Published in final edited form as:
Am J Manag Care. 2008 August ; 14(8): 530–539.

Agreement between Patient-reported Symptoms and their Documentation in the Medical Record

Serguei Pakhomov, PhD[*], **Steven J. Jacobsen, MD, PhD**[†], **Christopher G. Chute, MD, DrPH**[**], and **Véronique L. Roger, MD, MPH**[††]

Department of Pharmaceutical Care and Health Systems, University of Minnesota, MN

**Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester MN*

†*Department of Research and Evaluation, Kaiser Permanente, Pasadena, CA*

††*Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN*

Abstract

Objective—To determine the agreement between patient-reported symptoms of chest pain, dyspnea and cough and the documentation of these symptoms by physicians in the electronic medical record (EMR).

Methods—Symptoms reported by patients on patient provided information forms between January 1, 2006 and June 30, 2006 were compared to those identified with natural language processing (NLP) of the text of clinical notes from care providers. Terms that represent the three symptoms were used to search clinical notes electronically with subsequent manual identification of the context (e.g. affirmative, negated, family history) in which they occur. Results are reported using positive and negative agreement, and kappa statistics.

Results—Symptoms reported by 1,119 patients 18 years or older were compared to the non-negated terms identified in their clinical notes. Positive agreement was 74, 70 and 63 for chest pain, dyspnea, and cough, while negative agreement was 76, 76 and 75, respectively. Kappa statistics were 0.50 (95%CI 0.41-0.59) for chest pain, 0.46 (95%CI 0.37-0.54) for dyspnea and 0.38 (95%CI 0.28-0.48) for cough. Positive agreement was higher for older men ($p>0.05$) while negative agreement was higher for younger women ($p>0.05$).

Conclusions—We found discordance between patient self report and documentation of symptoms in the medical record. This has important implications for research studies that rely on symptom

Correspondence to Serguei V. Pakhomov, Ph.D., Department of Pharmaceutical Care and Health Systems, 7-125F Weaver-Densford Hall, 308 Harvard St. S.E., Minneapolis, MN; phone # (612) 624-1198, FAX # (612) 625-9931, e-mail: pakh0002@umn.edu.

Financial disclosure: There are no conflicts of interest.

Précis: In this article, we present evidence that that effective use of the EMR technology requires to examine the communication approaches of care providers and patients.

Take away points: The electronic medical record documentation of symptoms differs from the symptoms self-reported by patients. Efforts that rely on using symptoms information must take this into account.

MeSH: C23.888 (Signs and Symptoms); L01.224.065.580 (Natural Language Processing); N04.452.859.564.650 (Medical Records Systems, Computerized); N05.300.660.625 (Physician-Patient Relations)

information for patient identification and may have clinical implications that must be evaluated for potential impact on quality of care, patient safety and outcomes.

Keywords

angina pectoris; patient-reported symptoms; natural language processing; electronic medical record

---

Physical symptoms account for half of all outpatient visits in the US[1] and are commonly not diagnosed[2]. The verbal characterization of the symptoms conveyed by the patient, and recorded by care providers is central to the practice of clinical medicine, and increasing importance is attached to patient-centered clinical care[3]. With the increasing adoption of the electronic medical record, free text of the clinical history can now be subjected to automated analysis in ways which are impossible or uneconomic with paper based records[4-6]. A common, costly[7] example of a symptomatic condition in which history taking is central to management is chronic stable angina pectoris[8], the number one cause of death in the entire Western World[9]. In addition to its mortality burden, angina is associated with serious morbidity such as myocardial infarction and heart failure. Optimal methods of identifying stable angina patients remain unclear; many patients with typical symptoms are not diagnosed as angina[10] and age, sex, and ethnicity may influence physician's recommendations for diagnostic testing such as coronary angiography and the resulting ICD codes[11-14]. Our preliminary findings indicate that Natural Language Processing (NLP) of the free text of the EMR identifies patients with chronic angina pectoris[15] and heart failure[15] that were missed by traditional diagnostic coding approaches. NLP is a range of computational techniques aimed at extracting useful information from unstructured text. In the context of the EMR, NLP offers a promising method to automate the collection of a richer set of information for quality improvement and safety that would otherwise require manual chart abstraction.

Early identification of patients at risk for myocardial infarction is critical to its prevention and improved prognosis[16-18] and is possible using patient-reportable information[19]. Since the diagnosis of angina pectoris relies on the presentation of the symptoms that is conveyed by the patient to the physician[20], the ambiguous nature of verbal communication as well as the nature of coronary artery disease presenting itself differently in patients of different race and gender can make the diagnosis challenging. These challenges can lead to inconsistency and incompleteness in both the diagnosis and the information recorded in the EMR[21]. Currently, the nature of the processes that lead to the creation of the EMR is poorly understood; however, prior research indicates that these processes have critical implications for clinical care and may have a significant impact on patient outcomes.

The primary objective of the current study is to determine if there is disagreement between patient-reported symptoms of chest pain, dyspnea and cough and the documentation of these symptoms by care providers in the EMR. While in this study we focus on heart disease, which constitutes one of the top national health care priority areas[22], our findings have significant implications for any condition whose diagnosis and treatment relies on verbal presentation of symptoms.

METHODS

Sources of symptom documentation

For this study we identified two primary sources of symptom information: patient provided information forms and clinical notes. Both are part of the Mayo Clinic EMR.

Electronic Medical Record and Clinical Notes—Mayo Clinic maintains an Electronic Medical Record (EMR) for each in- and outpatient. A large part of the EMR consists of clinical notes which represent transcriptions of the dictations recorded by care providers after each contact with the patient and subsequently transcribed by trained medical transcriptionists. Clinical notes have been in use electronically since 1994; however, all services at the Mayo Clinic switched to electronic notes in 2005. The content of these notes complies with a nationally accepted standard, Health Level 7 (HL7) Clinical Document Architecture, and consists of standard sections including chief complaint, history of present illness, review of systems, impression/report/plan, among others. This standard is used by most major EMR vendors in the United States.

Current Visit Information Forms—Many health care providers including the Mayo Clinic ask their patients to fill out forms detailing their prior health and social history, current symptoms, medications and allergies. Figure 1 shows a portion of a patient provided information form asking the patient to indicate if he/she has ever experienced a symptom listed on the form. Subsequently, patient's responses are captured in a structured format by scanning the form into a database.

Study population

We used a subset of the database of patient provided information entered during 6 months between January 1, 2006 and June 30, 2006 as the convenience sample consisting of 121,891 patients who filled out the form at least once during the 6 month period. The patients in this sample are not restricted to any geographic location and represent general Mayo Clinic ambulatory and hospitalized populations. We selected records that indicated either positive or negative response to form items for chest pain, chest pressure, shortness of breath/dyspnea and cough[1]. The analysis presented in this manuscript is based on a combined symptom for chest pain that includes both chest pain and chest pressure responses.

Manual verification of symptoms

We manually examined clinical notes containing evidence of chest pain, dyspnea and cough according to the procedure illustrated in Figure 2. For each of the symptoms, we randomly selected 200 patients who marked the symptom on at least one of their forms and 200 patients who did not mark the symptom on any of their forms. No further restrictions other than age of >=18 years old were applied. Only records of patients who had one or more clinical note dictated during the study period (January 1, 2006 - June 30, 2006) were used in this study.

Each clinical note was searched automatically to identify and electronically mark the terms representing each of the symptoms and their orthographic variants and synonyms using search keywords arranged into natural language queries as shown in Figure 3. The keywords and methods were similar to those previously reported[15]. For example, a natural language query for "chest pain" identifies portions of clinical note text where one of the terms describing PAIN (e.g. pressure) either precedes or follows one of the terms describing the LOCATION (e.g. chest). Thus this query will find either "chest pressure" or "pressure in the chest." Subsequently, the text of the notes was manually examined by two nurse abstractors to determine the context in which each term appeared. The range of possible context labels that the abstractors could choose is displayed in Column 1 of Table 1. The "conditional" context label is selected when the term is mentioned in "conditional" context (e.g. "I recommended nitroglycerin if he should develop chest pain"). Terms manually identified as "negated", "family history", "conditional" or "unknown" were excluded from subsequent analysis.

[1]These codes are internal identifiers provided here for reference and have no relation to any standardized nomenclature of medical concepts.

In addition to the identification of the context in which the query terms appeared, the abstractors manually identified all symptoms of chest pain, dyspnea and cough in a random sample of 100 clinical notes. These manually identified symptoms were compared to those identified with automated natural language queries to determine their sensitivity.

Angina refers to a complex of symptoms one of which may be chest pain or discomfort. Similarly to dyspnea, orthopnea and PND (paroxysmal nocturnal dyspnea) refer to a specific kind of shortness of breath. The patient provided information form used for this study contains a question that covers this type of dyspnea (see “awakened with shortness of breath” in Figure 1). For the sensitivity analysis we added the terms “angina”, “orthopnea” and “PND” to the natural language queries to determine if these related terms have a measurable effect on the agreement between symptoms reported by patients and care providers.

Statistical analysis

For the primary analysis and the symptom reporting consistency, data were analyzed in terms of positive (Ppos) and negative (Pneg) agreement rather than sensitivity and specificity due to fact that neither the patient self-report nor the clinical notes could be considered a perfect criterion standard. Traditional kappa statistics are sensitive to imbalances in the marginal totals of 2×2 comparisons[23]. Positive and negative agreement measures have been proposed as a way to ensure the correct interpretation of kappa values[24] and have been used to assess the agreement between patient reported information and the medical record[25]. The computation of these measures is illustrated in Table 4 that represents a 2×2 table containing the counts where the patient responses were in concordance or discordance with the symptoms found in clinical notes for the same patient. Thus the positive agreement is a ratio of the concordances in positive responses to the difference between the concordances in positive and negative responses added to the total number of samples according to the following formula: 100*(2a/[N+(a-d)]). The negative agreement is a ratio of the concordances in negative responses to the difference between the concordances in positive and negative responses subtracted from the total number of samples according to the following formula: 100*(2d/[N-(a-d)]) [24]. In addition we also report on the false negative rate which, in our case, shows the proportion of times where the patient reported a symptom but the symptom did not appear in the physician’s note (i.e. false negative rate = 100*(c/(a+c))). We report standard measures of sensitivity and specificity instead of positive and negative rates for the assessment of reliability of identification of negation by the NLP system because we are using a manually created reference standard. Stratified analyses in gender and age subgroups were performed as well.

RESULTS

Prior to sampling, of the 121,891 patients who filled out a patient provided information form during 6 months between January 1, 2006 and June 30, 2006, 6,569 patients (5.39% 95%CI 5.26-5.52) reported “chest pain”, 6,166 patients (5.06% 95%CI 4.94-5.18) reported “chest pressure”, 13,924 patients (11.42% 95%CI 11.24-11.60) reported “dyspnea” and 11,670 patients (9.57% 95%CI 9.41-9.74) reported “cough”. Combining chest pain and chest pressure as synonymous terms yielded 10,518 patients (8.63% 95%CI 8.47-8.79) who reported either chest pain or pressure. These results are illustrated in Table 1. Random sampling resulted in a study population of 1,119 patients with 373 patients in the “chest pain” group, 391 patients in the “dyspnea” group, 337 patients in the “cough” group, and 18 patients (1.6%) with multiple group membership. Of these 1,119 patients, 499 (44.6%) were male, 127 (11.3%) were between 18 and 34 years old, 185 (16.5%) were between 35 and 49, 337 (30.1%) were between 50 and 64 and 470 (42.0%) were 65 or older. Twelve percent of the clinical notes for the 1,119 patients were marked as a “Hospital Admission” or “Hospital Dismissal” note thus representing an estimate of the notes from the hospitalized population. Only 6 patients filled out more than one patient provided information form during the study period. Overall, the clinical notes in this study sample represent a variety of clinical specialties including primary care (20%), cardiovascular services (13%), and physical medicine and rehabilitation (5%), hematology (4%), neurology (4%), endocrinology (4%), surgery (4%), psychology/psychiatry (4%), gastroenterology (3%) and emergency medicine (2%) among other less prevalent specialties.

Validation of natural language queries

The distribution of contexts illustrated in Table 1 shows that the terms denoting chest pain, dyspnea and cough occur in negated contexts between 18% and 30% of total occurrences. These observations are consistent with Chapman et al.[26] where 27% of findings and diseases were manually identified as negated in the process of creating a reference standard. We examined the precision of the NLP methodology for identification of the context (negated vs. affirmative) in which symptoms are mentioned in the free text of clinical reports. Complete results are presented in Table 3. The sensitivity of the NLP negation algorithm was 83% (95% CI 78-88), 84% (95%CI 80-87) and 88% (95%CI 84-90) for cough, dyspnea and chest pain respectively, while the specificity was 91% (95%CI 90-93), 93% (95%CI 91-94) and 92% (95%CI 91-94). The kappa values were 0.69 (95%CI 0.63-0.74) for cough, 0.74 (95%CI 0.71-0.78) for dyspnea and 0.78 (95%CI 0.75-0.81) for chest pain.

In the random samples of 100 notes for each of the 3 symptoms, the abstractors identified 34 mentions of chest pain, 46 mentions of dyspnea and 24 mentions of cough. Of the 34 chest pain mentions, 31 were also identified by the natural language queries yielding 91% (95%CI 82-100) sensitivity. Of the 46 dyspnea mentions, 45 were identified by the queries yielding 98% (95% CI 94-100) sensitivity. All 24 mentions of cough were also identified by automatic queries.

Agreement between Clinical Notes and patient provided information forms

The results are summarized in Table 2. Overall, the positive agreement (Ppos) for chest pain, dyspnea and cough was 74, 70 and 63 respectively, while the negative agreement (Pneg) was 78, 76 and 75. The kappa values were 0.52 (95% CI: 0.43-0.60) for chest pain, 0.46 (95% CI: 0.37-0.54) for dyspnea and 0.38 (95% CI:0.28-0.48) for cough. Additional analysis was carried out within different sex and age strata to determine if the concordance and discordance are influenced by these factors. Positive agreement was slightly lower for female patients across all three symptoms, and tended to be lower for younger subgroups. Negative agreement varied and tended to be lower for males and older subgroups. Kappa values ranged from 0.44 - 0.55 for chest pain, 0.27 - 0.52 for dyspnea and 0.31 - 0.56 for cough.

Including non-negated mentions of “angina” resulted in identification of 5 additional patients who self-reported chest pain and 3 additional patients who did not yielding positive agreement of 76 and negative agreement of 75. Including non-negated mentions of “orthopnea” and “PND” resulted in identifying one additional patient who did not self-report dyspnea yielding positive agreement of 70 and negative agreement of 75.

Distribution of clinical specialties for discordant results

We determined the distributions of clinical services that reflect the clinical specialty of care providers for the discordant results for each of the three symptoms (**Error! Reference source not found.**). The distributions were computed by counting the number of clinical notes originating in each of the top 10 most prevalent clinical services. Consistent with the overall distribution of clinical services in our sample, cardiovascular services and primary care/family medicine are the top two most prevalent specialties for the patients in the discordant groups. The majority of the notes for patients reporting chest pain that was not documented in the EMR were distributed between primary care, cardiovascular services and endocrinology. A similar distribution is found for dyspnea. For cough, the majority of the notes are distributed between primary care, cardiovascular and physical medicine and rehabilitation services. The majority of the notes for patients that had documentation of chest pain in the EMR but did not report it on patient provided information forms were distributed between primary care, psychology/ psychiatry, hematology and cardiovascular services. A similar distribution is found for cough with a large proportion of notes from the hematology service. For dyspnea, the majority of the notes originate from primary care and cardiovascular services.

DISCUSSION

Electronic Medical Records (EMR) offer a way to make improvements in the quality and safety of patient care by addressing the information challenge of organizing and making patient charts accessible and interoperable across health care providers[27]. The primary and most important use of the EMR is to facilitate and streamline care delivery; however, its secondary uses for clinical research and quality and safety assurance are also important[28]. From the standpoint of the primary use of the EMR, we currently do not have the data to determine if the discordances between the symptoms reported by patients and those documented by care providers have any significant clinical consequences. However, we do present evidence that that effective secondary uses of the EMR require to examine the communication approaches of care providers and patients. Our findings indicate substantial discordance between patient reporting and care provider documentation of the symptoms. Thus the two sources of the symptoms may complement each other and have implications for clinical studies and quality measurements that rely on the medical record for identification of symptoms. For example, to ensure completeness of identifying and recruiting participants with specific symptoms for clinical studies, it may be necessary to use the information reported by patients on self-entry forms in addition to other sources such as the clinical notes. It may also be necessary to use both sources of information for applications such as post-marketing medication safety surveillance where it may be necessary to detect alarming trends in symptoms in order to prevent more serous adverse drug reactions. The current study is the first step towards understanding the implications of symptom documentation practices for both primary and secondary uses of the EMR.

Validation of natural language queries

The results in Table 3 indicate that NLP is a valid tool for identification of symptoms that occur in negated contexts in the free text of clinical reports. These results are promising because they show that the NLP methods used in this pilot study may be applied retrospectively to an existing cohort of patients with known outcomes to determine if there is an association between the outcomes and symptom documentation. We also found that the methods used to identify the mentions of chest pain, dyspnea and cough are highly sensitive. Automated natural language queries missed “chest “tightness””, “sense of bruising in the lateral and posterior chest wall”, “anginal symptoms” and “dyspneic.” Capturing these types of cases will require only minor modifications to the algorithms used for term identification. Inclusion of terms “angina”,”orthopnea” and “PND” resulted in very minor changes in the positive and negative agreement thus not affecting the conclusions of the primary analysis. It is likely that these more specialized terms are used in conjunction with more generic terms such as chest pain and dyspnea and thus do not yield additional patients.

Agreement between Clinical Notes and patient provided information forms

In a previous study of the Mayo Clinic EMR, St. Sauver et al.[25] have shown that positive reports of cardiovascular disease risk factor information (blood pressure, triglycerides, cholesterol, history of heart rhythm, heart valve and arterial problems) by patients are largely inaccurate while negative reports were unlikely to be noted on the medical record by care providers thus showing a low level of agreement between patient-reported risk factors and physician documentation. This study, however, did not address symptom reporting and documentation, as this was not in the scope of their research questions. While patient self-report is a secondary source of information on lab results and past medical diagnoses with physician documented medical record being the primary source (thus perhaps contributing to the low agreement found by Sauver et al.[25]), the situation is reversed with respect to patient symptoms where the patient is the primary source and the care providers - secondary. The results of the current study are consistent with those reported by St. Sauver et al. and contribute to constructing a more complete picture of clinical process and documentation.

While we found discordance between patient and physician reported information, the causes of such discordances can only be hypothesized. Indeed, discordance may reflect the nature of the patient-physician interaction or other factors such as clinical specialty. For example, for discordances on chest pain reporting and documentation, one might expect to find the majority of clinical notes originating from non-cardiovascular services. Our data shows that while this is true, a relatively large proportion of the notes (24%) originated from cardiovascular services. The results are similar for dyspnea and cough. The majority of the notes for patients that reported chest pain that was not documented in the EMR originated in three services: cardiovascular, endocrinology and primary care. Direct examination of the patient-physician interaction would be necessary to determine the reasons for the discordance; however, they may differ by specialty. Furthermore, the discordance may also reflect the specifics of the documentation system. The Mayo Clinic uses an integrated EMR where all notes and forms (including the patient provided information forms) from all services for a given patient are available to any clinician working with the patient. Thus, clinician may decide to avoid repeating a symptom in his/her clinical note that is already documented elsewhere. Thus, the nature and the consequences of the interaction between the individual characteristics of the documentation systems, the care providers, and the patients warrant further investigation within the framework of patient-centered care.

Of interest are those instances where the patient reported a symptom not documented in the record (i.e., false negatives). Herein, 31% of patients with chest pain, 38% with dyspnea and 45% with cough did not have a positive mention of the symptom in the clinical notes. We could hypothesize that the outcomes of such patients may differ form those for whom symptoms were documented in the EMR. Conversely, when the patient does not report a symptom documented in the EMR, such discordance may reflect differential elicitation of symptoms by care providers by specialty. As these considerations remain speculative, further study is needed to examine the reasons for discordances and to determine if these affect clinical outcomes. The present study shows that the discordances exist, which has implications for clinical research that relies on the EMR to ascertain symptoms.

Limitations and strengths

The generalizability of the study depends on the availability of the EMR; however, the adoption of EMRs across the United States is growing[29]. Many medical centers already have their patient notes in electronic form and systems for accessing the information in the notes can be constructed even without a fully integrated EMR system. While the data sources used in this study are specific to the Mayo Clinic, collecting patient-reported information prior to visit and physician documentation of the visit are standard health care practices and similar resources are in use at other institutions across the United States. Thus, our results are generalizable to other institutions that are equipped with an EMR and can guide the improvement of existing EMR systems by suggesting new avenues for improved elicitation and capture of patient-specific information central to patient-care and research.

In this study, we did not look for an association between the literacy level or proficiency in the English language. We did not have these data available for this study but we recognize that these are important variables to consider since the forms require a certain level of knowledge of the English language. Care provider's demographic characteristics (as well as their interaction with patient characteristics) may also be important predictors of concordances and discordances in documentation. These characteristics were not available for the current study but will be considered in future work.

The use of the state-of-the-art EMR environment at the Mayo Clinic is a distinct strength of this study enabling an investigation of the discrepancies between patient reported symptoms and provider documentation. This study presents a novel approach that relies on automated and thus easily scalable investigation of the free text contained in provider documentation and thus serves to enable future large scale studies of the interaction between the patient and the care provider. Despite the fact that some natural language term variants (particularly those due to misspellings and non-standard descriptions) are likely to be missed by the automated natural language queries used in this project, our methodology is highly sensitive for identification of chest pain, dyspnea and cough.

Acknowledgements

We thank Ellen Koepsell and Diane Batzel for their dedicated efforts in annotating clinical notes. This work was supported in part by grants for the Public Health Service (RO1 HL 59205, HL 72435) (Bethesda, MD) and the National Institutes of Health Roadmap Multidisciplinary Clinical Research Career Development Award Grant (K12/NICHD)-HD49078 (Bethesda, MD).

Financial support: This work was supported by grants RO1-72435, GM14321 and AR30582, and the NIH Roadmap Multidisciplinary Clinical Research Career Development Award Grant (K12/NICHD-HD49078).

Abbreviations

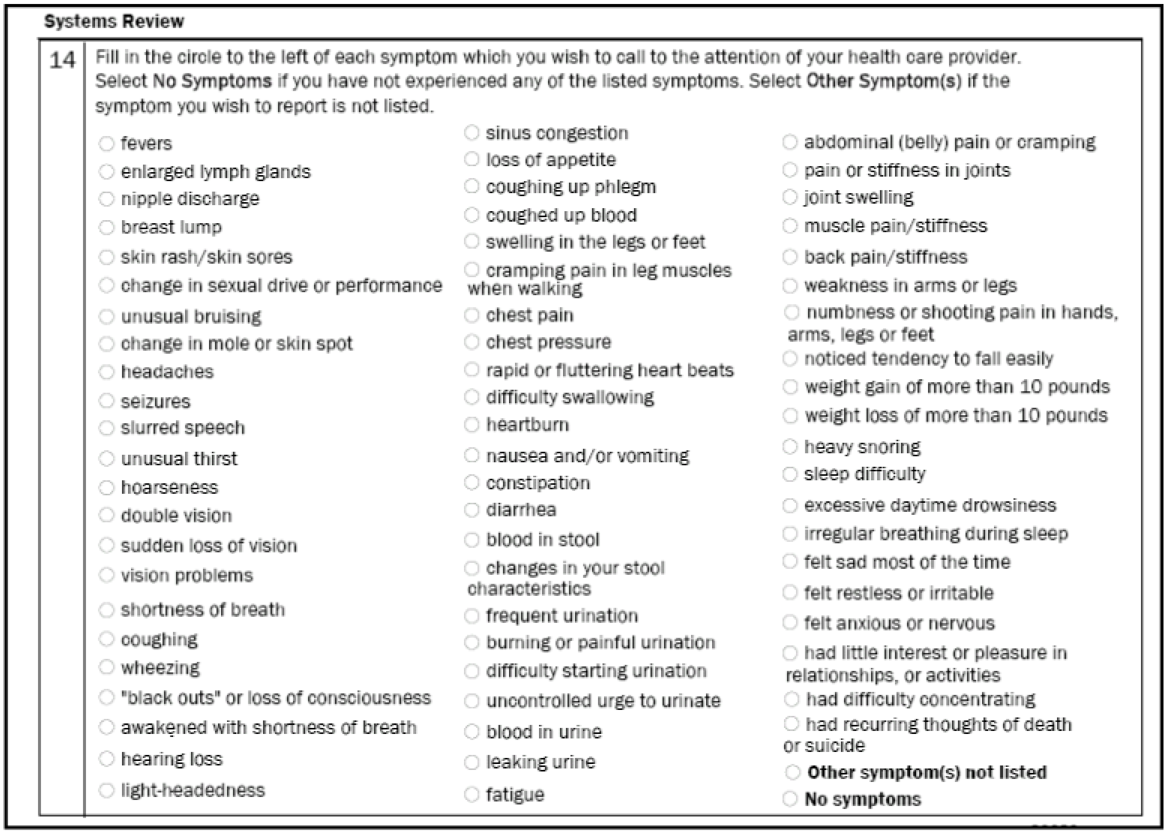
EMR, electronic medical record; ICD, international classification of disease; NLP, natural language processing; PND, paroxysmal nocturnal dyspnea.

References

1. Hing E, Cherry DK, Woodwell DA. National Ambulatory Medical Care Survey: 2004 summary. Adv Data 2006;(374):1–33.
2. Lamberg L. New mind/body tactics target medically unexplained physical symptoms and fears. JAMA 2005;294(17):2152–4. [PubMed: 16264150]
3. Crossing the Quality Chasm: A New Health System for the 21st Century. Institute of Medicine; Washington, DC: 2001.
4. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Inform Assoc 1994;1(2):142–60. [PubMed: 7719796]
5. Friedman C. A broad-coverage natural language processing system. Proc AMIA Symp 2000:270–4. [PubMed: 11079887]
6. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc 2005;12(4):448–57. [PubMed: 15802475]
7. Javitz HS, Ward MM, Watson JB, Jaana M. Cost of illness of chronic angina. Am J Manag Care 2004;10(11 Suppl):S358–69. [PubMed: 15603245]
8. Hemingway H, McCallum A, Shipley M, Manderbacka K, Martikainen P, Keskimaki I. Incidence and prognostic implications of stable angina pectoris among women and men. JAMA 2006;295(12):1404–11. [PubMed: 16551712]
9. Thom T, Haase N, Rosamond W, Howard VJ, Rumsfeld J, Manolio T. Heart Disease and Stroke Statistics--2006 Update. A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2006E-publication ahead of print

10. Philpott S, Boynton PM, Feder G, Hemingway H. Gender differences in descriptions of angina symptoms and health problems immediately prior to angiography: the ACRE study. Appropriateness of Coronary Revascularisation study. Soc Sci Med 2001;52(10):1565–75. [PubMed: 11314852]
11. Schulman KA, Berlin JA, Harless W, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med 1999;340(8):618–26. [PubMed: 10029647]
12. Hemingway H, Shipley M, Britton A, Page M, Macfarlane P, Marmot M. Prognosis of angina with and without a diagnosis: 11 year follow up in the Whitehall II prospective cohort study. BMJ 2003;327 (7420):895. [PubMed: 14563744]
13. Pakhomov S, Coden A, Chute C. Developing a Corpus of Clinical Notes Manually Annotated for Part-of-Speech. To appear in International Journal of Medical Informatics. 2005Special Issue on Natural Language Processing in Biomedical Applications
14. Ford EW, Menachemi N, Phillips MT. Predicting the adoption of electronic health records by physicians: when will health care be paperless? J Am Med Inform Assoc 2006;13(1):106–12. [PubMed: 16221936]
15. Pakhomov SS, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: role of natural language processing of the medical record. Am Heart J 2007;153(4): 666–73. [PubMed: 17383310]
16. GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. New England Journal of Medicine 1993;329:673–82. [PubMed: 8204123]
17. Second International Study of Infarct Survival Collaboartive Group. Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. The Lancet 1988;2:349–60.
18. Arciero T, Jacobsen S, Reeder G, et al. Temporal Trends in the Incidence of Coronary Disease. The American Journal of Medicine 2004;117:228–33. [PubMed: 15308431]
19. Wang SJ, Ohno-Machado L, Fraser H, Kennedy L. Using patient-reportable clinical history factors to predict myocardial infarction. Computers in Biology and Medicine 2001;31:1–13. [PubMed: 11058690]
20. Philpott S, Boynton P, Feder G, Hemingway H. Gender Differences in Descriptions of Angina Symptoms and Health Problems Immediately Prior to Angiography: the ACRE study. Social Science and Medicine 2001;52:1565–75. [PubMed: 11314852]
21. Schulman K, Berlin J, W H, et al. The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization. New England Journal of Medicine 1999;340:618–26. [PubMed: 10029647]
22. Priority Areas for National Action: Transforming Health Care Quality. Institute of Medicine; Washington, DC: 2003.
23. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. Journal of clinical epidemiology 1990;43(6):543–9. [PubMed: 2348207]
24. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. Journal of clinical epidemiology 1990;43(6):551–8. [PubMed: 2189948]
25. St Sauver JL, Hagen PT, Cha SS, et al. Agreement between patient reports of cardiovascular disease and patient medical records. Mayo Clinic proceedings 2005;80(2):203–10. [PubMed: 15704775]
26. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001;34(5):301–10. [PubMed: 12123149]
27. Dove JT. President's Page: Achieving the Heights of Quality With Electronic Health Records. J Am Coll Cardiol 2007;48:2218–22.
28. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care 2007;13(6 Part 1):277–8. [PubMed: 17567224]
29. Research finds low EHR adoption rates for physician groups. 2005. Accessed at http://www.mgma.com/press/EHR-adoptionstudy.cfm

NIH-PA Author Manuscript
NIH-PA Author Manuscript



Systems Review

14 Fill in the circle to the left of each symptom which you wish to call to the attention of your health care provider. Select **No Symptoms** if you have not experienced any of the listed symptoms. Select **Other Symptom(s)** if the symptom you wish to report is not listed.

- fevers
- enlarged lymph glands
- nipple discharge
- breast lump
- skin rash/skin sores
- change in sexual drive or performance
- unusual bruising
- change in mole or skin spot
- headaches
- seizures
- slurred speech
- unusual thirst
- hoarseness
- double vision
- sudden loss of vision
- vision problems
- shortness of breath
- coughing
- wheezing
- "black outs" or loss of consciousness
- awakened with shortness of breath
- hearing loss
- light-headedness
- sinus congestion
- loss of appetite
- coughing up phlegm
- coughed up blood
- swelling in the legs or feet
- cramping pain in leg muscles when walking
- chest pain
- chest pressure
- rapid or fluttering heart beats
- difficulty swallowing
- heartburn
- nausea and/or vomiting
- constipation
- diarrhea
- blood in stool
- changes in your stool characteristics
- frequent urination
- burning or painful urination
- difficulty starting urination
- uncontrolled urge to urinate
- blood in urine
- leaking urine
- fatigue
- abdominal (belly) pain or cramping
- pain or stiffness in joints
- joint swelling
- muscle pain/stiffness
- back pain/stiffness
- weakness in arms or legs
- numbness or shooting pain in hands, arms, legs or feet
- noticed tendency to fall easily
- weight gain of more than 10 pounds
- weight loss of more than 10 pounds
- heavy snoring
- sleep difficulty
- excessive daytime drowsiness
- irregular breathing during sleep
- felt sad most of the time
- felt restless or irritable
- felt anxious or nervous
- had little interest or pleasure in relationships, or activities
- had difficulty concentrating
- had recurring thoughts of death or suicide
- **Other symptom(s) not listed**
- **No symptoms**

Figure 1. "Review of Systems" part of a patient provided information form

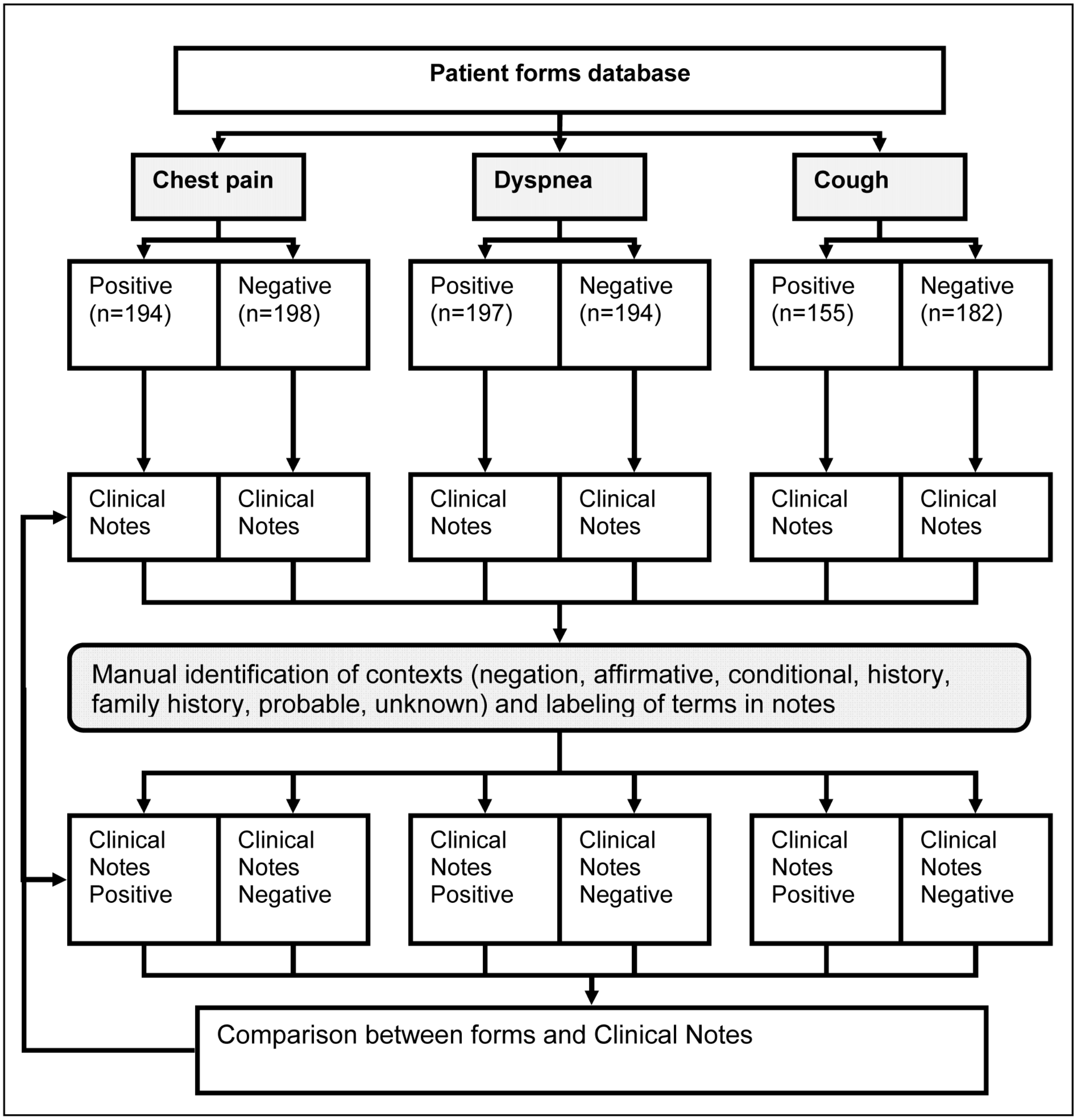


Figure 2. Diagram illustrating the study design

Chest pain	**PAIN** = (ache, aching, hurt, hurting, pain, pressure, discomfort, tightness, soreness, heaviness, squeezing, crushing, compression, tension, distress) **LOCATION** = (chest, breast†, thorax, rib cage, substernal, thoracic); **QUERY** = (**LOCATION PAIN**) or (**PAIN LOCATION**);
Dyspnea	**DIFFICULTY** = (difficulty, difficult, difficulties, hard, labored, strenuous, problem, gasping, short, shortness, out); **BREATH** = (breathing, breathes, breathe, breath)"; **ONEWORD** = (breathlessness, breathless, dyspnea, SOB)"; **QUERY** = ((**DIFFICULTY BREATH**) or (**BREATH DIFFICULTY**) or (**ONEWORD**))";
Cough	**QUERY** = (coughing, coughs, cough)

Figure 3. Natural language queries used to search clinical reports

†We recognize that, clinically, breast pain is distinct from chest pain; however, this study focuses on the patient presentation of symptoms. Thus including “breast pain” as a search term is motivated by the desire to capture clinician’s quoting of the patient as in the following real example: “I suspect that the patient’s breast pain is actually due to significant chest wall pain ...” In our sample, there were only 15 instances of breast pain; therefore, inclusion or exclusion of this term would not significantly affect the conclusions of this study.

Table 1

Distribution of contexts in which chest pain, dyspnea and cough occur

	Symptom		
Context	**Chest pain (N (%))**	**Dyspnea (N (%))**	**Cough (N (%))**
Affirmative	1178 (62)	1459 (65)	715 (57)
Negated	543 (29)	505 (23)	226 (18)
Probable	5 (< 1)	0	0
History	79 (4)	95 (4)	63 (5)
Family History	0	1 (< 1)	3 (< 1)
Conditional	78 (4)	162 (7)	243 (19)
Unknown	18 (< 1)	3 (< 1)	2 (< 1)
Total	1901 (100)	2225 (100)	1252 (100)

Table 2

Positive and negative agreement results as well as Kappa values for the comparison between patient self-report and care providers' notes (unstratified and stratified by age and sex)

	Chest Pain (n=373)			Dyspnea (n=391)			Cough (n=337)		
	Ppos	Pneg	Kappa (95%CI)	Ppos	Pneg	Kappa (95%CI)	Ppos	Pneg	Kappa (95%CI)
Unstratified									
	74	78	0.52 (0.44-0.60)	70	76	0.46 (0.37-0.54)	63	75	0.38 (0.28-0.48)
Sex									
Male	76	74	0.51 (0.38-0.64)	75	76	0.52 (0.40-0.65)	63	74	0.38 (0.24-0.53)
Female	72	76	0.48 (0.37-0.60)	64	75	0.40 (0.28-0.52)	61	75	0.37 (0.24-0.51)
Age									
18-34	65	79	0.44 (0.17-0.70)	62	85	0.48 (0.21-0.75)	74	81	0.56 (0.27-0.84)
35-49	78	76	0.55 (0.37-0.73)	49	76	0.27 (0.02-0.52)	45	77	0.23 (0.04-0.50)
50-64	76	74	0.51 (0.35-0.66)	69	75	0.45 (0.28-0.61)	50	79	0.31 (0.12-0.49)
>=65	71	76	0.47 (0.32-0.62)	75	72	0.47 (0.34-0.60)	69	66	0.37 (0.23-0.52)

Table 3

NLP context (negation) identification algorithm validation results

	N correctly identified	Sensitivity (%)	N correctly identified	Specificity (%)	Kappa
	Manual Annotation				
	CHEST PAIN				
	negated (n = 461)		**affirmative (n = 1188)**		
	404	88 (84-90)	1095	92 (91-94)	0.78 (0.75-0.81)
	DYSPNEA				
	negated (n = 460)		**affirmative (n = 1517)**		
NLP	385	84 (80-87)	1406	93 (91-94)	0.74 (0.71-0.78)
	COUGH				
	negated (n = 206)		**affirmative (n = 876)**		
	171	83 (78-88)	798	91 (90-93)	0.69 (0.63-0.74)

Table 4

A 2×2 table used to compute agreement measures

		Patient reported symptom		
		positive	**negative**	Marginal total
Clinical Notes symptom	**positive**	a	b	a+b
	negative	c	d	c+d
	Marginal total	a+c	b+d	N=a+b+c+d

Table 5

Proportion of notes generated by various services for patients with symptoms noted either on self-report and/or in the text of the EMR

	Clinical Specialty*										
	PSY	**PMR**	**GI**	**PC**	**HEM**	**NEURO**	**CV**	**EDPT**	**SURG**	**ENDO**	**Total**
	Chest pain										
TP(%)	5	18	3	29	6	4	25	3	4	3	100
FN(%)	2	4	5	**29**	2	8	**24**	1	8	17	100
TN(%)	6	8	9	39	3	6	5	2	14	9	100
FP(%)	**23**	1	0	**24**	18	5	16	2	7	3	100
	Dyspnea										
TP(%)	2	2	2	29	6	5	41	4	4	6	100
FN(%)	12	6	8	**23**	3	13	13	2	3	**17**	100
TN(%)	7	7	5	42	1	9	6	3	15	4	100
FP(%)	1	4	2	**46**	2	1	**27**	5	4	8	100
	Cough										
TP(%)	3	3	6	35	8	3	31	6	1	4	100
FN(%)	7	**21**	3	**31**	2	12	19	1	1	3	100
TN(%)	8	10	5	37	3	8	12	2	8	7	100
FP(%)	8	3	1	**21**	**35**	3	18	4	7	1	100

* Legend: PSY - psychology/psychiatry; PMR - physical medicine & rehabilitation; GI - gastroenterology; PC - primary care; HEM - hematology; NEURO - neurology; CV - cardiovascular; EDPT - emergency medicine; SURG - surgery; ENDO - endocrinology. TP - true positives (patient reported symptom and it was documented); FN - false negatives (patient reported symptom but not documented); TN - true negatives (patient did not report a symptom and no documentation found); FP - false positives (patient did not report a symptom but the symptom was found in documentation). The percentages highlighted in bold indicate the top 2 services with highest proportion of notes for discordant results.