Agreement Between Patient-reported Symptoms and Their Documentation in the Medical Record

Serguei V. Pakhomov, PhD; Steven J. Jacobsen, MD, PhD; Christopher G. Chute, MD, DrPH; and Véronique L. Roger, MD, MPH

Physical symptoms account for half of all outpatient visits in the United States\(^1\) 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 (EMR), free text of the clinical history can now be subjected to automated analysis in ways that are impossible or uneconomic with paper-based records.\(^4,5,6\) 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 1 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 patients with stable angina remain unclear. Many patients with typical symptoms are not diagnosed with angina,\(^10\) and age, sex, and ethnicity may influence physicians’ recommendations for diagnostic testing such as coronary angiography and the resulting International Classification of Diseases 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\(^15\) who were missed by traditional diagnostic coding approaches. Natural language processing 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.\(^19\) Because the diagnosis of angina pectoris relies on the patient conveying the symptoms to the physician,\(^12\) the ambiguous nature of verbal communication as well as the nature of coronary artery disease (ie, its presentation varies according to patient race and sex) can make the diagnosis challenging. These challenges can lead to inconsistency and incompleteness in both the diagnosis and the information recorded in the EMR.\(^11\) 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

For author information and disclosures, see end of text.
Patient-reported Symptoms in the EMR

critical implications for clinical care and may have a significant impact on patient outcomes.

The primary objective of the current study is to determine whether there was disagreement between patient-reported symptoms of chest pain, dyspnea, and cough and the documentation of these symptoms by care providers in the EMR. Although in this study we focus on heart disease, which constitutes one of the top national healthcare priority areas, 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 2 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. The Mayo Clinic maintains an EMR for each inpatient and outpatient. A large part of the EMR consists of clinical notes, which represent transcriptions by trained medical transcriptionists of the dictations recorded by care providers after each contact with the patient. 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 Clinical Document Architecture, and consists of standard sections including chief complaint, history of present illness, review of systems, and impression/report/plan, among others. This standard is used by most major EMR vendors in the United States.

Current Visit Information Forms. Many healthcare providers, including the Mayo Clinic, ask their patients to fill out forms detailing their prior health and social history, current symptoms, medications, and allergies. The Mayo Clinic’s form is reproduced in the eAppendix (available at www.ajmc.com). Figure 1 shows item 14 of this patient-provided information form, which asks the patient to indicate if he/she has ever experienced a symptom listed on the form. Subsequently, the 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 in-
formation 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 were not restricted to any geographic location and represent general Mayo Clinic ambulatory and hospitalized populations. We selected records that indicated either a positive or a negative response to questions about chest pain, chest pressure, shortness of breath/dyspnea, and cough. (These codes are internal identifiers provided here for reference and have no relation to any standardized nomenclature of medical concepts.) The analysis presented here is based on a combined symptom for chest pain that includes both the 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 1 of their forms and 200 patients who did not mark the symptom on any of their forms. No restrictions other than age ≥18 years were applied. Only records of patients who had 1 or more clinical notes dictated during the study period (January 1, 2006, through 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 (Figure 3). The keywords and methods were similar to those previously reported. For example, a natural language query for “chest pain” identifies portions of clinical note text where 1 of the terms describing PAIN (eg, pressure) either precedes or follows 1 of the terms describing the LOCATION (eg, formation 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 were not restricted to any geographic location and represent general Mayo Clinic ambulatory and hospitalized populations. We selected records that indicated either a positive or a negative response to questions about chest pain, chest pressure, shortness of breath/dyspnea, and cough. (These codes are internal identifiers provided here for reference and have no relation to any standardized nomenclature of medical concepts.) The analysis presented here is based on a combined symptom for chest pain that includes both the 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 1 of their forms and 200 patients who did not mark the symptom on any of their forms. No restrictions other than age ≥18 years were applied. Only records of patients who had 1 or more clinical notes dictated during the study period (January 1, 2006, through 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 (Figure 3). The keywords and methods were similar to those previously reported. For example, a natural language query for “chest pain” identifies portions of clinical note text where 1 of the terms describing PAIN (eg, pressure) either precedes or follows 1 of the terms describing the LOCATION (eg, formation 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 were not restricted to any geographic location and represent general Mayo Clinic ambulatory and hospitalized populations. We selected records that indicated either a positive or a negative response to questions about chest pain, chest pressure, shortness of breath/dyspnea, and cough. (These codes are internal identifiers provided here for reference and have no relation to any standardized nomenclature of medical concepts.) The analysis presented here is based on a combined symptom for chest pain that includes both the 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 1 of their forms and 200 patients who did not mark the symptom on any of their forms. No restrictions other than age ≥18 years were applied. Only records of patients who had 1 or more clinical notes dictated during the study period (January 1, 2006, through 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 (Figure 3). The keywords and methods were similar to those previously reported. For example, a natural language query for “chest pain” identifies portions of clinical note text where 1 of the terms describing PAIN (eg, pressure) either precedes or follows 1 of the terms describing the LOCATION (eg,
Patient-reported Symptoms in the EMR

In addition to 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 with those identified with automated natural language queries to determine their sensitivity.

Angina refers to a complex of symptoms, 1 of which may be chest pain or discomfort. Similarly to dyspnea, orthopnea and paroxysmal nocturnal dyspnea (PND) 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 whether these related terms had 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 and negative agreement rather than sensitivity and specificity because 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. Positive and negative agreement measures have been proposed as a way to ensure the correct interpretation of kappa values and have been used to assess the agreement between patient-reported information and the medical record. The computation of these measures is illustrated in Table 2, which represents a 2 × 2 table containing the counts where the patient responses were in concordance or discordance.

### Table 1. Distribution of Contexts in Which Chest Pain, Dyspnea, and Cough Occurred

<table>
<thead>
<tr>
<th>Context</th>
<th>Chest Pain</th>
<th>Dyspnea</th>
<th>Cough</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>1178 (62)</td>
<td>1459 (65)</td>
<td>715 (57)</td>
</tr>
<tr>
<td>Negated</td>
<td>543 (29)</td>
<td>505 (23)</td>
<td>226 (18)</td>
</tr>
<tr>
<td>Probable</td>
<td>5 (&lt;1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>History</td>
<td>79 (4)</td>
<td>95 (4)</td>
<td>63 (5)</td>
</tr>
<tr>
<td>Family history</td>
<td>0</td>
<td>1 (&lt;1)</td>
<td>3 (&lt;1)</td>
</tr>
<tr>
<td>Conditional</td>
<td>78 (4)</td>
<td>162 (7)</td>
<td>243 (19)</td>
</tr>
<tr>
<td>Unknown</td>
<td>18 (&lt;1)</td>
<td>3 (&lt;1)</td>
<td>2 (&lt;1)</td>
</tr>
<tr>
<td>Total</td>
<td>1901 (100)</td>
<td>2225 (100)</td>
<td>1252 (100)</td>
</tr>
</tbody>
</table>

*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.
dance 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: $100 \times \frac{2a}{2a + (a - d)}$. 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: $100 \times \frac{2d}{2d + (a - d)}$. In addition we also report on the false-negative rate, which in our case shows the proportion of times when the patient reported a symptom but the symptom did not appear in the physician’s note [ie, false-negative rate = $100 \times \frac{c}{a + c}$]. 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 used a manually created reference standard. Stratified analyses in sex and age subgroups were performed as well.

### RESULTS

Of the 121,891 patients who filled out a patient-provided information form during 6 months between January 1, 2006, and June 30, 2006, 6569 patients (5.39%; 95% confidence interval [CI] = 5.26, 5.52) reported chest pain, 6166 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.

Random sampling resulted in a study population of 1119 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%) in more than 1 group. Of these 1119 patients, 499 (44.6%) were male, 127 (11.3%) were between age 18 and 34 years, 185 (16.5%) were between age 35 and 49 years, 337 (30.1%) were between age 50 and 64 years, and 470 (42.0%) were age 65 years or older. Twelve percent of the clinical notes for the 1119 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 1 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%), 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 29% of the time. These observations are consistent with those of Chapman et al24; in their study, 27% of the 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 were 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. The specificity was 91% (95% CI = 90, 93), 93% (95% CI = 91, 94), and 92% (95% CI = 91, 94), for cough, dyspnea, and chest pain, respectively. 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.

---

### Table 2. Natural Language Queries Used to Search Clinical Reports

<table>
<thead>
<tr>
<th>Patient-reported Symptom</th>
<th>Clinical Notes Symptom</th>
<th>Positive</th>
<th>Negative</th>
<th>Marginal total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>$a$</td>
<td>$b$</td>
<td>$a + b$</td>
</tr>
<tr>
<td>Negative</td>
<td>$b$</td>
<td>$d$</td>
<td>$c + d$</td>
<td></td>
</tr>
<tr>
<td>Marginal total</td>
<td>$a + c$</td>
<td>$b + d$</td>
<td>$N = a + b + c + d$</td>
<td></td>
</tr>
</tbody>
</table>

---

### Table 3. Sensitivity and Specificity of the NLP Negation Algorithm

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough</td>
<td>83% (78, 88)</td>
<td>91% (90, 93)</td>
<td>0.69 (0.63, 0.74)</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>84% (80, 87)</td>
<td>93% (91, 94)</td>
<td>0.74 (0.71, 0.78)</td>
</tr>
<tr>
<td>Chest pain</td>
<td>88% (84, 90)</td>
<td>92% (91, 94)</td>
<td>0.78 (0.75, 0.81)</td>
</tr>
</tbody>
</table>
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 also were identified by automatic queries.

**Agreement Between Clinical Notes and Patient-provided Information Forms**

The results are summarized in Table 4. Overall, the positive agreement for chest pain, dyspnea, and cough was 74, 70, and 63, respectively, and the negative agreement was 78, 76, and 75, respectively. 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 whether the concordance and discordance are influenced by these factors. Positive agreement was slightly lower for female patients across all 3 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 to 0.55 for chest pain, 0.27 to 0.52 for dyspnea, and 0.31 to 0.56 for cough.

Including nonnegated 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 nonnegated mentions of orthopnea and PND resulted in identifying 1 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 3 symptoms (Table 5). 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 were the top 2 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 among primary care, cardiovascular services, and endocrinology. A similar distribution was found for dyspnea. For cough, the majority of the notes were distributed between primary care, cardiovascular services, and physical medicine and rehabilitation services. The majority of the notes for patients who had documentation of chest pain in the EMR but did not report it on patient-provided information forms were distributed among primary care, psychology/psychiatry, hematology, and cardiovascular services. A similar distribution was found for cough, with a large proportion of notes from the hematology service. For dyspnea, the majority of the notes originated from primary care and cardiovascular services.

**DISCUSSION**

Electronic medical records 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 healthcare providers. 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 also are important.26 From the standpoint of the primary use of the EMR, we currently do not have the data to determine whether the discordances between the symptoms reported by patients and those documented by care providers had any significant clinical consequences. However, we do present evidence that effective secondary uses of the EMR require examination of the communication approaches of care providers and patients.

Our findings indicate substantial discordance between patient reporting and care provider documentation of the symptoms. The 2 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 also may be necessary to use both sources of information for applications such as postmarketing medication safety surveillance, where it may be necessary to detect alarming trends in symptoms to prevent more serious adverse drug reactions. The current study is the first step toward 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 whether there was 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 would require only minor modifications to the algorithms used for term identification. Inclusion of the terms “angina,” “orthopnea,” and “PND” resulted in very minor changes in the positive and negative agreement, 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 showed 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.23 The study

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Chest Pain (n = 373)</th>
<th>Dyspnea (n = 391)</th>
<th>Cough (n = 337)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ppos</td>
<td>Pneg</td>
<td>Kappa (95% CI)</td>
</tr>
<tr>
<td>Unstratified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ppos</td>
<td>74</td>
<td>78</td>
<td>0.52 (0.44, 0.60)</td>
</tr>
<tr>
<td>Pneg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stratified by sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>76</td>
<td>74</td>
<td>0.51 (0.38, 0.64)</td>
</tr>
<tr>
<td>Female</td>
<td>72</td>
<td>76</td>
<td>0.48 (0.37, 0.60)</td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>65</td>
<td>79</td>
<td>0.44 (0.17, 0.70)</td>
</tr>
<tr>
<td>35-49</td>
<td>78</td>
<td>76</td>
<td>0.55 (0.37, 0.73)</td>
</tr>
<tr>
<td>50-64</td>
<td>76</td>
<td>74</td>
<td>0.51 (0.35, 0.66)</td>
</tr>
<tr>
<td>≥65</td>
<td>71</td>
<td>76</td>
<td>0.47 (0.32, 0.62)</td>
</tr>
</tbody>
</table>

CI indicates confidence interval; Pneg, negative agreement; Ppos, positive agreement.
Patient-reported Symptoms in the EMR

by St. Sauver et al, however, did not address symptom reporting and documentation, as that was not in the scope of their research questions. Although 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 St. Sauver et al), the situation is reversed for patient symptoms in cases where the patient is the primary source and the care provider is the secondary source. 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.

Although we found discordance between patient-reported 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 noncardiovascular services. Our data shows that although 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 who reported chest pain not documented in the EMR originated in 3 services: cardiovascular, endocrinology, and primary care. Direct examination of the patient-physician interaction would be necessary to determine the reasons for the discordance; however, the reasons may differ by specialty. Furthermore, the discordance also may 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. Clinicians may decide to avoid repeating a symptom in their 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 (ie, false negatives); 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

| 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
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Specialty, %</td>
</tr>
<tr>
<td>Result</td>
</tr>
<tr>
<td>Chest Pain</td>
</tr>
<tr>
<td>TP</td>
</tr>
<tr>
<td>FN</td>
</tr>
<tr>
<td>TN</td>
</tr>
<tr>
<td>FP</td>
</tr>
<tr>
<td>Dyspnea</td>
</tr>
<tr>
<td>TP</td>
</tr>
<tr>
<td>FN</td>
</tr>
<tr>
<td>TN</td>
</tr>
<tr>
<td>FP</td>
</tr>
<tr>
<td>Cough</td>
</tr>
<tr>
<td>TP</td>
</tr>
<tr>
<td>FN</td>
</tr>
<tr>
<td>TN</td>
</tr>
<tr>
<td>FP</td>
</tr>
</tbody>
</table>

CV indicates cardiovascular; EDPT, emergency medicine; EMR, electronic medical record; ENDO, endocrinology; FN, false negative (patient reported symptom but not documented); FP, false positive (patient did not report a symptom but the symptom was found in documentation); GI, gastroenterology; HEM, hematology; NEURO, neurology; PC, primary care; PMR, physical medicine & rehabilitation; PSY, psychology/psychiatry; SURG, surgery; TN, true negative (patient did not report a symptom and no documentation found); TP, true positive (patient reported symptom and it was documented). The percentages highlighted in blue indicate the top 2 services with highest proportion of notes for discordant results.
that the outcomes of such patients may differ from 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 whether 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.27 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. Although the data sources used in this study are specific to the Mayo Clinic, collecting patient-reported information before a visit and physician documentation of the visit are standard healthcare 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 because the forms require a certain level of knowledge of the English language. Care providers’ demographic characteristics (as well as the interaction of these characteristics with patient characteristics) also may 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; this approach can be used in 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 nonstandard 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.

Acknowledgments

We thank Ellen Koepsell and Diane Batal for their dedicated efforts in annotating clinical notes.

Author Affiliations: From the Department of Pharmaceutical Care and Health Systems, University of Minnesota (SVP), Minneapolis; the Department of Research and Evaluation, Kaiser Permanente (SJJ), Pasadena, CA; and the Department of Health Sciences Research (CGC) and the Division of Cardiovascular Diseases (VLR), Mayo Clinic, Rochester, MN.

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

Author Disclosure: The authors (SVP, SJJ, CGC, VLR) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (SVP, SJJ, CGC); acquisition of data (SVP, VLR); analysis and interpretation of data (SVP, SJJ, VLR); drafting of the manuscript (SVP, VLR); critical revision of the manuscript for important intellectual content (SVP, SJJ, CGC); statistical analysis (SVP, VLR); provision of study materials or patients (SVP); obtaining funding (SVP); administrative, technical, or logistic support (SVP, CGC); and supervision (SVP, CGC, VLR).

Address correspondence to: Serguei V. Pakhomov, PhD, Department of Pharmaceutical Care and Health Systems, University of Minnesota, 308 Harvard St, SE, 7-125F Weaver-Densford Hall, Minneapolis, MN 55401. E-mail: pakhov002@umn.edu.

REFERENCES

Patient-reported Symptoms in the EMR


Current Visit Information

This questionnaire collects information about your current state of health to assist the medical team with your care and help us meet requirements established by Medicare and other insurers. This information will be stored electronically in your medical record.

Using a black pen, please answer ALL questions by filling in the appropriate circle(s) like ●, and by PRINTING in the appropriate box.

Patient Information - if any of this information is incorrect, please inform the receptionist.

1. Patient Name: (Last, First, Middle Initial)  Mayo Clinic Number:  Gender:  Date of Birth:

2. Who completed this form?
   ○ Patient  ○ Family Member: Parent, Spouse, Son, Daughter, etc.  ○ Domestic Partner  ○ Legal Guardian  ○ Other

3. Do you have a living will or other advance directive?
   ○ Yes, up to date and on file at Mayo Clinic Rochester.
   ○ No, not interested.
   ○ Yes, but may need to be updated at Mayo Clinic Rochester.
   ○ No, but would like more information.
   ○ I do not know.  
   ➤ Additional information is available at the Patient Education Center, or ask the receptionist. (If hospitalized, ask nurse.)

4. Do you have cultural or religious preferences that you feel we should know about during your care?  ○ No  ○ Yes

5. Do you have any special dietary preferences or requirements?  ○ No  ○ Yes

Health Care Provider Information

6. Do you have a health care provider outside of Mayo Clinic Rochester? If yes, please fill out the following information and proceed to question 7 on page 2.

Name of Requesting Provider - the provider outside of Mayo Clinic Rochester who requested or arranged your evaluation at Mayo Clinic.

Date last seen by this provider

mm/dd/yyyy

Street Address

Phone - (area code) and number
(   )

Fax - (area code) and number
(   )

City

State or Province  Country  ZIP Code or Postal Code

Primary Provider (if different from requesting provider) - the provider outside of Mayo Clinic Rochester who manages your total care or to whom you go for routine health problems.

Date last seen by this provider

mm/dd/yyyy

Street Address

Phone - (area code) and number
(   )

Fax - (area code) and number
(   )

City

State or Province  Country  ZIP Code or Postal Code

Continue on Next Page

MC0985-23rev0605

23623

Copyright 2004 Mayo Press
### Medications

<table>
<thead>
<tr>
<th>Name of Medication</th>
<th>Strength of Each Dose</th>
<th>Number of Doses at a Time</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: <strong>Amoxicillin</strong></td>
<td>250 mg</td>
<td>1 pill</td>
<td>3 times daily</td>
</tr>
</tbody>
</table>

If you are taking more medications, please discuss with your provider.

8. Have you taken aspirin-containing products in the last 14 days?  
   - No  
   - Yes

9. Have you taken any oral, inhaled, injected steroid or cortisone-type drugs in the past 30 days?  
   - No  
   - Yes

10. Have you ever been instructed or advised to take antibiotics prior to any dental work or medical procedures?  
    - No  
    - Yes

11. Have you ever been told NOT to take or to STOP taking a medication?  
    - No  
    - Yes

Continue on Next Page
### Allergies

12 Fill in the circle if you have ever had an allergy or sensitivity to each of the following items:

- Latex or rubber
- Specific foods
- Influenza (flu) vaccination
- Adhesive tape
- Betadine or skin disinfectant
- Iodine or X-ray contrast dye
- Other vaccines - Tetanus, Measles, Polio, etc.
- "-caine" anesthetics such as Xylocaine, Novocaine
- I have other allergies not listed.
- No allergy to any of these items

13 List all medications, substances, foods, dusts, fumes, and animals to which you have allergies or unpleasant side effects.

<table>
<thead>
<tr>
<th>List Drug or Item</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: Sulfa</td>
<td>rash</td>
</tr>
</tbody>
</table>

### 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
## Systems Review, continued

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>15</strong> Are you having difficulty with pain?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>16</strong> Have you had the following immunizations:</td>
<td></td>
</tr>
<tr>
<td>Pneumococcal (for pneumonia)?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td>Tetanus/Diphtheria within the last 10 years?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td>Influenza shot within the last 12 months?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>17</strong> Have you ever had a colon or rectum examination?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td>If yes, how long ago was it?</td>
<td>○ Less than 1 year ○ 1-3 years ○ 4-5 years ○ Over 5 years</td>
</tr>
<tr>
<td><strong>18</strong> Do you feel you might be at risk for HIV or AIDS?</td>
<td>○ No ○ Yes</td>
</tr>
<tr>
<td><strong>19</strong> Have you ever had tuberculosis (TB) or had exposure to someone who had TB?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>20</strong> Do you have a communicable infectious disease (such as hepatitis)?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
</tbody>
</table>

### Questions 21-31 are to be completed for FEMALE patients ONLY.

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>21</strong> How long has it been since your last PAP smear and pelvic exam?</td>
<td>○ Less than 1 year ○ 1 to 2 years ○ More than 2 years ○ 4-5 years ○ Never had a PAP smear ○ Don't know</td>
</tr>
<tr>
<td><strong>22</strong> Have you <strong>EVER</strong> had an abnormal PAP smear?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>23</strong> How long has it been since your last mammogram?</td>
<td>○ Less than 1 year ○ 1 to 2 years ○ More than 2 years ○ Never had a mammogram ○ Don't know</td>
</tr>
<tr>
<td><strong>24</strong> How many pregnancies have you had?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8, 9+</td>
</tr>
<tr>
<td><strong>25</strong> How many live births have you delivered?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8, 9+</td>
</tr>
<tr>
<td><strong>26</strong> Might you be pregnant at this time?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>27</strong> Are you currently breast-feeding?</td>
<td>○ No ○ Yes</td>
</tr>
<tr>
<td><strong>28</strong> Have your menstrual periods changed in any way or become abnormal to you?</td>
<td>○ No ○ Yes ○ No, I am past menopause ○ No, I have had a hysterectomy ○ Yes</td>
</tr>
<tr>
<td><strong>29</strong> Are you experiencing unusual vaginal bleeding?</td>
<td>○ No ○ Yes</td>
</tr>
<tr>
<td><strong>30</strong> Have you had a tubal ligation?</td>
<td>○ No ○ Yes</td>
</tr>
<tr>
<td><strong>31</strong> Date of onset of your last menstrual period mm/dd/yyyy</td>
<td></td>
</tr>
</tbody>
</table>

### Questions 32-34 to be completed by MALE patients ONLY.

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>32</strong> How long has it been since your last prostate exam?</td>
<td>○ Less than 1 year ○ 1 to 2 years ○ More than 2 years ○ Never had a prostate exam ○ Don't know</td>
</tr>
<tr>
<td><strong>33</strong> Have you <strong>EVER</strong> had an abnormal prostate exam?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
<tr>
<td><strong>34</strong> Have you had a vasectomy?</td>
<td>○ No ○ Yes ○ Don't know</td>
</tr>
</tbody>
</table>
### Social Context

35. Select the highest level of schooling you have completed:
- ○ 8th grade or less
- ○ High school graduate or GED
- ○ 4-year college graduate
- ○ Some high school, but didn't graduate
- ○ Some college or 2-year degree
- ○ Post graduate studies

36. What is your current employment status?
- ○ Employed
- ○ Unemployed
- ○ Work disabled
- ○ Student
- ○ Retired
- ○ Self-employed
- ○ Full-time homemaker
- ○ Other

37. List most recent occupation:

38. What is your current relationship status?
- ○ Divorced
- ○ Married
- ○ Separated
- ○ Single
- ○ Widowed
- ○ Committed relationship
- ○ Other

39. Has your relationship status changed in the past 12 months?  ○ No  ○ Yes

40. Do you ever feel afraid in your home?  ○ No  ○ Yes  ○ Don't know

### Substance Use History

41. Have you ever felt the need to cut down on your alcohol consumption?  ○ No  ○ Yes

42. Do relatives/friends worry or complain about your alcohol consumption?  ○ No  ○ Yes

43. Do you currently smoke or use other tobacco products?
- ○ No, never used any
- ○ No, I quit all use
- ○ Yes, if yes, mark all that apply:  ○ Cigarettes  ○ Pipe  ○ Cigar  ○ Chewing tobacco  ○ Other

44. If you previously smoked or used other tobacco products and have quit, how long ago did you quit?
- ○ Within the past 30 days
- ○ 1-12 months ago
- ○ 2-3 years ago
- ○ 4-10 years ago
- ○ 11 or more years ago

45. If you do smoke or use other tobacco products, how motivated are you to quit?
- ○ Not motivated
- ○ Somewhat motivated
- ○ Very motivated

### Self-Care / Home Environment Assessment

46. Can you climb two flights of stairs without stopping to rest?
- ○ Yes, with no difficulty
- ○ Yes, with difficulty
- ○ No, can't do at all
- ○ Don't know

47. Are you dependent on a device for normal breathing (CPAP, nasal oxygen)?  ○ No  ○ Yes

48. Fill in the circle to the left of each activity which you have difficulty performing on your own:
- ○ Preparing meals
- ○ Using the toilet
- ○ Bathing
- ○ Getting in and out of bed
- ○ Feeding yourself
- ○ Housekeeping
- ○ Walking
- ○ Managing medications
- ○ Dressing
- ○ Climbing stairs
- ○ Using transportation
- ○ No difficulty with any of these activities

49. Which of the following describes your living environment?
- ○ House
- ○ Apartment
- ○ Assisted Living
- ○ Nursing Home
- ○ Other

50. With whom do you live?
- ○ Live alone
- ○ Spouse
- ○ Domestic partner
- ○ Family
- ○ Other

51. Do you have assistance for your home care from family, friends, or others should you require it?  ○ No  ○ Yes

52. Do you depend on any assistive devices such as a cane, wheelchair, braces, walker, or assistance from other people to perform activities important to you in your daily life?  ○ No  ○ Yes

53. Do you wear dentures?  ○ No  ○ Yes

54. Do you have hearing aid(s)?  ○ No  ○ Yes

See the instruction page for information about returning this form.