Leveraging the electronic problem list for public health research and quality improvement

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Courtney Lee Hebert, MD

Graduate Program in Public Health

The Ohio State University

2013

Master’s Examination Committee:

Philip R.O. Payne, PhD, Advisor

Peter J. Embi, MD

Randi Foraker, PhD
Copyright by
Courtney Lee Hebert, MD
2013
Abstract

**Background:** With the growth in electronic health records (EHR) throughout the country we have an unprecedented opportunity to leverage the data that are provided by patients and their providers every day in order to improve healthcare on a large scale. Health outcomes research, disease surveillance programs and predictive modeling can all make use of data from comprehensive EHRs. For in-hospital risk-prediction, up-to-date data from an EHR can provide the ability to perform real-time risk prediction. However, there are many challenges to using EHR data in real time. One of these is the challenge of identifying up-to-date comorbidities in an automated, reliable way.

**Objective:** This study is an in-depth examination of one important aspect of the EHR, the electronic problem list, and an assessment of its potential use in real-time risk-modeling.

**Setting:** Large Midwestern tertiary care facility

**Participants:** All patients admitted with a discharge diagnosis of pneumonia from May of 2012 through January 2013.

**Methods:** The study has three parts. The first is to determine how comorbidities derived from the problem list compare to comorbidities from administrative billing codes, the usual method for automatically detecting comorbidities. The second is to
determine how the problem list changes during a hospitalization. The third is to
determine how the risk of readmission (calculated by a previously developed
prediction tool) changes per day of hospitalization and at what point in a
hospitalization adequate risk prediction can be made.

Results: 96.1% of pneumonia patients had at least one problem on their problem
list during their encounter. 72.4% of patients with a problem list had pneumonia
listed as a problem during their encounter. Most patients had problems added to
their list during the hospitalization. For the majority of major comorbidities kappa
statistics were very good comparing the problem list to administrative billing.
Finally, the performance of the risk-model did not vary significantly at different
points during the hospitalization.

Conclusions: This study suggests that the problem list is a reliable source of
comorbidity data for major comorbidities at our institution and that it is reasonable
to use the list for risk-modeling throughout the hospitalization.
Vita

1997.................................................. Bishop Watterson, High School

2001.................................................. B.A. Theology, Georgetown University

2005.................................................. M.D, The Ohio State University

2008.................................................. Residency, Internal Medicine, University of Chicago

2011.................................................. Fellowship, Infectious Diseases, University of Chicago

2011.................................................. Post-doctoral researcher, the department of Biomedical Informatics, The Ohio State University

2013.................................................. Instructor, the department of Biomedical Informatics, The Ohio State University

Publications


**Fields of Study**

Major Field: Public Health

Specialization in Biomedical Informatics
Table of Contents

Abstract ................................................................................................................................. ii

Vita ........................................................................................................................................ iv

Publications .......................................................................................................................... iv

Fields of Study ...................................................................................................................... v

List of Tables ........................................................................................................................ viii

List of Figures ...................................................................................................................... ix

Chapter 1: Introduction ........................................................................................................ 1

Chapter 2: Background ......................................................................................................... 3

Data Sources for Comorbidity Detection ............................................................................ 3

Previous research on the EPL ............................................................................................. 4

A use case for real-time risk modeling ............................................................................... 6

Chapter 3: Methods .............................................................................................................. 9

Overview ............................................................................................................................... 9

Settings and Participants ..................................................................................................... 10

Study Design and definitions .............................................................................................. 10

The Problem List ................................................................................................................ 12
Derivation dataset for predictive model ................................................................. 14
Statistical Analysis ................................................................................................. 14
Comparison of administrative to EPL comorbidities ........................................ 14
Derivation of the predictive model ....................................................................... 15
Validation with new data ...................................................................................... 16
Chapter 4: Results ................................................................................................. 18
Validation dataset ................................................................................................. 18
Electronic Problem List ......................................................................................... 18
Kappa statistics ..................................................................................................... 21
Derivation dataset ................................................................................................. 23
Validation of model .............................................................................................. 24
Chapter 5: Discussion ........................................................................................... 27
Limitations ............................................................................................................. 29
Conclusion ............................................................................................................ 31
References ............................................................................................................ 33
List of Tables

Table 1: Advantages and Disadvantages of available comorbidity sources ................... 3
Table 2: Assumptions made for this study ................................................................. 15
Table 3: Demographics of the cohort ................................................................. 18
Table 4: Evolution of the EPL over hospitalization and a comparison of pre- and post-billing EPL ................................................................. 20
Table 5: Kappa statistics and percentage of patients with comorbidities comparing EPL to administrative coding data ................................................................. 22
Table 6: Final derivation model ............................................................................ 24
List of Figures

Figure 1: Overview of Methodology .................................................................................. 17
Figure 2: Percentage of patients without an EPL by day of hospitalization................. 19
Figure 3: Scatter plot comparing Charlson scores from EPL on the day of discharge to administrative data ........................................................................................................... 23
Figure 4: Change in AUC of model at different points during the hospitalization........ 25
Chapter 1: Introduction

The increase in the availability of electronically collected healthcare data has created new opportunities for personalized and predictive care in the nation's medical centers. One of the data sources to which many institutions now have access is a list of a patient's problems that resides in the EHR. This is called the electronic problem list (EPL). The EPL is unique in that it has the potential to provide up-to-date comorbidity data which can be used in a variety of applications. For example, comorbidity data can be used to create risk models that predict the occurrence of a clinical outcome such as death, readmission to the hospital, or future infection (1-4). Many of these risk-models are currently calculated by hand by a care provider, however improvements in technology and the availability of the EPL have made it possible to automate the models. This is especially important for models that target high-risk patients for in-hospital interventions which require risk-scoring to be available in real-time, at the point-of-care.

This study aims to address and study some of the common challenges when using the EPL for comorbidity detection. This is an important area of study in the field of biomedical informatics, given that comorbidities are included in many risk scores, data on comorbidities can help in automated cohort detection for research
studies, and accurate comorbidity identification can aide in decision support to improve individual patient care (2, 5, 6).

The three main goals of this study are 1. To determine how the EPL is used, and how its composition evolves over the course of a hospitalization; 2. To compare comorbidities documented in the EPL to those documented by administrative coding; and 3. To determine whether the data on comorbidities from the EPL can be used for real-time risk prediction in an actual use case of readmission prediction at our institution.
Chapter 2: Background

Data Sources for Comorbidity Detection

Comorbidities are recorded in clinical notes, administrative coded data and the EPL. In addition, comorbidities can be inferred from laboratory abnormalities (for example a hemoglobin AIC above 6.5% may indicate that the patient has diabetes). There are advantages and disadvantages to each of these methods as can be seen in Table 1.

Table 1: Advantages and Disadvantages of available comorbidity sources

<table>
<thead>
<tr>
<th></th>
<th>Coded</th>
<th>Available in real-time</th>
<th>Clinician-documentd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free text notes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Administrative billing data</td>
<td>Yes</td>
<td>No</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Laboratory data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Electronic problem List</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Clinical notes are written by a clinician and are likely the richest source of patient data. Unfortunately, to automate these notes requires natural language processing methods, which have not yet been fully adopted for mainstream use (7). Administrative billing codes are available at most institutions, can be easily automated, but up-to-date administrative codes are usually not available in real-time. Instead, most of this coding is done after a hospitalization is complete.
Laboratory measurements are easy to work with and coded to allow for automation, however they can be misleading when it comes to identifying comorbidities. These measurements are prone to laboratory error, may be only temporarily abnormal and often require clinical context to understand their relevance.

The EPL, on the other hand, is clinician curated, available at the point of care, and coded for automation, suggesting that it could be an ideal source of comorbidity data for real-time risk prediction. Unfortunately there is very little data on the reliability of the EPL, and many studies report major limitations when using this source. In the following section I will summarize these studies as well as show where more study is needed.

*Previous research on the EPL*

Much of the research on the EPL focuses on comparing it to other, better-studied data sources. Szetzo and colleagues compared the EPL from the Veteran’s administration (VA) EHR system to administrative billing codes and manual chart review (8). They found that the sensitivity of administrative data was 80% for 8 out of 9 comorbidities versus 49% for the EPL, and specificity was 91-100% for administrative data and 98-100% for the EPL. This was a very small study, with only 148 patients included.

Wright and colleagues studied the EPL at their institution in an attempt to create automated methods for detecting comorbidities (9). They looked at 17 individual comorbid conditions and found that the sensitivity of the EPL ranged from 4.7% for renal failure to 78.5% for breast cancer. On average the sensitivity
was 51.6%. The specificity of each of the problems was reported at 100% however the authors made the assumption that if a problem was on the EPL then it was considered a true positive, not warranting chart review. The sensitivity for administrative data ranged from 43.3% for renal failure to 100% for hemophilia with specificities from 95-100%.

Galanter and colleagues compared the EPL to manual chart review for inpatient records and found that only 25% of the problems found by chart review were on the EPL and 46% of patients had no problems documented on their EPL at all (10). Similarly, Meystre and colleagues studied 247 patients at their institution soon after the EPL was implemented and found very low sensitivity when compared to manual chart review (11).

In general the limited number of studies available suggests that the EPL generally has low sensitivity but high specificity in detecting comorbidities and that the list is underused. There are several reasons why this might be the case. Many studies have shown that there are a lack of clear standards on how to manage an EPL and that practices and attitudes differ from clinician to clinician (12, 13). In addition, because the EPL is generally for clinical use, a clinician may not have an incentive to document a problem that is present but not currently active, or is not a problem that they are currently treating. For example, if the patient mainly sees a cardiologist at one institution, the cardiologist may not feel comfortable documenting the details of the patient’s seizure disorder, so may leave it off the list.
Although much of the literature has cast doubt on the reliability of the EPL, there are several reasons why the EPL deserves more study. The literature is limited and in most of the studies the sample size was small. Some of the studies (11) evaluated the performance of the EPL soon after it was implemented, not allowing time for clinicians to fully incorporate it into practice. Many of these studies are older and with the rapid changes in EHR technology, they may no longer be relevant. Finally, these studies have limited generalizability, often studying home-grown or unique systems (for example the VA system). This study is unique in that it includes a relatively large number of recently admitted patients at a large tertiary care center using an EHR system that has 678 total installations and 13.7% of all installations in acute-care hospitals (14). In addition, the EHR system had been in place for at least 7 months prior to the start of the study.

The next section explains the motivation for the current study by presenting an actual use-case at our institution for which real-time comorbidity detection was necessary.

*A use case for real-time risk modeling*

The use case that I consider in this paper is a case of real-time 30-day readmission risk prediction at the Ohio State University Wexner Medical Center (OSUWMC). Readmission to the hospital within 30 days from a previous inpatient admission is considered by the Centers for Medicare and Medicaid services to be a marker of poor quality care (15). Medicare fines hospitals that have higher rates of readmission for key diagnoses (congestive heart failure, pneumonia and acute
myocardial infarction) and more diagnoses are being added in the future (16). This creates a huge incentive for hospitals to try to reduce readmissions. One way to do this is to find those patients that are at highest risk for readmission and target them for readmission-reduction strategies. In this environment researchers within the department of Biomedical Informatics were charged with creating a readmission-risk prediction tool to use on recently discharged patients.

The tool was created with data collected with a previous EHR platform. This platform did not have an EPL, so we identified comorbidities using the administrative discharge codes for the initial hospitalization. Because our comorbidity variables were not available until the day of discharge or later, our point of reference for the model was the end of the hospitalization. So the variables were created using the entirety of the data available during the hospitalization. For example, the variable $\text{hemoglobin} < 10 = 1$ if a patient had ever had a hemoglobin measurement that was less than 10g/dL during the whole hospitalization, 0 otherwise.

Three models were created: one for congestive heart failure, one for acute myocardial infarction and one for pneumonia. They were validated on a held out random sample and generally performed well. However, when we went to implement these models we encountered two major problems:

1. The end-users of the models (nurses, case-managers, patient navigators) wanted to use the models early in the hospital course so that they would have time to intervene if necessary.
2. We had recently switched to a new EHR platform which had an EPL that
could potentially be used for comorbidity identification, but there had been
little study of its reliability or usage at our hospital.

This study aims to respond to these two problems. I will test the reliability of
the EPL when compared to administrative data in order to determine if the EPL
comorbidities can be reasonably substituted for the administrative comorbidities.
Then, I will determine how the EPL evolves over the course of hospitalization, and
finally evaluate whether or not the model created using cumulative, end of
hospitalization data would be predictive of outcome earlier in the course of the stay.

I chose to look primarily at one diagnosis, pneumonia, for several reasons.
First, pneumonia patients can be cared for in multiple settings by many different
teams, so I would get a better sample of hospital practice than by using the
congestive heart failure or acute myocardial infarction patients most of whom
would have been cared for by a small number of cardiologists. Second, the
pneumonia model performed well on the previous validation cohort. Third, the
pneumonia model included variables that would be likely to vary over the course of
a hospitalization (*number of medications*) and included several comorbidity
variables.

I hypothesized that the EPL would approximate the reliability of
administrative billing sources of comorbidities, and that although there may be
fluctuations in risk scoring and EPL content, scores obtained early on in a hospital
course will be predictive of later scores.
Chapter 3: Methods

Overview

This was a retrospective cohort study. The main objectives of the study were to evaluate how the EPL is used during an inpatient visit at our institution and to determine the challenges and potential benefits to its use for real-time risk prediction of 30-day readmission during an inpatient hospital stay. In summary, I will accomplish this by:

1. **Comparing the comorbidities present on the EPL to administrative coding** in order to determine the reliability of the EPL (by the last day of hospitalization), and assess whether the EPL could be used in place of the traditional coded data source, administrative billing codes.

2. **Studying the evolution of the EPL over the course of hospitalization** in order to ensure that important comorbidities are added to the EPL in a timely manner and that the EPL is actually being updated during a hospitalization. If this is not true then it would not be a good source for real-time data.

3. **Comparing the performance of the risk-prediction model on different days of hospitalization** in order to determine whether the risk-score early in the course of hospitalization is predictive of the risk-score at discharge.
**Settings and Participants**

The Ohio State University Wexner Medical Center (OSUWMC) is a 976-bed Midwestern tertiary care center located in Columbus, OH. The medical center includes six hospitals that provide general medical services, comprehensive cancer care and transplant medicine. OSUWMC has an integrated inpatient and outpatient EHR (EpicCare, Epic Systems, Verona, WI). The current outpatient EHR has been in use since 2006 and the current inpatient EHR first went live in October of 2011. The data for this study were extracted from the OSUWMC information warehouse which includes comprehensive administrative and clinical data on all OSUWMC patients.

All adult (≥18) patients admitted to OSUWMC from 5/1/2012 to 1/29/2013 with a primary discharge diagnosis of pneumonia (International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes: 480.x, 481, 482.x, 483.x, 484.x, 485, 486, 487.0, 488.01, 48.11) were eligible for inclusion into the study. This cohort will be referred to as the *validation cohort*.

**Study Design and definitions**

All data collected were extracted from discrete data fields in the EHR; no free-text sources were used. Data were collected on demographics, social history, laboratory measurements, medication administration, administrative billing data and EPL data. Patients were excluded from analysis if they died during the course of the hospitalization (n=27) since this made them not available for readmission. Prisoners were excluded (n=25). Those who left against medical advice were excluded (n=11), as were patients who were discharged to another acute short-term
care facility (n=5). These exclusion criteria were chosen to reflect the exclusion criteria of the original study which created the risk-prediction model. Because this study focused on how the EPL was used at our hospital, patients without any entry on their EPL for the entirety of their hospitalization were excluded (n=17). In addition to these exclusions, two patients who were erroneously missing all medication administration data in our system were excluded.

An index encounter was defined as the first admission for each patient that fulfilled the criteria above. Readmission within 30 days was defined as an inpatient hospital stay 30 days or less after discharge from an index encounter. Any readmission regardless of cause was included. Only the first readmission within 30 days was counted as a readmission (excluded 30 subsequent admissions).

Medication data was collected per day of hospitalization. For example, the variable number of medications was calculated independently for each day of hospitalization and steroid use included just steroid use on that particular day. The medication data were from administration records, so only those medications that were explicitly “given” that day were included.

Laboratory data was collected per day of hospitalization as well. However, the variable hemoglobin less than 10 was defined as a hemoglobin measurement that was less than 10g/dL at any point, up to and including, the current date.

Administrative sourced comorbidities were extracted from discharge diagnosis ICD-9-CM codes associated with the index hospitalization.
The Problem List

In order to understand the complexities involved in using the EPL, it is important to understand its role at our institution.

The EPL is a feature of the EHR at OSUWMC that is managed by clinicians. Each patient has one EPL, which may or may not be populated with problems. The problems are added via a text search and then drop down menu, from which the clinician chooses. There are two dates associated with all problems that are entered. The first is a first entered date which is the date that a clinician enters the problem. There is also a noted date. This date is by default the same as the first entered date but can be manually altered by a clinician if they choose to instead document when a problem began. All problems are linked to an ICD-9-CM code. These ICD-9-CM codes are mapped to intelligent medical objects (IMO, Intelligent Medical Objects, Inc., Northbrook, IL) problem terminology. This mapping is intended to help physicians by providing terminology that they are familiar with (as opposed to the standard text associated with ICD-9-CM coding) and help them to better communicate a patient’s current problems to other health care providers. For example, two different IMO phrases “heart attack with ST Elevation on EKG” and “heart attack without ST Elevation on EKG” may be mapped to the same ICD-9-CM code 410.9 “Acute Myocardial Infarction of unspecified site” (17). These two problems have very different meanings clinically, despite being mapped to the same ICD-9-CM code.
Problems can be added during an outpatient or inpatient visit and these problems together create a patient’s EPL. The only way a problem can be removed from the EPL is by a clinician deleting a problem or resolving a problem. Currently, OSUWMC does not have a specific guiding policy on how the EPL should be maintained.

In 2012 the EPL began to be used for inpatient physician billing. This means that after a physician saw a patient they would have to bill for their services using the EHR and link their treatment to one or more problems on the EPL. If a relevant problem was not already available, they would have to add a new one. Inpatient physician billing was rolled-out over several months. By 8/24/12 all internal medicine and hospitalists were using the EPL for billing. In order to determine if administrative changes over this time period may have affected the quality of the EPL I compared the number of problems on the list and the number of patients that had pneumonia on their EPL from the time period before all physicians were using the EPL for billing to after 8/24/12.

A patient was considered to have a problem on their EPL on a particular day if the first entered date was less than or equal to the date of interest and the resolved date was greater than the date of interest or there was no resolved date.

Both administrative and EPL codes, were mapped to a set of 33 comorbidities of interest adapted from ICD-9-CM groupings previously published (18). In addition, a modified Charlson comorbidity score, which is a cumulative
score of major comorbid conditions, was calculated using the same ICD-9-CM groupings (18, 19).

Derivation dataset for predictive model

The predictive model used in the study was derived for a previous study at our institution. The data for this model were obtained from a retrospective cohort of patients admitted between August 1, 2009 and July 31, 2011 with a primary discharge diagnosis of pneumonia. The data were collected from a time period before the transition to a new EHR system. As discussed previously, the model was made with data that were available after the hospitalization since up-to-date comorbidity data was not available until that time. This dataset will be referred to as the derivation dataset. The data definitions, outlined above are consistent with those in this dataset with a few exceptions. In order to create a dataset that could be used as a validation of the derivation dataset, and would also allow risk prediction per-day of hospitalization, the assumptions seen in Table 2 were made.

Statistical Analysis

All tests were performed using Stata (StataCorp. 2011 Stata Statistical Software: Release 12. College Station, TX: StataCorp LP).

Comparison of administrative to EPL comorbidities

After mapping the ICD-9-CM codes from both these sources to the 33 comorbidity categories, kappa statistics were calculated based on the presence or absence of the comorbidity in administrative data compared to the EPL on the day
of discharge. Charlson comorbidities derived from administrative codes and EPL were correlated using Spearman’s rank correlation because the distribution was right skewed. The means were compared using a paired t-test. *Number of problems on the EPL* and *number of medications* were treated as continuous variables and compared using a paired t-test. Categorical variables such as *patients with pneumonia on EPL* were analyzed using a Chi-square test.

**Derivation of the predictive model**

In order to take into account the reliability of the EPL-derived variables, only variables with a kappa statistic above 0.6 comparing EPL to administrative coding were eligible for inclusion in the regression model. Using the derivation dataset,
variables with a p-value of <0.2 in univariate analysis were included in a logistic regression model against the binary variable, readmission within 30 days. Variables were removed in a backwards fashion from the regression model until all p-values were <0.05. The resulting regression model was evaluated using the area under the receiver operating curve (AUC) and the Hosmer-Lemeshow goodness of fit test (GOF). The logistic function \( P(y|x) = 1/(1+(e^{b_0+b_1x})) \) was calculated for each patient and patients were divided into high, medium and low risk categories based on tertiles of their predicted probability of readmission.

*Validation with new data*

The regression model was then applied to the validation dataset for each day of hospitalization. The derivation model coefficients and risk-cutoffs were used on the validation dataset. AUC and GOF were calculated for each day of hospitalization across all patients’ hospital days. The AUCs from different days were compared using a non-parametric test of AUC equality. Figure 1 shows an overview of the methodology.

IRB approval was obtained for all data collection related to this project.
Step 1: Develop a dataset with data organized per day of hospitalization
Step 2: Compare comorbidities on the problem list on the last day of hospitalization to those documented in administrative discharge coding for this hospitalization
Step 3: Develop a derivation model from the derivation dataset using demographic factors, laboratory values, social history variables, medication variables, and comorbidities from administrative discharge coding, as long as the calculated kappa statistic is greater than 0.6.
Step 4: Use the derivation model to predict 30-day readmission on the validation dataset.
Chapter 4: Results

Validation dataset

418 patients met inclusion criteria and were included in the final validation dataset. The 30-day readmission rate was 20.6%. The basic demographics of this cohort can be seen in Table 3.

Table 3: Demographics of the cohort

<table>
<thead>
<tr>
<th></th>
<th>Readmitted within 30 days (%)</th>
<th>Not readmitted within 30 days (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (median, IQR)</td>
<td>61.2 (51.3-69.4)</td>
<td>59.7 (50.3-72.9)</td>
</tr>
<tr>
<td>Gender female (%)</td>
<td>31 (36.1)</td>
<td>165 (49.7)</td>
</tr>
<tr>
<td>Race, black (%)</td>
<td>31 (36.1)</td>
<td>98 (29.5)</td>
</tr>
<tr>
<td>Length of stay (days) (median, IQR)</td>
<td>5 (3-8)</td>
<td>5 (3-8)</td>
</tr>
<tr>
<td>Admission in last 30 days (%)</td>
<td>24 (27.9)</td>
<td>59 (17.8)</td>
</tr>
<tr>
<td>Number of medications administered on day of admission (median, IQR)</td>
<td>7 (3-12)(^1)</td>
<td>6 (3-11)(^2)</td>
</tr>
<tr>
<td>Number of medications administered on day of discharge (median, IQR)</td>
<td>11 (8-15)</td>
<td>10 (7-13)</td>
</tr>
</tbody>
</table>

Abbreviations: IQR, interquartile range.
\(^1\)17 data points missing
\(^2\)45 data points missing

Electronic Problem List

Only 17(4%) patients, of those who met all other exclusion criteria, had no items on their EPL for the entirety of the study and were excluded from the data analysis. Of those undergoing analysis, on the day of admission 376 (90.0%) patients had a problem on their EPL. By day 15 of hospitalization all patients had at least one problem on the EPL (Figure 2).
Table 4 shows the change in the number of problems on the EPL on the day of admission compared to the day of discharge. The median number of problems increased from 9 to 11. A t-test comparing the mean number of problems on these days showed that this was a significant difference with p < 0.01. Most patients had more problems on the day of discharge than on the day of admission 289 (69.1%). 15 (3.6%) had fewer problems on the day of discharge, and 114 (27.2%) had the same number of problems on the day of admission than on the day of discharge. The
The median number of problems added during the course of admission was 3 with a range of 1-19.

Table 4: Evolution of the EPL over hospitalization and a comparison of pre- and post-billing EPL

<table>
<thead>
<tr>
<th></th>
<th>All patients</th>
<th>Readmitted within 30 days (%)</th>
<th>Not readmitted within 30 days (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of problems on EPL on day of admission (median, IQR)</strong></td>
<td>9 (4-15)</td>
<td>12 (6-18)</td>
<td>8 (4-14)</td>
</tr>
<tr>
<td><strong>Number of problems on EPL on day of discharge (median, IQR)</strong></td>
<td>11 (7-18)</td>
<td>14 (9-20)</td>
<td>11 (6-18)</td>
</tr>
<tr>
<td><strong>Charlson score by EPL on day of admission (excluding those with missing EPL)</strong></td>
<td>2 (1-3)</td>
<td>3 (2-5)</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td><strong>Charlson score by EPL list on day of discharge (median, IQR)</strong></td>
<td>2 (1-4)</td>
<td>3 (2-5)</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td><strong>Charlson score by administrative coding (median, IQR)</strong></td>
<td>2 (1-4)</td>
<td>3 (2-5)</td>
<td>2 (1-4)</td>
</tr>
<tr>
<td><strong>Number of problems on EPL prior to billing (median, IQR)</strong></td>
<td>9 (6-15)</td>
<td>13.5 (8.5-18.5)</td>
<td>8 (5-14)</td>
</tr>
<tr>
<td><strong>Number of problems on EPL after billing (median, IQR)</strong></td>
<td>13 (8-20)</td>
<td>14 (10-22)</td>
<td>13 (7-19)</td>
</tr>
<tr>
<td><strong>Patients with pneumonia on EPL before billing (%)</strong></td>
<td>124 (67.0)</td>
<td>25 (62.5)</td>
<td>99 (68.3)</td>
</tr>
<tr>
<td><strong>Patients with pneumonia on EPL after billing (%)</strong></td>
<td>179 (76.8)</td>
<td>33 (71.7)</td>
<td>146 (78.1)</td>
</tr>
</tbody>
</table>

Abbreviations: IQR, interquartile range; EPL, electronic problem list

The difference between the Charlson score on admission and discharge was statistically significant with a p-value for paired t-test of <0.01, however they were highly correlated with a Spearman’s coefficient of 0.93.

Table 4 also shows the difference in the number of problems on the EPL from before it was uniformly used for physician billing to after. The median number of
problems increased from 9 to 13 between these two time periods (p-value from paired t-test <0.01).

In addition to there being more problems on the list, the list appears to have become more relevant. Overall 72% of patients in this cohort had pneumonia on their EPL at some point during the hospitalization. Before billing this percentage was 67% and after it rose to 77% (p-value 0.03 from Chi-square test).

**Kappa statistics**

There was a wide range of kappa statistics comparing comorbidities derived from administrative codes to those from the EPL (see Table 5). Most major comorbidities (cancer, diabetes, renal failure, HIV, pulmonary disease, liver disease) had kappa values in the good to excellent range. In general, less granular comorbidity categories (diabetes) had higher kappa statistics then their subcategories (uncomplicated or complicated diabetes). Most psychiatric/social comorbidities were in the moderate category. Obesity was the only comorbidity that was below the 0.6 cutoff but was statistically significant by univariate analysis in the derivation dataset.

For 22/33 comorbidities the percentage of patients with a comorbidity by EPL was lower that from administrative codes.
Table 5: Kappa statistics and percentage of patients with comorbidities comparing EPL to administrative coding data

<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>Kappa statistic</th>
<th>EPL</th>
<th>Administrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphoma</td>
<td>0.96</td>
<td>6.9%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.90</td>
<td>29.7%</td>
<td>31.6%</td>
</tr>
<tr>
<td>Rheumatologic</td>
<td>0.81</td>
<td>4.1%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Renal failure</td>
<td>0.80</td>
<td>17.0%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Human immunodeficiency virus</td>
<td>0.80</td>
<td>2.4%</td>
<td>2.4%</td>
</tr>
<tr>
<td>Neurologic disorder</td>
<td>0.77</td>
<td>8.1%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Uncomplicated diabetes</td>
<td>0.76</td>
<td>27.3%</td>
<td>24.2%</td>
</tr>
<tr>
<td>Liver disease</td>
<td>0.75</td>
<td>5.5%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Paralysis</td>
<td>0.73</td>
<td>2.4%</td>
<td>2.2%</td>
</tr>
<tr>
<td>Solid tumor</td>
<td>0.72</td>
<td>21.5%</td>
<td>15.1%</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>0.68</td>
<td>10.5%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>0.68</td>
<td>19.6%</td>
<td>23.4%</td>
</tr>
<tr>
<td>Chronic pulmonary disease</td>
<td>0.66</td>
<td>34.4%</td>
<td>43.8%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.66</td>
<td>50.2%</td>
<td>61.7%</td>
</tr>
<tr>
<td>Complicated diabetes</td>
<td>0.63</td>
<td>6.2%</td>
<td>7.7%</td>
</tr>
<tr>
<td>Metastasis</td>
<td>0.63</td>
<td>6.0%</td>
<td>8.6%</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>0.60</td>
<td>2.4%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.58</td>
<td>11.7%</td>
<td>18.2%</td>
</tr>
<tr>
<td>Pulmonary hypertension/embolism</td>
<td>0.53</td>
<td>8.1%</td>
<td>7.9%</td>
</tr>
<tr>
<td>Drug abuse</td>
<td>0.52</td>
<td>3.8%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Obesity</td>
<td>0.51</td>
<td>10.3%</td>
<td>12.7%</td>
</tr>
<tr>
<td>Depression</td>
<td>0.50</td>
<td>18.2%</td>
<td>17.5%</td>
</tr>
<tr>
<td>Valvular disease</td>
<td>0.45</td>
<td>2.4%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Psychiatric issues</td>
<td>0.44</td>
<td>2.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Uncomplicated hypertension</td>
<td>0.41</td>
<td>49.3%</td>
<td>45.5%</td>
</tr>
<tr>
<td>Abnormal weight loss</td>
<td>0.36</td>
<td>6.7%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Iron deficiency anemia</td>
<td>0.32</td>
<td>3.3%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>0.30</td>
<td>4.1%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Anemia</td>
<td>0.29</td>
<td>3.3%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Electrolyte abnormalities</td>
<td>0.28</td>
<td>3.3%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Complicated hypertension</td>
<td>0.28</td>
<td>3.1%</td>
<td>16.7%</td>
</tr>
<tr>
<td>Coagulopathy</td>
<td>0.19</td>
<td>4.1%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>-0.01</td>
<td>0.5%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>
The EPL-derived Charlson score correlated well with the administrative coded Charlson score. Spearman’s coefficient was 0.72 with p value of <0.01. See Figure 3.

![Figure 3: Scatter plot comparing Charlson scores from EPL on the day of discharge to administrative data](image)

The size of the marker is related to the frequency of that pair of observations

**Derivation dataset**

There were 1042 patients who met criteria in the derivation dataset. The overall readmission rate was 18.7%. The only variable removed due to a
suboptimal kappa statistics was obesity. The resulting model, seen in table 6, had an AUC of 0.71 on the derivation dataset and no evidence of a lack of fit. Separating the model into tertiles resulted in a cutoff for low risk of <0.11, between 0.20 and 0.11 for medium risk and >0.20 for high risk.

Table 6: Final derivation model

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin &lt;10g/dL</td>
<td>1.91</td>
<td>1.35-2.71</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Prior admission</td>
<td>1.95</td>
<td>1.31-2.90</td>
<td>0.001</td>
</tr>
<tr>
<td>Anti-arrhythmic medication given</td>
<td>2.48</td>
<td>1.22-5.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Solid tumor</td>
<td>1.47</td>
<td>1.60-3.79</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Neurologic disease</td>
<td>1.85</td>
<td>1.08-3.17</td>
<td>0.03</td>
</tr>
<tr>
<td>Steroid medication given</td>
<td>1.47</td>
<td>1.01-2.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Number of medications</td>
<td>1.04</td>
<td>1.01-1.07</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Validation of model

The model was initially applied to the last day of hospitalization since this application took into account all that had happened up until that point and was most consistent with the derivation methodology. The AUC of the model on the last day of admission was poor at 0.60, with no evidence of a lack of fit. The model was then applied to the day of admission, the day after admission, the day before discharge and the halfway point of the hospitalization. Figure 4 shows the change in AUC over each of these days of hospitalization. Overall, the AUCs of the model on different
days were not statistically different with a p-value of 0.10 using an ROC test of equality.

![Graph showing change in AUC at different points during hospitalization]

Figure 4: Change in AUC of model at different points during the hospitalization

Of the 356 patients with a calculable risk-score (no missing values) on the day of admission, 10 (2.8%) had no change in their risk score from the day of admission to the day of discharge. 274 (76.8%) patients had an increase in their score, 72 (20.2%) patients had a decrease in the score. The mean decrease was 0.026 (range 0.002-0.122) and mean increase 0.073 (range 0.002 - 0.364). 64 (17.9%) of these patients went from a score of low or medium on the day of
admission to a score of high on the day of discharge. Only 18 (5%) patients went from a score of high or medium on the day of admission to a score of low by discharge. 241 (67.5%) of these patients had no change in their risk category between the day of admission and day of discharge.
Chapter 5: Discussion

In this study, the main goals were to assess the reliability of the EPL, study how the EPL changed over the hospitalization and see whether or not a patient’s risk-score early in the hospitalization would be similar to one at discharge. These are the major findings:

1. Most patients had at least one problem on their EPL by the day of discharge
2. The majority of the patients had more problems on their EPL on the day of discharge then the day of admission
3. Although the number of problems increased, major comorbidity, measured by the Charlson comorbidity score did not have a clinically relevant increase over the hospitalization.
4. Kappa scores comparing the reliability of the EPL versus administrative data were high for most major comorbidities. The reliability was lower for more granular descriptions of disease, and for mental health and drug use variables. The Charlson score by EPL and by administrative billing correlated well.
5. The predictive model performed poorly but its performance varied minimally over the hospitalization.

Despite reports in the literature of severe underuse of the EPL, in this cohort 96% of our patients had at least one problem documented on their EPL by the last day of the index admission. Most patients had items added to their EPL during their hospitalization however the Charlson score on admission and discharge was very similar. This suggests that the problems that were added were transient, rather than major chronic issues and that most major comorbidities were added on the first day
of hospitalization or prior to hospitalization, during an outpatient encounter. In
addition, the number of problems and relevance of the problems increased from
before physicians started using the EPL for billing to after. This suggests that
administrative changes that encourage the use of the EPL may improve the
completeness of the list.

Because the study design did not include a “gold standard” comorbidity
source I was unable to calculate sensitivity and specificity of the EPL. However, for
the majority of the comorbidities, the EPL had good reliability when compared to
administrative codes. The percentage of patients with a comorbidity documented in
the EPL was lower than the percentage of patients with a comorbidity documented
from administrative coding for the majority of comorbidities. This is likely reflective
of the findings of previous investigations, that the EPL is often lacking in sensitivity.
However, the case could be made that since clinicians manage the EPL, it may
actually include more clinically relevant or active comorbidities compared to
administrative coding which is often done by professional coders.

In addition, Table 5 shows that kappa statistics seemed to be higher for more
general categories such as diabetes, rather than the more specific categories of
uncomplicated or complicated diabetes. This likely reflects the fact that physicians
are not professional coders and may not strive to find the most appropriate code if it
does not matter clinically. This has broad implications for the use of this data source
for other applications. For example, if one was trying to find a cohort of patients
with complicated diabetes they may want to use the diabetes code on the EPL along
with other codes that suggest complications, such as *renal disease* rather than just relying on the *complicated diabetes* code which would likely miss many eligible patients.

In general, social comorbidities such as drug use, alcohol use, obesity and depression had lower kappa statistics than many major medical comorbidities. This finding may suggest that clinicians do not think of these as problems that should be listed in an EPL. This is a topic that would be important to address in a hospital-wide policy on EPL best-practices.

The predictive model did not validate well on the validation dataset. The difficulty of creating useful readmission models has been reported in the literature (20). This is likely due to the heterogeneity of reasons why patients return to the hospital (financial, worsening underlying disease, poor quality care), the difficulty in collecting discrete data on social factors that might influence readmission, and the change in risk factors over time as readmission reduction strategies are put in place. Despite the poor performance, the results show that the majority of patients had a similar risk prediction on day one of hospitalization as on the day of discharge. These results provide support for using the readmission model in real-time, during the course of the hospitalization.

*Limitations*

There are several limitations to this study. The first limitation is the study's generalizability. Although it was an in-depth study of the use of the EPL at our institution, it only included a small number of patients, over a relatively short time
period, all of whom were admitted with pneumonia. However, the patients were spread among three hospitals and many different providers so these data likely provided a reasonable reflection of the workflow at our institution over this time period. In addition, one of the reasons that a short time period was chosen was that with the rapid development and evolution of technology at our institution, it is difficult to pick a long time period where conditions are relatively stable in which to study. The second limitation is the lack of manual chart review comparison. Manual review of 418 charts would be a massive undertaking and would hold its own limitations. Because most of the notes in our institution have the EPL embedded in the note it would be extremely hard to blind the reader to the contents of the EPL. If the reader was blinded, they might incorrectly assume that a problem was not mentioned in a note, even though a physician may have assumed that is was documented because it was present on the embedded EPL and so did not think to restate the problem. Despite these limitations, an extensive chart review is a future plan for this study. The third limitation is the poor performance of the predictive model on this cohort of patients. This is likely due to the fact that I was validating the model on a cohort admitted two years later than the cohort that the model was derived on, that I was validating the model on data from a completely different EHR system, and that I took a static model that was focused on one day of the hospitalization and used it for per-day risk prediction. Finally, although I studied the change in number of problems on the EPL I did not study the actual content of the EPL in order to determine its quality. It is possible that although problems are being
added during the course of the hospitalization these may not add to patient care (could be redundant). I also did not study whether clinicians are appropriately removing problems from the EPL when they are no longer active. Although this is a major limitation if I were studying the quality of the EPL, since I was looking only at the reliability of the EPL for comorbidity detection this is less of a limitation. Most comorbidities are unlikely to be removed from the EPL because of their inherent chronicity.

This study provides the basis for many future investigations. One potential avenue of study would be to repeat the study with other predictive models to see whether the findings of this study are reproducible for other models, especially ones that are known to reliably predict outcomes. Another ongoing activity is to study the use of this predictive model prospectively to determine whether or not the prediction is helpful to users and leads to an actual reduction of readmissions.

Conclusion

The findings of this study help to address our three aims. The findings show that the EPL has good reliability for major comorbidities when compared to administrative coding. The EPL is added to during a hospitalization however the general burden of chronic comorbidities does not seem to change significantly from the beginning to the end of the hospital stay. Finally, the performance of the predictive model did not vary significantly over different days of hospitalization.
These findings suggest that it is reasonable to use the EPL to calculate real-time readmission risk throughout the hospital stay.
References


