Informatics in Primary Care 2004;12:139–45 © 2004 PHCSG, British Computer Society

Refereed papers

A first step towards translating evidence into practice: heart failure in a community practice-based research network

Mihai Onofrei PharmD
Clinical Pharmacy Specialist

Jacquelyn Hunt PharmD
Director of Primary Care Research

Joseph Siemienczuk MD
Chief Medical Officer

Providence Health System, Portland, Oregon, USA

Daniel R Touchette PharmD MA
Assistant Professor, Oregon State University College of Pharmacy, Portland, Oregon, USA

Blackford Middleton MD MPH MSc
Director of Clinical Informatics Research and Development, Partners HealthCare System and Assistant Professor of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT

Objective To determine the validity of an electronic health record (EHR) in the identification of patients with left ventricular dysfunction in a primary care setting.
Design A cross-sectional study.
Setting Nine clinics participating from the Providence Research Network (PRN) comprising 75 physicians serving approximately 200 000 patients. All clinics utilise the Logician™ EHR for all patient care activities.
Patients The study included all PRN patients with an active chart.
Interventions All patients with a heart failure diagnosis in the problem list were identified by database query. Left ventricular ejection fraction (LVEF) data were identified through query of local cardiology and hospital echocardiography databases. Additional LVEF data were sought in a manual search of paper charts.
Measurements and main results To determine the problem list coding accuracy for a heart failure (HF) diagnosis we evaluated sensitivity, positive predictive value and related derived statistical measures using documented LVEF as the 'gold standard'. Of 205 755 active PRN patients, 1731 were identified with a problem list entry of HF. Based on comparison with documented LVEF, the sensitivity for problem list entry was 43.9% and 54.4% when HF was defined as an LVEF $\leqslant$55% and $\leqslant$40%, respectively.
Conclusion The validity of an EHR problem list entry of HF was poor. The problem list validity could be enhanced through reconciliation with other data sources. Inaccurate EHR problem lists may have clinical consequences, including under-prescribing of beneficial therapies.

Keywords: data quality, heart failure, medical record system, quality improvement

Introduction

A major focus of practice-based research networks (PBRN) is the translation of existing evidence into clinical practice. Through the rigorous evaluation of methods of translating research into practice in community settings, PBRNs have the potential to identify and assess implementation strategies that are most likely to be effective and sustainable.[1] Of particular interest among PBRNs is the use of information technology, such as electronic health records (EHRs), in translating research findings into improvements in care.

Disease states that are optimal candidates for translational research are highly prevalent, associated with known effective therapies and suffer from undertreatment. Accordingly, the Providence Research Network (PRN) teamed researchers with practising physicians interested in improving care in the area of heart failure.

Strategies aimed at improving heart failure management begin with the accurate identification of candidate patients. Commonly, identification of patient populations with a specific condition (such as heart failure [HF]) can be accomplished by query of administrative or claims databases. The validity of such identification strategies can be determined using measures such as sensitivity, specificity and predictive values. This approach necessitates comparison of data with a 'gold standard' measurement. Sources of gold standard measurements include manual medical record abstraction, patient survey and electronic databases, alone or in combination.[2–5] In one study, reliance on administrative data resulted in exclusion of approximately one-third of patients with clinical evidence of HF.[6]

It has been suggested that EHRs represent a source of structured clinical data that complement or could even replace administrative data sources for the purpose of identifying patient populations and measuring quality of care.[7] Since the patient's diagnosis list is created and maintained in the EHR by their practitioner, it has the potential to be a very sensitive and specific identification source. Although several papers have described the validity of administrative data to identify patient populations, much less is known about the validity of EHR-derived data.[8,9] Accordingly, the aim of this study was to determine the validity of EHR for identification of patients with systolic dysfunction using sensitivity, positive predictive value and related derived measures.

Methods

Study setting

The PRN is a community-based network of internal medicine and family practice clinics in Portland, Oregon. Network clinics participating in this study consisted of approximately 75 primary care providers caring for more than 200 000 patients in nine clinic locations. All practices share a common EHR, Logician™ (GE Medical Systems Information Technology, 20540 NW Evergreen Parkway, Hillsboro, OR 97124) that was implemented in most clinics by 1997. All office encounters and telephone contacts by all physicians and staff are documented in the EHR in a text-based format. The EHR problem lists are generated and maintained through physician entry of patient diagnosis, analogous to the traditional paper chart. Problem list entries are stored in a searchable format, based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) system. Additional patient data, including demographics, insurance status, medications and laboratory values, are available in a searchable database. Interfaces to the EHR allow providers to access laboratory and radiology results as well as hospital and emergency room records. During the study period external documents, such as records transferred from previous physicians, specialty consult letters and non-interfaced test results, were stored in a complementary paper chart.

The study population consists of all patients with an active record in the PRN's medical record database. Patients were excluded if they left a PRN clinic or died during the study period. Patients were defined as inactive and were also excluded if their electronic chart lacked documentation of any contact, including evidence of an office visit, telephone communication or prescription refill, between 1 January 1998 and 31 December 2000.

Methods

This study was a cross-sectional analysis comparing patient EHRs with independently maintained cardiology laboratory records. The study was approved by the local institutional review board in February 2001. Physician-entered diagnosis of HF in the EHR was reconciled against left ventricular ejection fraction (LVEF) as measured by echocardiography, angiography or nuclear imaging. These measurements of LVEF were used as the gold standard indicators for determining whether a patient had HF due to left ventricular dysfunction. Owing to ambiguity surrounding the definition of systolic dysfunction, two analyses were performed. Depressed left ventricular function was defined as an LVEF $\leqslant$55% for the first analysis and $\leqslant$40% for the second. The physician-entered diagnosis of HF was defined by the following search terms: ICD-9-CM 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.xx and 428.xx (see Table 1). At the time of this study, ICD-9-CM codes did not support differentiation among systolic dysfunction and other categories of HF. Accordingly, an inclusive list of HF codes was assembled.

The gold standard measure of LVEFs was constructed through several processes to ensure completeness. First, LVEF values were extracted from the

Table 1 Frequency of heart failure-related diagnosis and ICD-9-CM codes in the PRN EHR

ICD-9-CM codes	Description	Frequency no.	(%)
402.01	Hypertensive heart disease, malignant with CHF	7	(0.43)
402.11	Hypertensive heart disease, benign with CHF	16	(0.92)
402.91	Hypertensive heart disease, NOS with CHF	10	(0.59)
404.01	Hypertensive heart/renal disease, malignant with CHF	1	(0.05)
404.03	Hypertensive heart/renal disease, malignant with CHF + renal failure	1	(0.05)
404.11	Hypertensive heart/renal disease, benign with CHF	0	(0)
404.13	Hypertensive heart/renal disease, benign with CHF + renal failure	2	(0.12)
404.91	Hypertensive heart/renal disease, NOS with CHF	0	(0)
404.93	Hypertensive heart/renal disease, NOS with CHF + renal failure	0	(0)
425.xx	Cardiomyopathy	241	(13.92)
428.xx	Heart failure	1453	(83.92)
	Total	1731	(100)

NOS, not otherwise specified.

echocardiography databases from all local laboratories utilised by PRN practitioners, including hospital and cardiology practice laboratories. It is unlikely that LVEF assessment would be performed outside of the local laboratories. Second, manual review was performed on all charts of patients identified with HF or an HF-related diagnosis in the problem list of the patient record. Any evidence of a test yielding a measured LVEF derived from an outside cardiology laboratory or any results mentioned in previous chart records or consult letters were extracted. Lastly, HF patients who were not identified in the previous steps were sought through a query of billing data and chart medication lists. The practice management software of PRN clinics was queried for evidence of billing for a service related to any of the referenced HF-related ICD-9-CM codes. Patients were also screened if carvedilol was present as an active or historical prescription in the chart medication list. Carvedilol was selected as an identifier because it is prescribed solely for HF. A chart review was then conducted to confirm or exclude the presence of a depressed LVEF. LVEF values from all of these sources were then aggregated to create the gold standard list of patients with confirmed systolic dysfunction. In the case of multiple LVEF values, the lowest value was selected for the purpose of the study.

Statistical analysis

The active population, stratified by presence or absence of a documented chart diagnosis of HF, was crossed with the gold standard list of patients with confirmed systolic dysfunction. A two-by-two table was used to calculate the following standard statistics applicable to diagnostic tests. The EHR problem list was treated as a diagnostic test for systolic dysfunction. A problem list entry for HF was considered a positive result for systolic dysfunction. Since determination of LVEF is a prerequisite for a diagnosis of systolic dysfunction, the absence of an LVEF entry was defined as a negative result.

The completeness and accuracy of the patient problem list were inferred by measures of sensitivity and positive predictive value, respectively.[5] The likelihood ratio for a positive result was the ratio of the chance of a positive result if the patient had the disease to the chance of a positive result if the disease was absent. In addition, we report the true positive to false negative ratio (TPFN), and the number of false negatives to total number of patients in the database ratio times 10 000 ($DBFind^{10\,000}$), which are two newly developed and published statistical tools.[5] The $DBFind^{10\,000}$ represents the number of false negatives in a database of 10 000 patients. The lower the TPFN ratio and the higher the $DBFind^{10\,000}$ results, the greater the number of false negative cases in the database.

Results

Review of all sources of measured LVEF yielded 1403 patients in the PRN with an LVEF value ≤55%, of whom 793 had an LVEF ≤40%. The majority of LVEF values were identified from local echocardiography databases (n=713), followed by manual chart review (n=558), billing data (n=78), and carvedilol use (n=54). The EHR database relevant to the participating PRN clinics contained 205 755 active patient records. Of active patient records, practitioners had entered an HF diagnosis in the problem list of 1731 patient charts.

There were 616 patients identified with both a problem list diagnosis of HF and a confirmed LVEF ⩽55%. This true positive number drops to 431 when LVEF threshold is ⩽40%. Patients with an LVEF <55% and <40% whose chart did not contain a problem list entry for HF diagnosis (false negatives) totalled 787 and 362, respectively. Patients whose chart contained an HF diagnosis but lacked documentation of depressed LVEF were defined as false positives. Patients without documentation of depressed LVEF either had no evidence of LVEF assessment or a test result that demonstrated preserved left ventricular function. Using the 55% LVEF threshold, 214 patients had documentation of preserved LVEF, while 902 had no evidence of LVEF assessment. The findings for the 40% threshold were 398 and 902, respectively.

Sensitivity and positive predictive values are variable based on the LVEF definition of systolic dysfunction. When systolic dysfunction is defined as an LVEF ⩽55%, sensitivity was 43.9%, positive predictive value was 35.5% and the likelihood ratio for a positive result was 80.3 (see Table 2a). Comparatively, when the LVEF definition is decreased to ⩽40%, sensitivity increased to 54.4%, while positive predictive value fell to 24.9% (see Table 2b). The likelihood ratio for a positive result increased modestly, to 85.7. Specificity of the EHR was 99.5% and 99.4% for an LVEF of ⩽55% and ⩽40% respectively, due largely to the low prevalence of HF.

TPFN, the ratio of true positives to false negatives, was 0.8 with an LVEF definition ⩽55%, and increased to 1.2 when LVEF ⩽40%. $DBFind^{10\,000}$ decreased from 38.3 to 17.6 when the LVEF threshold was changed from ⩽55% to ⩽40%.

Discussion

Our analysis assessed the utility of a problem list entry for heart failure in an EHR to identify candidates for management of systolic dysfunction. Currently, accurate identification of individuals with systolic

Table 2a Results for patients with an LVEF ⩽55%

		Evidence of depressed LVEF (LVEF ⩽55%)		Total
		Present	Absent	
Diagnosis of HF in problem list	Present	True '+' (A) 616	False '+' (B) 1116	(A + B) 1732
	Absent	False '–' (C) 787	True '–' (D) 203 237	(C + D) 204 024
Total		(A + C) 1403	(B + D) 204 353	(A + B + C + D) 205 756

		95% Confidence interval
Sensitivity A/(A+C)	43.9%	43.7%–44.1%
Specificity D/(B+D)	99.5%	99.4%–99.5%
Likelihood ratio '+'	80.3	
Likelihood ratio '–'	0.6	
Positive predictive value A/(A+B)	35.5%	35.4%–35.8%
Negative predictive value D/(C+D)	99.6%	99.6%–99.6%
TPFN ratio	0.8	
$DBFind^{10\,000}$	38.3	

LVEF = left ventricular ejection fraction

TPFN = true positive to false negative

Table 2b Results for patients with an LVEF ≤40%

		Evidence of depressed LVEF (LVEF ≤40%)		Total
		Present	Absent	
Diagnosis of HF in problem list	Present	True '+' (A) 431	False '+' (B) 1300	(A + B) 1731
	Absent	False '–' (C) 362	True '–' (D) 203 662	(C + D) 204 024
Total		(A + C) 793	(B + D) 204 962	(A + B + C + D) 205 755

		95% Confidence interval
Sensitivity A/(A+C)	54.4%	54.1%–54.6%
Specificity D/(B+D)	99.4%	99.3%–99.4%
Likelihood ratio '+'	85.7	
Likelihood ratio '–'	0.5	
Positive predictive value A/(A+B)	24.9%	24.7%–25.1%
Negative predictive value D/(C+D)	99.8%	99.8%–99.8%
TPFN ratio	1.2	
DBFind$^{10\ 000}$	17.6	

LVEF = left ventricular ejection fraction
TPFN = true positive to false negative

dysfunction for the purpose of providing targeted interventions is a difficult task. Structured medical data stored by an EHR may be expected to complement other data sources for a variety of uses, including quality improvement processes, disease management programmes and clinical research. Validation of the accuracy of EHR patient problem lists not only describes current physician practice, but is also a prerequisite in defining the broader utility of this information source. It has been suggested that the validity of these data is best characterised by statistical measures often applied to diagnostic tests, such as sensitivity, positive predictive value and likelihood ratios.[5] This study characterises the utility of a problem list database for identification of candidates for heart failure disease management interventions among 205 755 patients in a community-based primary care PBRN.

Sensitivity and positive predictive value were dependent upon the LVEF value used to define systolic dysfunction, and are reported here for LVEF thresholds of 40% and 55%. In both cases sensitivity and positive predictive value are disappointing. The sensitivity of 54.4%, obtained even with the LVEF threshold set at 40%, implies that nearly half of the systolic dysfunction population would be overlooked using this search strategy. Likewise, the positive predictive value of 24.9%, determined at the same LVEF threshold, implies that a patient with a problem list entry for HF has only a one-quarter likelihood of documented systolic dysfunction.

The results of the derived statistics TPFN ratio and DBFind$^{10\ 000}$ also suggest poor validity of the problem list database. The latter statistic represents the number of false negatives in a database of 10 000 patients, and can therefore be used to quantify and compare the potential benefit of reconciliation of the database. Thus, with validation of the database one could expect to identify 17–38 additional patients with systolic dysfunction in every 10 000 patients.

These results are of increasing value based on the increasing prevalence of EHR use and the comparable processes for physician entry of diagnosis that exist across EHR types. However, the findings may not be reflective of data accuracy in countries such as the United Kingdom with a 99% prevalence of EHRs in primary care together with dedicated efforts to improve data quality and information management.[10,11] The search capability of the EHR offers potential new uses for medical record data, including problem lists. One source of problem list inaccuracy may relate to the behavioural component of data entry. Physician attention to coding completeness and accuracy may be expected to increase with growing appreciation of the potential uses of clinical databases.

Finally, the most accurate data may be compiled through the aggregation of multiple data sources. In this study of a heart failure population, combining the EHR problem list data with electronic LVEF values from the cardiology laboratory would be expected to provide the most valid identification strategy. EHRs are capable of importing clinical data, such as LVEF values, from external sources. The routine application of such interfaces should enhance validity.

This study has several limitations. The gold standard for defining a diagnosis of systolic dysfunction was a measured LVEF. As a result appropriately coded instances of diastolic dysfunction, in the presence of preserved or unmeasured systolic function, are deemed false positives. At the time of this analysis, HF diagnosis in the EHR problem list was incapable of distinguishing isolated diastolic dysfunction from systolic dysfunction, as there were no differentiating ICD-9-CM codes. The most recent revisions to the ICD-9-CM system now include codes specific for systolic (428.2x), diastolic (428.3x) and combined systolic and diastolic heart failure (428.4x).[12] Adoption of these newly revised codes in both clinical and administrative applications will benefit quality improvement, disease management and research initiatives. This study also was not designed to quantify occult systolic dysfunction, so problem list inaccuracy associated with undiagnosed heart failure remains undetermined. Finally, the criteria used to identify active patients are arbitrary, and this may have resulted in the inclusion of some patients no longer receiving care in PRN clinics.

Although the potential utility of accurate patient problem lists for such activities as disease management has been discussed, the implications of problem list inaccuracy in conventional practice remain undetermined. This may be elucidated through comparison of patterns of drug utilisation between patients in the true positive and false negative categories. It will be beneficial to determine whether false negative patients are less likely to receive valuable therapies, such as ACE inhibitors, beta blockers and spironolactone, or more likely to receive relatively contraindicated therapies, such as non-steroidal anti-inflammatory agents and dihydropyridine calcium channel antagonists.

Conclusion

In conclusion, the results of this validation study of EHR-stored patient problem list data in the primary care setting are disappointing. Assessment of accuracy is an important step in the evolving application of EHR data to patient care, disease management, clinical research and related activities. EHR database validity may be enhanced through integration with other clinical data sources. The ramifications of incomplete and inaccurate EHR databases are unknown and merit investigation.

REFERENCES

1 Green LA and Dovey SM. Practice-based primary care research networks. *British Medical Journal* 2001;322:567–8.
2 Whitelaw FG, Nevin SL, Milne RM *et al*. Completeness and accuracy of morbidity and repeat prescribing records held on general practice computers in Scotland. *British Journal of General Practice* 1996;46:181–6.
3 Whitelaw FG, Nevin SL, Taylor RJ *et al*. Morbidity and prescribing patterns for the middle-aged population of Scotland. *British Journal of General Practice* 1996;46:707–14.
4 Mant J, Mant F and Winner S. How good is routine information? Validation of coding for acute stroke in Oxford hospitals. *Health Trends* 1997–8;29:96–9.
5 Hassey A, Gerrett D and Wilson A. A survey of validity and utility of electronic patient records in a general practice. *British Medical Journal* 2001;322:1401–5.
6 Goff DC, Pandey DK, Chan FA *et al*. Congestive heart failure in the United States. *Archives of Internal Medicine* 2000;160:197–202.
7 Hunt JS, Siemienczuk J and Middleton B. Use of an electronic medical record in disease management programs: a case study in hyperlipidemia. *Medinfo* 2001;10 (Pt 1):825–9.
8 Taylor DH, Fillenbaum GG and Ezell ME. The accuracy of Medicare claims data in identifying Alzheimer's disease. *Journal of Clinical Epidemiology* 2002;55:929–37.
9 Andrade SE, Gurwitz JH, Chan KA *et al*. Validation of diagnoses of peptic ulcers and bleeding from administrative databases: a multi-health maintenance organization study. *Journal of Clinical Epidemiology* 2002;55:310–13.
10 Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N and Wright L. Methodology for assessing the usefulness of general practice data for research in one research network. *Health Informatics Journal* 2004;10:91–109.
11 Horsfield P and Teasdale S. Generating information from electronic patient records in general practice: a description of clinical care and gender inequalities in coronary heart disease using data from over 2 million patient records. *Informatics in Primary Care* 2003;11:137–44.
12 Hart A and Hopkins C (eds). *2003 ICD-9-CM Expert for Physicians, Volumes 1 and 2. International Classification of Diseases, 9th revision, Clinical Modification* (6e). Salt Lake City, Utah: Ingenix, 2002.

CONFLICTS OF INTEREST

None.

FUNDING

Unfunded.

ADDRESS FOR CORRESPONDENCE

Jacquelyn Hunt PharmD
Director of Primary Care Research
Providence Medical Group
3601 SW Murray Blvd, Suite #45
Beaverton, Oregon 97005
USA
Tel: +1 503 574 9850
Fax: +1 503 574 9863
Email: Jacquelyn.Hunt@providence.org

Accepted August 2004