

# Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial

Adam Wright,<sup>1,2,3</sup> Justine Pang,<sup>1,2</sup> Joshua C Feblowitz,<sup>1,2</sup> Francine L Maloney,<sup>2</sup> Allison R Wilcox,<sup>2</sup> Karen Sax McLoughlin,<sup>1</sup> Harley Ramelson,<sup>1,2,3</sup> Louise Schneider,<sup>1,3</sup> David W Bates<sup>1,2,3,4</sup>

► An additional appendix is published online only. To view this file please visit the journal online ([www.jamia.bmj.com/content/19/4.toc](http://www.jamia.bmj.com/content/19/4.toc)).

<sup>1</sup>Division of General Internal Medicine, Brigham & Women's Hospital, Boston, Massachusetts, USA

<sup>2</sup>Partners HealthCare, Boston, Massachusetts, USA

<sup>3</sup>Harvard Medical School, Boston, Massachusetts, USA

<sup>4</sup>Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts, USA

## Correspondence to

Dr Adam Wright, Brigham and Women's Hospital, 1620 Tremont St, Boston, MA 02115, USA; [awright5@partners.org](mailto:awright5@partners.org)

Received 2 August 2011

Accepted 5 December 2011

Published Online First

3 January 2012

## ABSTRACT

**Background** Accurate clinical problem lists are critical for patient care, clinical decision support, population reporting, quality improvement, and research. However, problem lists are often incomplete or out of date.

**Objective** To determine whether a clinical alerting system, which uses inference rules to notify providers of undocumented problems, improves problem list documentation.

**Study Design and Methods** Inference rules for 17 conditions were constructed and an electronic health record-based intervention was evaluated to improve problem documentation. A cluster randomized trial was conducted of 11 participating clinics affiliated with a large academic medical center, totaling 28 primary care clinical areas, with 14 receiving the intervention and 14 as controls. The intervention was a clinical alert directed to the provider that suggested adding a problem to the electronic problem list based on inference rules. The primary outcome measure was acceptance of the alert. The number of study problems added in each arm as a pre-specified secondary outcome was also assessed. Data were collected during 6-month pre-intervention (11/2009–5/2010) and intervention (5/2010–11/2010) periods.

**Results** 17 043 alerts were presented, of which 41.1% were accepted. In the intervention arm, providers documented significantly more study problems (adjusted OR=3.4,  $p<0.001$ ), with an absolute difference of 6277 additional problems. In the intervention group, 70.4% of all study problems were added via the problem list alerts. Significant increases in problem notation were observed for 13 of 17 conditions.

**Conclusion** Problem inference alerts significantly increase notation of important patient problems in primary care, which in turn has the potential to facilitate quality improvement.

**Trial Registration** [ClinicalTrials.gov](http://ClinicalTrials.gov): NCT01105923.

## INTRODUCTION AND BACKGROUND

An accurate and up-to-date patient problem list represents the cornerstone of the problem-oriented medical record, especially in internal medicine. It serves as a valuable tool for providers attempting to familiarize themselves with a patient's clinical status and provides a means of succinctly communicating this information between providers. In addition, an accurate problem list has been associated with higher-quality care.<sup>1</sup> For example,

Hartung *et al* found that patients with 'congestive heart failure' (CHF) on their problem list were more likely to receive ACE inhibitors or angiotensin-II receptor blockers than CHF patients without 'CHF' listed on their problem list. Further, many clinical decision support (CDS) rules use problem list entries to make inferences about patients,<sup>2</sup> so a complete, accurate list may facilitate more effective CDS. Conversely, an incomplete or inaccurate problem list could lead to delayed or inappropriate care. Finally, an accurate and comprehensive problem list would help to correctly identify patient populations and create patient registries for the conduct of quality improvement activities and research.

Despite these numerous benefits, problem lists are often inaccurate, incomplete, and out of date.<sup>3–5</sup> In previous research, we showed that problem list completeness in one network ranged from 4.7% for renal insufficiency or failure to 50.7% for hypertension, 61.9% for diabetes, to a maximum of 78.5% for breast cancer,<sup>6</sup> and other institutions have found similar results.<sup>3–5</sup> In addition, we have found in previous qualitative studies that provider attitudes toward, and use of, the problem list vary widely.<sup>7, 8</sup>

Beginning in 2011, in order to be considered 'meaningful users' of an electronic health record (EHR) and qualify to receive federal stimulus grants under the HITECH Act, which can total US \$44 000 through Medicare and US\$63 750 through Medicaid, providers must, among other things, 'maintain an up-to-date problem list of current and active diagnoses,' with 80% of patients having at least one problem recorded or an indication of 'no known problems.'<sup>9–11</sup> Given wide variation in problem list use by providers,<sup>7, 8</sup> new tools are needed to help providers meet this goal.

Researchers have used a variety of strategies in an attempt to detect patient problems and increase problem list use. In general, these methods fall into two broad categories: problem inference (or proxy) rules and natural language processing (NLP) techniques. Problem inference techniques use related clinical information such as laboratory tests, medications, and billing codes to infer problems (eg, a patient receiving metformin who has had multiple abnormal HbA1c tests is likely to have diabetes). In contrast, NLP strategies use algorithms designed to process and code free-text entries such as progress notes. Several groups have used data mining techniques and clinical associations to



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jamia.bmj.com/site/about/unlocked.xhtml>

predict patient problems.<sup>12–14</sup> Others have reported success using NLP techniques to automate the problem list.<sup>15–17</sup> Prior efforts have generally been evaluated in a laboratory setting, and focused on a single or small number of problems.

In this study, we performed a cluster randomized, controlled trial of a clinical alerting system that used inference rules to detect and notify providers of undocumented problems, giving them the opportunity to correct these gaps and increase problem list completeness. Our goal was to assess whether or not this system would improve problem notation for a broad array of patient conditions.

## METHODS

### Design overview

In a prior study, we presented a novel method for developing and validating problem-inference rules,<sup>6</sup> as well as a knowledge base containing validated rules for 17 clinically important conditions (henceforth referred to as ‘study problems’). These rules were based on previous work using data-mining techniques to identify medication-problem associations and laboratory-problem associations.<sup>14</sup> The rules take into account problem list entries (free-text and coded), billing diagnosis codes, laboratory results, medications, and vital signs to identify likely gaps in the problem list. Rule development and validation is described in detail in our previous work.<sup>6</sup> To summarize, rule development occurred in six steps: (1) identification of problem associations with structured data; (2) selection of specific problems; (3) development of preliminary rules; (4) characterization of preliminary rules and alternatives; (5) selection of a final rule; and (6) validation of the final rule. Using these rules, the average sensitivity and positive predictive value (PPV) for the training set were 83.4% and 91.1%, respectively; for the validation set, average sensitivity and PPV were comparable at 83.9% and 91.7%, respectively. Importantly, the inference rules were more sensitive than the problem list itself and had a higher PPV than billing codes. The performance of the rules is fully described in our prior paper,<sup>6</sup> and the contents of the rules are presented in the online appendix. As we developed the rules, we prioritized PPV and specificity in order to minimize the occurrence of false positive alerts, which might annoy users; however, for most conditions, we were able to achieve good performance on all four metrics: PPV, negative predictive value, specificity and sensitivity.

In four cases, we developed rules for groups of clinically similar entities: asthma/chronic obstructive pulmonary disease (COPD), congenital coagulopathy (hemophilia, congenital factor XI deficiency and von Willebrand disorder), osteoporosis/osteopenia, and renal failure/insufficiency. We created these groupings because, although we were able to determine, with a high degree of certainty, that the patient had one of the conditions (eg, asthma or COPD), we could not reliably discriminate between the conditions because of similar diagnostic criteria or treatment approaches. The 17 rules developed were:

- ▶ Attention deficit hyperactivity disorder
- ▶ Asthma/COPD
- ▶ Breast cancer
- ▶ Coronary artery disease (CAD)
- ▶ Congenital coagulopathy (hemophilia, congenital factor XI deficiency and von Willebrand disorder)
- ▶ CHF
- ▶ Diabetes mellitus
- ▶ Glaucoma
- ▶ Hypertension
- ▶ Hyperthyroidism

- ▶ Hypothyroidism
- ▶ Myasthenia gravis
- ▶ Osteoporosis/osteopenia
- ▶ Rheumatoid arthritis
- ▶ Renal failure/insufficiency
- ▶ Sickle cell disease
- ▶ Stroke

For each condition, problem synonyms were identified (eg, diabetes mellitus, type 2 diabetes, non-insulin-dependent diabetes mellitus). The alert would only fire if neither the problem itself nor any synonyms were present on the patient’s problem list. However, hierarchically related problems did not cause suppression of the alert (eg, hyperglycemia on the problem list did not prevent the diabetes mellitus alert from displaying, nor did nephropathy prevent renal insufficiency from being suggested). The complete set of rules for the study problems is described in detail in the online appendix, which uses standard codes (LOINC, SNOMED and ICD-9) to maximize its usefulness to sites wishing to replicate our study. The longitudinal medical record (LMR), a proprietary, full-featured, outpatient EHR<sup>18</sup> uses proprietary codes for laboratory results and problems, and our internal rules used these codes. However, the proprietary code systems are directly mapped to LOINC and SNOMED, respectively, and we used these pre-existing mappings to create the online appendix, so the description of the rules in the appendix matches the internal logic of our system exactly.

### Setting and participants

Participating clinics (n=11) included all primary care practices affiliated with Brigham and Women’s Hospital, an academic medical center in Boston, Massachusetts. Each practice used the LMR, which allows providers to record patient problems on an electronic problem list from a database of coded problems or as free-text entries. Participating clinics were divided into a total of 28 ‘clinical areas’ based on pre-existing administrative divisions within the clinics (eg, suites A, B, and C or pediatric vs adult medicine).

Participating practices included both urban and suburban clinics and a diverse mixture of primary care clinics in hospital and community settings across the greater-Boston area. These practices serve a racially and socioeconomically diverse population of patients.

### Randomization and interventions

We developed an electronic alert in the LMR which notifies providers when there appears to be an undocumented problem. At the time, a provider saves a typed note or reviews a dictation, and our system analyzes the patient’s medications, laboratory results, billing codes, and vital signs and uses the knowledge base to determine whether a patient is likely to have any of the 17 study problems. If the system detects one or more potential problems, it reviews the problem list to determine whether the problem is documented, and, if not, an actionable alert is shown onscreen (figure 1). If more than one undocumented problem is detected, alerts for all undocumented problems are displayed in a single window. To the right of each suggested problem is a reason why the alert is appearing. To the left is a check-box, which providers can use to select problems to add. Problems are ‘pre-checked’ for ease-of-use. Providers can accept the alert (in which case the problem will be added to the problem list), ignore the alert (in which case it will be presented the next time a note is completed for that patient), or over-ride the alert (in which case the alert is

**Figure 1** Screenshot of problem inference alerts.

**Problem List Suggestion**

Based on patient's clinical and billing data, the patient may have the following problems. Upon save, checked items will be added to the problem list. Unchecked items will not be added, and you will not be prompted again. Expand All

Add	Problem Description	
<input checked="" type="checkbox"/>	Coronary arteriosclerosis: Patient is taking a platelet aggregation inhibitor and has been billed at least once for CAD.	Enter Customizable Description.
<input checked="" type="checkbox"/>	Diabetes mellitus: Patient has a HbA1c >= 7.0%.	Enter Customizable Description. <span style="float: right;">Related terms</span>
<input checked="" type="checkbox"/>	Hypertensive disorder: Patient has been billed for hypertension and is on an antihypertensive agent.	Enter Customizable Description. <span style="float: right;">Related terms</span>
<input checked="" type="checkbox"/>	Hypothyroidism: Patient is on thyroid hormone.	Enter Customizable Description.
<input checked="" type="checkbox"/>	Osteoporosis: Patient has been billed at least twice for osteoporosis or osteopenia.	Enter Customizable Description.
<input type="checkbox"/>	or Osteopenia: Patient has been billed at least twice for osteoporosis or osteopenia.	Enter Customizable Description.
<input checked="" type="checkbox"/>	Chronic renal impairment: Patient has at least three low GFRs, and their most recent GFR is also low.	Enter Customizable Description. <span style="float: right;">Related terms</span>
<input type="checkbox"/>	or Chronic renal failure syndrome: Patient has at least three low GFRs, and their most recent GFR is also low.	Enter Customizable Description. <span style="float: right;">Related terms</span>

Explanations
Save
Cancel

suppressed for the duration of the study). When the provider adds a problem, he or she is also given the opportunity to add additional details or select a related term (eg, 'gestational diabetes' or 'diabetes mellitus type 2' instead of simply 'diabetes mellitus').

We conducted a randomized controlled trial of this intervention for a 6-month period, and also collected baseline data before the intervention in order to provide a second control. To reduce the risk of contamination, we used a cluster randomization method.

Clusters (n=28) were designated on the basis of pre-existing administrative divisions within the clinics. For example, one primary care clinic is divided into adult medicine, family medicine, and pediatric medicine, and another is divided into separate suites, A, B, and C. In both cases, these subunits were treated as separate clusters. Clusters were then grouped into three bands: hospital-based, community and federally qualified health center. Once grouped into the three bands, the clusters within each band were randomly allocated to the control or intervention arms, with 14 clinics randomized to the control arm and 14 to the intervention arm.

Providers were not aware to which arm their subclinic group was assigned until the intervention was implemented. Patients were not made aware of the intervention. No pre-intervention orientation or training took place in the intervention arm. Blinding was not possible given the nature of this intervention. Data were collected over a 6-month pre-intervention period and a subsequent 6-month intervention period. The system went live on May 16, 2010 in the intervention group clinics, and post-period data were collected prospectively for 183 days (6 months) for both arms, concluding on November 14, 2010. In addition, 183 days (6 months) of pre-period data from both arms were collected retrospectively to act as a baseline. The study was approved by the Partners HealthCare Human Research Committee and was registered with ClinicalTrials.gov (NCT01105923).

### Outcomes and follow-up

The primary outcome of this study was the acceptance rate of the alert, defined as number of alerts accepted divided by number of unique alerts presented. In certain instances, providers might see the same alert serially, so we aggregated

presentations and acceptances of the same alert for the same patients in our calculation of the acceptance rate.

As a secondary outcome, we measured the number of study problems documented in the two groups during the two time periods, and calculated the unadjusted relative rate of problem notation in the intervention group by comparing the number of problems recorded in the intervention arm during the intervention period to all other groups. The unadjusted relative rate was defined as the ratio ( $\text{problems}_{\text{intervention-post}} / \text{problems}_{\text{control-post}} / (\text{problems}_{\text{intervention-pre}} / \text{problems}_{\text{control-pre}})$ ).

### Statistical analysis

For the primary outcome, we calculated the acceptance rate of the alert for each of the 17 conditions, as well as an overall acceptance rate. For the secondary outcome of problem addition, which consisted of comparisons of count data, we modeled our data as Poisson-distributed counts. The unadjusted relative rate was calculated as described above, and tested for equality with one using a normal approximation.

In addition to this unadjusted relative rate, we used Poisson regression with an interrupted time series approach to control for potential exogenous temporal effects. Specifically, we used five coefficients and a scale parameter to model six features: starting rate, four slopes (pre and post period for the control and intervention arms), and a parameter for effect of the intervention. The effect parameter was an OR for the immediate effect of the intervention. In the case where differences between the control and intervention groups were non-significant, we removed the related terms from our model. This resulted in a new parameter for the intervention, which instead measured the overall effect of the intervention. This parameter has a similar interpretation to our unadjusted relative rate, and was compared for equality with one using a  $\chi^2$  test.

Finally, in order to counteract a possible problem of multiple comparisons, we used a Bonferroni correction. This correction maintains the error rate by testing each hypothesis against a lower  $\alpha$  value, where the new cut-off for statistical significance is  $\alpha/n$ , where  $n$  is the number of independent tests. In our case, the new cut-off was calculated to be 0.0029 (0.05/17 rules).

Demographic data were analyzed using a  $\chi^2$  test for categorical data and Student t test for continuous variables. Study data were analyzed using SAS V.9.2.

**Role of the funding source**

This work was supported by a grant from the Partners Community HealthCare Incorporated System Improvement Grant Program. Partners Community HealthCare Incorporated was not involved in the design, execution or analysis of the study, or in the preparation of the manuscript.

**RESULTS**

**Participant flow**

All 28 participating clinics completed the study, and there was no loss to follow-up. Overall, 41 039 patients were seen in the control clinics during the entire study period, and 38 025 patients were seen in the intervention clinics. A small number of patients (n=3894, 5.2%) were seen in both intervention and control clinics, and thus appear in both arms of the study. Figure 2 shows the flow of subclinics through the study.

**Demographic and baseline data**

Intervention and control groups appeared clinically similar across a range of demographic and clinical variables (table 1). During the 6-month pre-intervention period, greater problem list use was observed in the control group. A total of 3230 study problems (17.8 problems/day) were added in the intervention group, and 3597 study problems (19.8 problems/day) were added in the control group (p<0.001).

**Primary outcome: acceptance rate**

Problem inference rules fired a total of 17 043 times during the intervention period for a total of 11 508 patients in the intervention arm. The overall acceptance rate for problem inference

**Table 1** Demographics of patients seen in control and intervention clinics

Demographic characteristic	Control	Intervention	p Value
Age (years), mean (SD)	49.6 (20.0)	47.7 (19.6)	<0.0001
Sex (female)	61.4%	68.0%	<0.0001
Race or ethnicity			
American Indian	0.12%	0.17%	<0.0001
Asian/Pacific Islander/Native Hawaiian	2.4%	3.0%	
Black	14.2%	11.4%	
Hispanic	17.1%	17.6%	
White	57.6%	58.6%	
Other	0.61%	0.55%	
Declined/unknown	7.9%	8.7%	
Language			
English	85.7%	85.3%	<0.0001
Spanish	8.8%	10.0%	
Other	5.5%	4.6%	
Primary insurance			
Commercial	59.6%	64.1%	<0.0001
Medicare	22.3%	17.7%	
Medicaid	14.0%	14.6%	
Other/self pay	4.0%	3.6%	
Income (US\$), mean (SD)	56 350 (20 686)	59 663 (23 737)	<0.0001

alerts was 41.1%. The highest acceptance rate of the 17 conditions was 55.7% for glaucoma alerts (table 2). Alerts for myasthenia gravis and sickle cell disease were infrequently presented and infrequently accepted.

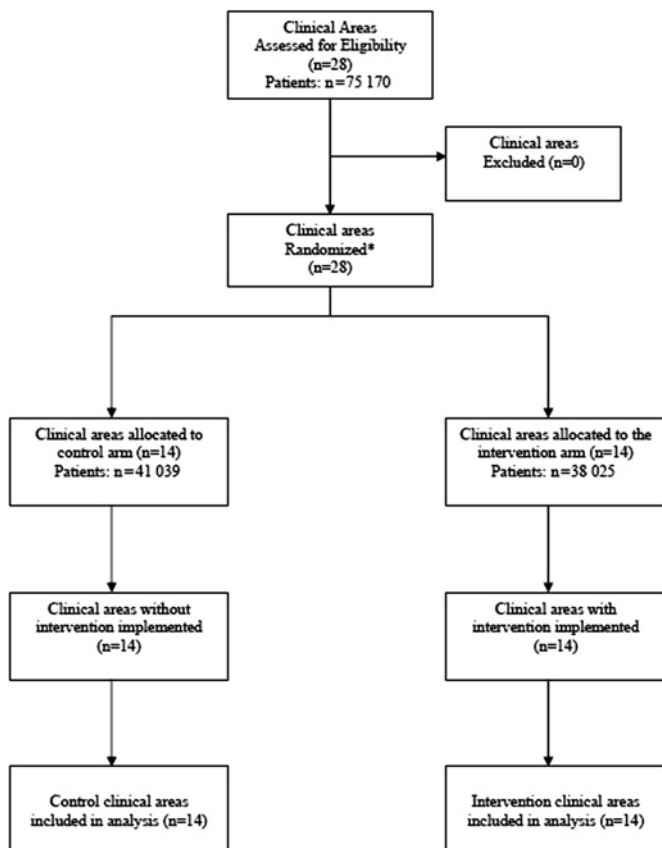
**Pre-specified secondary outcome: problem notation in the problem list**

During the intervention period, 10 016 study problems were added in the intervention group compared with 3739 added in the control group—an absolute difference of 6277 problems (compared with 367 fewer problems added in the intervention group during the pre-intervention period, p<0.0001). The unadjusted relative rate of study problem addition was 2.98 times more problem notation in the intervention group (p<0.0001), and the adjusted OR was 3.43 (p<0.0001).

The cumulative number of study problems added over the course of the entire study is shown in figure 3. As reflected in the figure, the rate of study problem notation during the pre-intervention period was slightly lower in the intervention group than in the control group. The inflection point in the intervention group line was coincident with the initiation of the study intervention in that group and, by the completion of the study, the intervention group had added significantly more problems than the control group.

Table 3 shows the rate of problem list addition for each of the 17 study problems. Using the unadjusted differences measure, statistically significant increases in problem notation were seen for 15 of 17 study problems using an uncorrected threshold of p<0.05. When the Bonferroni correction was applied to the threshold, two of the 15 problems were no longer statistically significant (congenital coagulopathy and hyperthyroidism). Relative rates of problem notation (for statistically significant conditions) ranged from 1.54 times more notation for hyperthyroidism (p=0.031) to 6.89 times more notation for renal failure and insufficiency (p<0.0001).

In addition to the unadjusted difference, we used Poisson regression and interrupted time series analysis to control for temporal trends. We began with a model with four slopes (results not shown). Outcomes from this model were similar to



**Figure 2** Participant flow for study clusters (subclinic randomization). Because randomization was carried out at the subclinical level, a small number of patients (n=3849) appear in both arms of the study.

**Table 2** Alert acceptance rates by condition

Disease	Unique rule firings	Number of alerts accepted	Number of alerts over-ridden	Number of alerts ignored	Overall acceptance rate (%)*
Attention deficit hyperactivity disorder	225	102	31	194	45.3
Asthma/chronic obstructive pulmonary disease	2452	842	585	1867	34.3
Breast cancer	245	116	36	209	47.3
Coronary artery disease	1069	439	274	795	41.1
Congenital coagulopathy	45	15	19	26	33.3
Congestive heart failure	914	331	279	635	36.2
Diabetes mellitus	1330	519	324	1006	39.0
Glaucoma	336	187	40	296	55.7
Hypertension	5362	2281	1029	4333	42.5
Hyperthyroidism	141	46	39	102	32.6
Hypothyroidism	1291	639	220	1071	49.5
Myasthenia gravis	15	3	6	9	20.0
Osteoporosis/osteopenia	2285	962	475	1810	42.1
Rheumatoid arthritis	231	61	78	153	26.4
Renal failure/insufficiency	991	413	148	843	41.7
Sickle cell disease	12	1	7	5	8.3
Stroke	99	54	14	85	54.5
Total	17 043	7011	3604	13 439	41.1

\*Overall acceptance rate combines alerts that were accepted after being displayed multiple times (number of alerts accepted/unique rule firings).

unadjusted results; however, the increases for congenital coagulopathy and hyperthyroidism were no longer statistically significant (at either  $p < 0.05$  or the Bonferroni-corrected threshold of  $p < 0.0029$ ). After removal of non-significant model components (yielding a simplified two-slope model shown on the right-hand side of table 3), the difference for congenital coagulopathy was once again statistically significant; however, when the Bonferroni correction was used, this study problem was not statistically significant. ORs from the final model were mostly similar to the unadjusted relative rates, and the overall OR for intervention effect on problem list notation was 3.43 ( $p < 0.0001$ ).

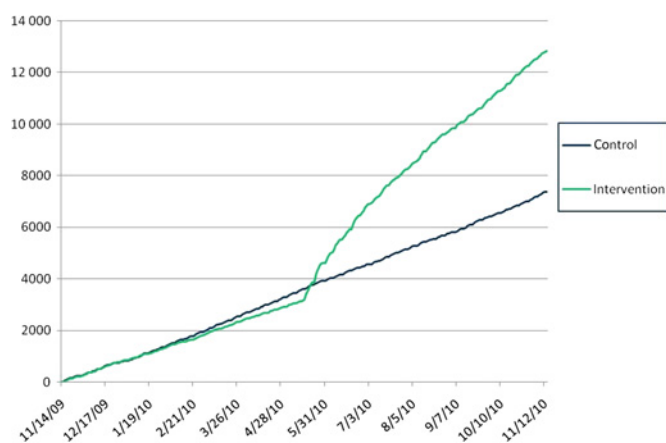
To assess the accuracy of problems added as the result of the intervention, we also conducted an audit of a random selection of accepted alerts ( $n=1178$ ). In order to form a representative sample, we used a weighting strategy. Each of the 17 may have been suggested on the basis of one or more condition sets (eg, a diabetes suggestion could be triggered by the HbA1c value, medications, billing codes, or a combination of these features). For each condition set, we reviewed the accuracy of up to 30

alerts (less if there were fewer than 30 total accepted alerts for a given condition set). Study staff (FM) conducted a manual chart review, including a review of the notes. The gold standard was free-text documentation of the problem in any stored physician notes. We computed a weighted accuracy score by taking the accuracy for each condition set and weighting it according to how often that condition set triggered an alert. The weighted accuracy of all accepted alerts was found to be 89.8%. The 10.2% of alert acceptances not associated with documentation had a variety of causes, including patients near to diagnosis of a disease (eg, patients with pre-diabetes or metabolic syndrome on the cusp of diagnosis with diabetes), patients who appeared to actually meet diagnostic criteria for the disease, but for whom the diagnosis was not discussed in the record (eg, patients who met diagnostic criteria for chronic kidney disease based on glomerular filtration rate or hypertension based on serial blood pressure measures but without documentation of the condition in their notes), as well as some potentially erroneous additions.

## DISCUSSION

We found that electronic problem list alerts were often accepted by users, and resulted in a substantial increase in study problem notation. The rate of notation of study problems increased dramatically during the intervention period as a result of this simple alert-based intervention. Overall, study problems were approximately three times more likely to be documented when alerts were shown. This increase is clinically important, since many of these problems are used for quality improvement and CDS.

Importantly, 14 out of 17 study problems were more often recorded in the intervention group than the control group. Only three conditions, myasthenia gravis, sickle cell disease, and hyperthyroidism, had similar rates between the two groups; however, even though the difference for hyperthyroidism was not statistically significant with Bonferroni correction, one could infer that there may be a trend for possible statistical significance with a larger sample size. Since our previous research validated the algorithm for the study problems,<sup>6</sup> it is probable



**Figure 3** Cumulative number of study problems added during pre-intervention and intervention periods.

**Table 3** Total study problems recorded during pre-intervention and intervention periods with unadjusted and adjusted ORs

Problem	Pre-intervention period		Intervention period		Unadjusted* comparison		Adjusted† comparison	
	Control	Intervention	Control	Intervention	Relative rate	p Value	OR	p Value
Attention deficit hyperactivity disorder	66	72	69	157	2.09	<0.0001	2.23	<0.0001
Asthma/chronic obstructive pulmonary disease	498	503	529	1291	2.42	<0.0001	2.98	<0.0001
Breast cancer	151	123	180	246	1.68	0.0004	1.78	<0.0001
Coronary artery disease	164	134	178	576	3.96	<0.0001	4.66	<0.0001
Congenital coagulopathy	4	4	5	19	3.80	0.0133	2.06	0.0384
Congestive heart failure	64	50	97	373	4.92	<0.0001	7.56	<0.0001
Diabetes mellitus	597	446	535	814	2.04	<0.0001	1.97	<0.0001
Glaucoma	53	74	61	263	3.09	<0.0001	3.78	<0.0001
Hypertension	1019	863	1031	3082	3.53	<0.0001	4.12	<0.0001
Hyperthyroidism	86	96	72	124	1.54	0.0308	1.30	0.2928
Hypothyroidism	207	237	205	823	3.51	<0.0001	3.99	<0.0001
Myasthenia gravis	4	2	3	5	3.33	0.2850	2.10	0.114
Osteoporosis/osteopenia	513	483	582	1521	2.78	<0.0001	3.40	<0.0001
Rheumatoid arthritis	27	21	24	75	4.02	<0.0001	3.97	<0.0001
Renal failure/insufficiency	84	73	87	521	6.89	<0.0001	8.22	<0.0001
Sickle cell disease	9	12	13	23	1.33	0.3538	1.66	0.2897
Stroke	51	37	68	103	2.09	0.0023	2.35	0.0002
Total	3597	3230	3739	10 016	2.98	<0.0001	3.43	<0.0001

\*Unadjusted comparison based on unadjusted relative rate of problem list addition, as described in the methods sections.

†Adjusted comparison based on Poisson regression model for an interrupted time series.

that the overall low prevalence of myasthenia gravis and sickle cell disease is responsible for the lack of any difference in notation between study arms.

Our results suggest that problem inference rules such as these are a valuable tool for improving problem list completeness and thus may be beneficial for improving patient care. A more complete problem list makes it easier for providers to obtain an accurate picture of a patient’s issues, which is especially important when an unfamiliar patient is being seen, such as in the case of urgent care or emergency visits, or in inpatient wards. Additionally, since problems are used for CDS, identification of patients for research studies, and quality measurement, these types of rules show great potential for improving quality and reducing costs.

One important question is how the observed increase in the notation of problems would ultimately benefit patients. Assuming that a given alert was correct, there were two potential scenarios for each alert reminder: (1) the alert called attention to an undocumented problem that the provider was not aware of and (2) the alert recommended a problem that the provider was aware of but had not documented in the problem list. Although the first scenario may provide a particular immediate clinical impact (making the provider aware of an unknown diagnosis), it is also likely to be less common. However, both cases provide significant positive clinical benefit, including enabling CDS (such as relevant preventive care reminders), facilitating quality measurement and research, and promoting awareness of a patient’s active problems among the entire care team (including providers that may not know the patient well).

An additional implication of this study may be to help providers achieve ‘meaningful use’ of EHRs, as one of the stage 1 and 2 meaningful use goals is to demonstrate problem list use for 80% of patients over the next few years.<sup>9 10</sup> By meeting the meaningful use criteria, clinicians would receive incentive funds that could offset the expenses of implementing and maintaining the LMR. A tool such as that described here may be highly valuable for encouraging problem list use and increasing

accuracy in the near term, especially for the large numbers of providers who are just starting with electronic records and are struggling to populate their problem lists.

Given these promising results and diverse potential applications, we hope to dramatically expand the problem inference knowledge base in the future. Ultimately, rules such as these may be used in tandem with provider documentation to increase the accuracy of the problem list, with the potential to improve patient care. However, additional provider engagement will also be required, and some problem list maintenance tasks (such as the removal of resolved problems or consolidation of duplicates) are beyond the scope of our described intervention.

### Limitations

Our investigation has several potential limitations. First, problem inference rules were developed, validated, and tested at a single site. Further research will need to be carried out to assess the generalizability of these results. Additionally, we had the benefit of a self-developed EHR, giving us the ability to extract the necessary data to develop and validate our knowledge base, as well as the ability to design a novel intervention. In contrast, most institutions use commercial EHR systems, which may not have this degree of flexibility. Although we encourage other institutions to develop and validate their own rules when feasible, we have also made our full knowledge base freely available for use by other organizations, including vendors.<sup>6</sup> Another possible limitation of this approach is that imperfect accuracy of the problem inference rules could lead to erroneous alerts, which, if accepted, would result in inaccurate problems being added to the problem list. An audit of a random selection of accepted alerts revealed a global weighted accuracy of 89.8%. Although the accuracy of accepted alerts was very high overall, this finding nevertheless reveals the presence of a number of problems erroneously added to the problem list as a result of the alerts. Many of these instances appeared to be the result of borderline conditions (eg, metabolic syndrome, white-coat hypertension). However, the potential problem of providers accepting erroneous clinical alerts merits further study.

Additionally, the overall acceptance rate for the alerts was 41.1%, with the rest of the alerts being either over-ridden or ignored by physicians. Since the accuracy of the accepted alerts was shown to be high in the above-described audit, future research should look into the reasons why alerts are either over-ridden or ignored. Finally, the intervention was limited to primary care providers; extending this tool to specialties may require more focus on the types of alerts presented.

## CONCLUSION

Problem inference alerts appear to be a powerful tool for improving notation of patient problems, and may thus in turn help improve quality of care. The use of problem inference alerts dramatically increased the notation of patient problems in the intervention group. Healthcare providers seeking to increase problem list completeness for meaningful use or other reasons should consider implementing such alerts. Future studies should focus on whether implementing such alerts has a direct effect on patient outcomes.

**Acknowledgments** We thank the following: Don Bugbee, Betti Rozenman, Regina Breyt, Janak Joshi and Bill Hanley for designing, developing and implementing the MAPLE software; Teal Aroy, Sue Smith, Janet Cygielnik, Dave Dubois, Kira Tsivkin and Roberto Rocha for assistance in knowledge base creation; Howard Goldberg, Mike Vashevko and Molly Schaeffer for providing the problem terminology service; Julie Fiskio, Deb Williams, Yelena Kruglova, Brian Hingston and Alex Turchin for extracting data; E John Orav for statistical consulting; Blackford Middleton for assisting with development and implementation management; and Stanislav Henkin for help with editing of the manuscript.

**Funding** Partners Community HealthCare Incorporated (PCHI) System Improvement Grant Program.

**Competing interests** None.

**Ethics approval** Ethics approval was provided by Partners HealthCare Human Research Committee.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** AW had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## REFERENCES

1. **Hartung DM**, Hunt J, Siemenczuk J, *et al*. Clinical implications of an accurate problem list on heart failure treatment. *J Gen Intern Med* 2005;**20**:143–7.
2. **Wright A**, Goldberg H, Hongsermeier T, *et al*. A description and functional taxonomy of rule-based decision support content at a large integrated delivery network. *J Am Med Inform Assoc* 2007;**14**:489–96.
3. **Kaplan DM**. Clear writing, clear thinking and the disappearing art of the problem list. *J Hosp Med* 2007;**2**:199–202.
4. **Szeto HC**, Coleman RK, Gholami P, *et al*. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. *Am J Manag Care* 2002;**8**:37–43.
5. **Tang PC**, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. *J Am Med Inform Assoc* 1999;**6**:245–51.
6. **Wright A**, Pang J, Febowitz JC, *et al*. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. *J Am Med Inform Assoc* 2011;**18**:859–67.
7. **Febowitz J**, Wright A. *The Patient Problem List: an Ethnographic Study of Primary Care Provider Use and Attitudes*. Washington, DC: AMIA 2011 Annual Symposium, 2011. under review.
8. **Wright A**, Maloney FL, Febowitz JC. Clinician attitudes toward and use of electronic problem lists: a Thematic analysis. *BMC Med Inform Decis Mak* 2011;**11**:36.
9. *Comparison of Meaningful Use Objectives Between the Proposed Rule to the Final Rule*. 2010. [https://www.cms.gov/EHRIncentivePrograms/Downloads/NPRM\\_vs\\_FR\\_Table\\_Comparison\\_Final.pdf](https://www.cms.gov/EHRIncentivePrograms/Downloads/NPRM_vs_FR_Table_Comparison_Final.pdf) (accessed 13 Jun 2010).
10. **Blumenthal D**, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med* 2010;**363**:501–4.
11. **Jha AK**. Meaningful use of electronic health records: the road ahead. *JAMA* 2010;**304**:1709–10.
12. **Burton MM**, Simonaitis L, Schadow G. Medication and indication linkage: a practical therapy for the problem list? *AMIA Annu Symp Proc* 2008:86–90.
13. **Cao H**, Markatou M, Melton GB, *et al*. Mining a clinical data warehouse to discover disease-finding associations using co-occurrence statistics. *AMIA Annu Symp Proc* 2005:106–10.
14. **Wright A**, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. *J Biomed Inform* 2010;**43**:891–901.
15. **Chase HS**, Radhakrishnan J, Shirazian S, *et al*. Under-documentation of chronic kidney disease in the electronic health record in outpatients. *J Am Med Inform Assoc* 2010;**17**:588–94.
16. **Meystre S**, Haug P. Improving the sensitivity of the problem list in an intensive care unit by using natural language processing. *AMIA Annu Symp Proc* 2006:554–8.
17. **Meystre SM**, Haug PJ. Randomized controlled trial of an automated problem list with improved sensitivity. *Int J Med Inform* 2008;**77**:602–12.
18. **Poon EG**, Wald J, Bates DW, *et al*. Supporting patient care beyond the clinical encounter: three informatics innovations from partners health care. *AMIA Annu Symp Proc* 2003:1072.