpatient encounter data in the VISTA system. The minimum dataset mandated by Veterans Health Administration is fulfilled through entries in the V files. The local PTF file contains a pointer to the Patient file. In contrast to the V files, the Patient file is not mandated nationally, although individual facilities may mandate its use. The local VISTA entries form the ceiling for completeness of computer diagnoses in the national VA databases in Austin, TX, since it is these local entries that are submitted to the national VA database. The information entered into the VPOV file updates the Outpatient Encounter (409.68) and the Outpatient Diagnosis (409.43) files. These files are then extracted and put into the national Patient Care Database.

**Table 2.** Prevalence of the Study Diagnoses in Clinic Charts

	n (Total = 148)	Prevalence
Congestive heart failure	15	0.10
Atrial fibrillation	17	0.11
Chronic obstructive pulmonary disease	24	0.16
Benign prostatic hyperplasia	24	0.16
Gastroesophageal reflux disease	25	0.17
Coronary artery disease	37	0.25
Diabetes	37	0.25
Hyperlipidemia	62	0.42
Hypertension	93	0.63

**Figure.** Distribution of Study Diagnoses



and false negatives were diagnoses found only in the medical chart. Sensitivity, specificity, PPV, and NPV were determined for each diagnosis in the 2 outpatient data files. (Sensitivity = true positive/[true positive + false negative]; specificity = true negative/[true negative + false positive]; PPV = true positive/[true positive + false positive]; and NPV = true negative/[true negative + false positive].) The normal approximation for the binomial distribution was used as the test statistic to determine the level of significance for the difference in sensitivity, specificity, PPV, and NPV for each diagnosis between the VPOV and problem files. The Wilcoxon signed rank test was used to determine significance of differences between the files across the diagnoses as a group.<sup>16</sup>

... RESULTS ...

**Table 2** lists the study diagnoses in increasing order of prevalence as

**Table 3.** Sensitivity and Specificity of Study Diagnoses in Problem Lists vs Veteran Purpose of Visit (VPOV) File\*

	Sensitivity			Specificity		
	Problem List	VPOV	P	Problem List	VPOV	P
Atrial fibrillation	0.75	0.94	.06	1.00	0.99	.16
Benign prostatic hyperplasia	0.42	0.50	.29	0.98	0.99	.72
Coronary artery disease	0.49	0.89	<.01	1.00	0.98	.08
Congestive heart failure	0.73	0.87	.17	1.00	1.00	—
Chronic obstructive pulmonary disease	0.55	0.95	<.01	1.00	0.97	.02
Diabetes	0.81	1.00	<.01	1.00	0.99	.16
Gastroesophageal reflux disease	0.56	0.80	.03	1.00	0.97	.04
Hyperlipidemia	0.74	0.95	<.01	0.99	0.91	.01
Hypertension	0.73	1.00	<.01	0.98	0.94	.15

\*The VPOV file is an administrative file that lists the diagnoses from outpatient clinic visit encounter forms.



determined by medical chart review. Prevalence ranged from 10% for congestive heart failure to 63% for hypertension. The mean patient age was 64 years, and 96% were men; 84% of the study patients had 1 to 4 study diagnoses, with a mean of 2.3 diagnoses (Figure).

The sensitivity of the problem list ranged from 42% for benign prostate hyperplasia to 81% for diabetes (Table 3). The sensitivities of the administrative data (the VPOV file) were better, ranging from 50% for benign prostate hyperplasia to 100% for diabetes and hypertension. There was a statistically significant difference between the 2 files in 6 of the 9 diagnoses. Furthermore, the VPOV file was more sensitive than the problem list when all 9 diagnoses were considered together ( $P < .008$ ).

The specificity for the problem list was quite high (98% to 100%). The specificities for the administrative VPOV files were also very good (91% to 100%). The problem file was significantly more specific than the VPOV file in only 3 of the 9 individual diagnoses (Table 3), but the problem list was significantly more specific when the diagnoses were considered as a group ( $P < .02$ ). Four additional diagnoses were found in the PTF (inpatient file) that were not found in the VPOV file. There was one case each of benign prostatic hyperplasia, coronary artery disease, congestive heart failure, and gastroesophageal reflux disease. Because of the small yield from the PTF it was not included in further analyses.

The PPVs of the problem list (Table 4) ranged from 83% to 100% and were not significantly different than those of the VPOV file, which ranged from

83% to 100%. The PPV was low for benign prostatic hyperplasia due in part to the relatively low prevalence of this condition such that the few false positives had a dramatic effect on the overall PPV. The specificity was comparable to that of the other diagnoses. Moreover, for the high-prevalence diagnoses, such as hypertension, hyperlipidemia, diabetes, and coronary artery disease, there is only a small difference in the PPVs between the problem list and the VPOV file, but the VPOV file provides a significantly higher NPV.

... DISCUSSION ...

We found that outpatient administrative and clinical diagnostic codes relevant to hypertension recorded in VISTA had high specificity, in particular in the clinical files. Some diagnostic codes have very high sensitivity when encounter form data are examined over multiple visits. The VPOV file (an administrative file of encounter form diagnoses) was more sensitive but less specific than the problem file (a clinical file), but both may be affected by the underlying prevalence of the condition.

In our sample, the administrative file identified all patients with a medical chart diagnosis of hypertension, but it also identified as hypertensive several patients whose medical charts did not support that diagnosis. As expected, the administrative file and the clinical problem list file contain significantly different information. For the 9 study diagnoses, the problem list file was less sensitive but more specific than the VPOV file.

**Table 4.** Positive and Negative Predictive Value of Study Diagnoses in Problem Lists and VPOV File

	Positive Predictive Value			Negative Predictive Value		
	Problem List	VPOV	P	Problem List	VPOV	P
Atrial fibrillation	1.00	0.94	.15	0.97	0.99	.09
Benign prostatic hyperplasia	0.83	0.92	.75	0.89	0.90	.35
Coronary artery disease	1.00	0.94	.07	0.84	0.96	<.01
Congestive heart failure	1.00	1.00	—	0.98	0.98	.36
Chronic obstructive pulmonary disease	1.00	0.83	.01	0.93	0.99	.01
Diabetes	1.00	0.97	.16	0.94	1.00	<.01
Gastroesophageal reflux disease	1.00	0.87	.03	0.91	0.96	.07
Hyperlipidemia	0.98	0.89	.06	0.84	0.96	<.01
Hypertension	0.99	0.97	.33	0.65	1.00	<.01

The choice of which particular file to use will not only be affected by the ramifications of a false-negative relative to a false-positive designation but may depend on the underlying prevalence of the particular condition. The problem list file can be used with confidence to select patients with true-positive diagnoses of hypertension, eg, if quality improvement efforts are to be directed toward patients with known hypertension. The administrative file can be used across populations of patients, eg, to examine drug choice for management of hypertension based on specific comorbidities. Such analyses can be used to identify subgroups with the largest deviations from best practice so that interventions can be targeted toward those subgroups.

Although our study was limited to an individual site and a fixed set of chronic conditions, we believe that it is potentially applicable to most large health-care organizations. It is common practice to collect *ICD-9* codes for every clinical encounter and to maintain databases with this administrative information; in addition, most electronic medical record systems include an option for use of a problem list maintained by the clinician. In our study, the sensitivities of the administrative data files for these 9 chronic conditions were very high. We understand that coding practices may differ both within and outside the VA system, but it is likely that the same pattern—administrative data capturing more diagnoses but with less specificity than the clinical files—exists across systems. Large multisite studies would be needed to confirm this.

We realize that some database systems do not maintain a separate clinical problem list. Although we found that the problem list was more specific than the administrative file, one would need to assess whether the potential benefit would justify the added expense associated with maintaining such a file. The differences in specificity were statistically significant but not dramatic. Within the VA, the decision to maintain a problem file is determined locally.

Our findings suggest design features for electronic medical record systems that could potentially enhance the accuracy of the diagnostic data with little added expense. An electronic medical record system could take advantage of the high sensitivity of the administrative outpatient encounter diagnoses by using them to generate initial or updated lists of diagnoses as a draft version of the clinician-maintained problem list, and could then present the problem list during clinic visits with a prompt to the clinician to edit and sign-off on the problem list, to maximize specificity.

There are some limitations to this study. The sample was drawn from actual clinic visits and might oversample for more complicated patients who have a greater number of provider encounters, so our sample might underrepresent healthier patients. At a teaching hospital, patients may be more likely to receive care from multiple providers than in settings with strict gatekeeping. This would potentially increase the likelihood of any single diagnosis being entered into the VPOV file, but not the problem file, because diagnostic codes are entered into the VPOV file from all encounters regardless of the provider, whereas the problem list is generally maintained by the primary care provider. It is likely, however, that the patterns of referral and care by multiple providers for patients with comorbidities are not very different in our setting than from most other large health-care systems.

Our review was limited to a fixed set of common chronic diagnoses for which patients are followed up for long periods, thereby increasing the likelihood that each condition is coded. These findings may be less applicable to other diagnoses, particularly acute illnesses. We did not examine the coding accuracy for each individual encounter.

Last, the medical chart, which was used as the standard for comparison, may contain errors in documentation. Our review was limited to diagnoses recorded in the clinic notes and hospital discharge summaries and did not include laboratory or medication profiles, or blood pressure measurements, which may enhance the accuracy of medical chart diagnoses but may also introduce new sources of error. We selected the medical chart note as the basis for comparison because it is the standard source for quality assessment reviews.

---

### ... CONCLUSIONS ...

Computerized outpatient administrative data across multiple visits may be an accurate and efficient method for ascertaining chronic medical diagnoses. Computerized clinical diagnostic data are more specific, but less sensitive. The choice of whether to use administrative or clinical files should be based on the needs of the specific research or quality assessment project. Because outpatient administrative data are commonly collected by large healthcare organizations, they have the potential of being an inexpensive, readily available source of diagnoses if their accuracy can be verified.

**Acknowledgments**

We thank the VA Information Resource Center (VIREC), a Health Services Research and Development Field unit, Hines, IL, for comments on earlier drafts of this manuscript.

---

... REFERENCES ...

1. **Barbour GL.** Development of a quality improvement checklist for the Department of Veterans Affairs. *Jt Comm J Qual Improv* 1994;20:127-139.
2. **Ashton CM, Petersen NJ, Wray NP, Yu HJ.** The Veterans Affairs medical care system: Hospital and clinic utilization statistics for 1994. *Med Care* 1998;36:793-803.
3. **Menke TJ, Ashton CM, Krumhaus S.** Research organizations in the Department of Veterans Affairs that focus on database analysis. *Med Care* 1996;34:MS103-MS110.
4. **Iezzoni LI.** Assessing quality using administrative data. *Ann Intern Med* 1997;127:666-674.
5. **Tucker AM, Weiner JP, Honigfeld S, Parton RA.** Profiling primary care physician resource use: Examining the application of case mix adjustment. *J Ambulatory Care Manage* 1996;19:60-80.
6. **Powe NR, Weiner JP, Starfield B, Stuart M, Baker A, Steinwachs DM.** Systemwide provider performance in a Medicaid program: Profiling the care of patients with chronic illnesses. *Med Care* 1996;34:798-810.
7. **Weiner JP, Parente ST, Garnick DW, Fowles J, Lawthers AG, Palmer RH.** Variation in office-based quality: A claims-based profile of care provided to Medicare patients with diabetes. *JAMA* 1995;273:1503-1508.
8. **Weiner JP, Starfield BH, Powe NR, Stuart ME, Steinwachs DM.** Ambulatory care practice variation within a Medicaid program. *Health Serv Res* 1996;30:751-770.
9. **Jollis JG, Ancukiewicz M, Delong ER, Pryor DB, Muhlbaier LH, Mark DB.** Discordance of databases designed for claims payment versus clinical information system: Implications for outcomes research. *Ann Intern Med* 1993;119:844-850.
10. **Lloyd SS, Rissing JP.** Physician and coding errors in patient records. *JAMA* 1985;254:1330-1336.
11. **Kashner TM.** Agreement between administrative files and written medical records: A case of the Department of Veterans Affairs. *Med Care* 1998;36:1324-1336.
12. **Mendelsohn A, Whittle J.** Accuracy of coding for cardiac catheterization and percutaneous transluminal coronary angioplasty at a Department of Veterans Affairs Medical Center. *J Am Health Inform Manage Assoc* 1996;67:64-70.
13. **Goff DC, Pandey DK, Chan FA, et al, for the Corpus Christi Heart Project.** Congestive heart failure in the United States: Is there more than meets the I(CD code)? *Arch Intern Med* 2000;160:197-202.
14. **Psaty BM, Boineau R, Kuller LH, Luepker RV.** The potential costs of upcoding for heart failure in the United States. *Am J Cardiol* 1999;84:108-109.
15. **Kolodner RM, ed.** *Computerizing Large Integrated Health Networks: The VA Success.* In: Hannah KJ, Ball MJ, eds. *Computers in Health Care.* New York, NY: Springer; 1997.
16. **Snedecor G, Cochran W.** *Statistical Methods.* Ames, IA: Iowa State University Press; 1989.