

Automatic detection of omissions in medication lists

Sharique Hasan,¹ George T Duncan,² Daniel B Neill,² Rema Padman²

¹Graduate School of Business, Stanford University, Stanford, California, USA

²H John Heinz III College, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Correspondence to

Dr Sharique Hasan, Graduate School of Business, Stanford University, 518 Memorial Way, Stanford, CA 94305, USA; hasan_sharique@gsb.stanford.edu

Received 27 October 2009

Accepted 8 February 2011

Published Online First

29 March 2011

ABSTRACT

Objective Evidence suggests that the medication lists of patients are often incomplete and could negatively affect patient outcomes. In this article, the authors propose the application of collaborative filtering methods to the medication reconciliation task. Given a current medication list for a patient, the authors employ collaborative filtering approaches to predict drugs the patient could be taking but are missing from their observed list.

Design The collaborative filtering approach presented in this paper emerges from the insight that an omission in a medication list is analogous to an item a consumer might purchase from a product list. Online retailers use collaborative filtering to recommend relevant products using retrospective purchase data. In this article, the authors argue that patient information in electronic medical records, combined with artificial intelligence methods, can enhance medication reconciliation. The authors formulate the detection of omissions in medication lists as a collaborative filtering problem. Detection of omissions is accomplished using several machine-learning approaches. The effectiveness of these approaches is evaluated using medication data from three long-term care centers. The authors also propose several decision-theoretic extensions to the methodology for incorporating medical knowledge into recommendations.

Results Results show that collaborative filtering identifies the missing drug in the top-10 list about 40–50% of the time and the therapeutic class of the missing drug 50%–65% of the time at the three clinics in this study.

Conclusion Results suggest that collaborative filtering can be a valuable tool for reconciling medication lists, complementing currently recommended process-driven approaches. However, a one-size-fits-all approach is not optimal, and consideration should be given to context (eg, types of patients and drug regimens) and consequence (eg, the impact of omission on outcomes).

INTRODUCTION

Proper prescribing of medication depends on inputs compiled from various sources, such as a patient's demographic characteristics, diagnoses, allergies and their current and past lists of prescription and non-prescription medications. The physician incorporates this information with clinical knowledge to arrive at a decision about which drugs will best address the patient's condition. A lack of relevant information, as well as a failure to process it properly, can negatively affect the prescribing decision, perhaps resulting in adverse drug events (ADEs).

Estimates provided in a 2006 Institute of Medicine report indicate that errors in medication prescription and dispensation cause nearly 1.5

million preventable adverse drug events each year, with approximately 800 000 in long-term care facilities, with financial costs reaching upwards of \$3.5 billion per annum.¹ The failure to reconcile medications at various transition points including admission, transfer between units and discharge has been recognized as contributing to nearly 50% of these errors.² To remedy these failures, medication reconciliation, the process of creating the most accurate list of all medications a patient is taking is the recommended course of action by the Joint Commission on the Accreditation of Healthcare Organizations.

Medication reconciliation consists of three steps: (1) verification, collecting a patient's complete and accurate medication history, including the patient's list of active medications; (2) clarification, determining the appropriateness of medications and doses; and (3) reconciliation, documenting any changes in medication orders before further clinical action is taken.³

In this paper, we focus on the verification step in the medication reconciliation process. This step is the foundation for the remainder of the reconciliation process. Recent studies indicate that there are significant discrepancies between clinic-derived medication histories, admissions orders, patient self-reports and claims data, all sources that could be relied upon when constructing a patient's medication history.^{4–7} For instance, a study comparing self-reported drug consumption against medical records data found that 80.4% of patients had discrepancies, with nearly three discrepancies per patient.⁸ The omission of drugs from a patient's list constitutes the majority of discrepancies, followed by commissions—the presence of drugs which should, in fact, be absent.^{7–9} The medical consequences of discrepancies are not trivial by any means. Discomfort, clinical deterioration or worse can occur in patients if discrepancies are not adequately resolved.^{10–11}

Medication reconciliation

A variety of approaches have been proposed for resolving discrepancies in medication lists.¹² Most focus on improving organizational processes with emphasis on proper assignment of responsibilities, improving communication and increasing access to timely and relevant patient information. The most common approach involves using reconciliation forms that ask patients about their medication history, allergies, and other pertinent health information at different points in the care-giving process. Medical staff are then responsible for ensuring that forms are completed and verified. Form-based interventions do improve list accuracy, and studies evaluating the effectiveness of medication reconciliation find the rate of medication errors

is reduced by 70%, and the rate of adverse drug events reduced by 80%.² While the results are promising, the quantity and impact of adverse drug events are still unacceptably high. Moreover, even well-planned reconciliation efforts are sometimes hindered by inconsistent application and the inability to sustain processes over time.¹³

Recent advances in information technology have also aided medication reconciliation efforts on several fronts. Electronic medical records, prescribing systems, and computerized physician order entry applications provide a means to store medication information in a structured and easily accessible format. Decision-support modules with preprogrammed rules that alert prescribers about potentially harmful interactions are also being integrated into these systems to help reduce errors. The effectiveness of these alerts, however, depend critically on the accuracy of the stored patient information, which in turn depends on the robustness of the reconciliation process.¹⁴

In addition to traditional healthcare information systems, technologies and modules have been designed specifically for the medication reconciliation task.¹⁵ For instance, one implementation of the module electronically incorporates the three-step reconciliation process. In the first step, the system compiles a list of medications prescribed by physicians and recorded in the EMR. The physician or nurse who is conducting the reconciliation is required to query the patient about compliance with the drug regimen. Next, the system asks the nurse to enter any medications that the patient is taking, but are not currently recorded in the EMR's database, including drugs prescribed outside the current setting, herbal supplements, and over-the-counter drugs. This information is obtained through patient self-report, a mode that has been shown to be prone to significant errors and discrepancies.^{6 16 17} After the electronic verification step, the clinician then clarifies and reconciles the list with current orders.

Finally, the linking of health information across organizations could improve medication reconciliation efforts. Healthcare organizations are currently focusing on connecting databases between clinical settings, between clinical settings and pharmacies, and between clinical settings and insurance companies. This type of networking will undoubtedly improve efforts to reconcile medications, but these efforts are still limited by information stored in databases. As such, current approaches to medication reconciliation focus on procedural and organizational issues that provide structure within which verification, clarification, and reconciliation can effectively occur.

Relationship between EMR and medication reconciliation

Patient data stored in electronic medication records and form-based reconciliation processes are complementary processes that can jointly reduce discrepancies. Reconciliation forms allow the reporting of medications not currently reflected in patient records. New information from completed forms could be used to create a more accurate and accessible list of medications for a patient. The less obvious fact is that there may be a relationship in the opposite direction. An electronic record is a repository of medication lists for many patients, with a list, for a given patient, containing drugs recorded for them in a given clinical setting. Electronic patient records also generally include patient demographics, diagnoses, allergies, and other pertinent health information such as laboratory test results. Even basic information, if processed appropriately, could provide insights about potential discrepancies in a patient's medication list. As such, electronic records provide us with the capacity to use medication information from a large population of patients to increase the

accuracy of an individual patient's list. For example, if many patients in the database who were prescribed drug A were also prescribed drug B, the occurrence of drug A in a given patient's list suggests that drug B may also be present but not recorded.

Collaborative filtering

A class of methods, together identified as collaborative filtering, is often used for processing information about entities to make inferences or predictions about the information of other entities. With the growth of the internet and emergence of large databases of user purchases, online retailers are increasingly using collaborative filtering to make predictions about products that an individual may enjoy based on aggregated information from users with similar observed tastes.¹⁸ Successful applications of collaborative filtering include movie recommendations by Netflix and product recommendations on <http://Amazon.com/>. For instance, <http://Amazon.com/> may recommend a product Y to a customer who has recently bought product X, since many people who have previously bought product X have also tended to buy product Y. In this paper, we frame the task of judging the correctness of an existing medication list analogously. Often, the clinician has some, possibly incomplete, information about a patient's medical history, including medications. Based on the list of medications observed by the clinician, say drugs A, we would— analogously to the product recommendation applications—like to infer what other drugs B that patient may also be taking.

In the following sections, we outline a framework for using collaborative filtering as a methodology for the detection of potential omissions of medications from a patient's list. Using five traditional methods for collaborative filtering, we attempt to answer the following question: if a patient's medication list is incomplete, what drugs are most likely to be missing? The output of each method is an ordered list of drugs considered to have the highest likelihood of being omitted from a patient's record. In practical terms, a subset of the ordered list of potentially omitted drugs (eg, the top 5 or 10) can be used to develop individualized memory aids shown to improve recall and could potentially improve reconciliation efforts.¹⁹

This article presents a significant expansion of the work presented in our earlier conference paper.²⁰ The model formulation of the current paper is extended by incorporating two additional sets of predictors, demographic data and diagnoses, in addition to medications. Furthermore, four new models based on the logistic regression approach to collaborative filtering are presented. The logistic regression models not only incorporate the new predictors but also offer two kinds of approaches for dealing with the high dimensionality of the set of predictors—dimensionality reduction using principal-component analysis and a covariate penalization approach using the methodology developed in Park and Hastie.²¹ Two additional detailed case studies in two new clinics are also added to the article with a more general analysis of the larger sample of clinics using the three computationally efficient algorithms—popular, K-nearest neighbor (KNN), and co-occurrence to examine how characteristics of clinics affect the effectiveness of these algorithms.

The remainder of the paper is organized as follows. The section 'Model formulation' presents our formulation of the verification step of medication reconciliation as a collaborative filtering problem using data on patient medications, demographics, and diagnoses. In the section 'Collaborative filtering methods,' we describe five computational and statistical methods for collaborative filtering. The sections 'Medication data for model validation' and 'Simulation experiment and cross-validation' provide

a description of these data and our simulation experiment. In the sections ‘Discussion and future work’ and ‘Implications and future work’, we provide a discussion of our results and their implications for the medication reconciliation process, and provide a discussion of our contributions, some limitations, and potential directions for future research.

MODEL FORMULATION

Formal representation of the medication list

The central piece of information in the medication reconciliation task is a patient’s list of medications. This list is a set of entities, where each entity represents a drug. The most granular view of a drug entity is a brand name drug with an associated dose and route (eg, Tylenol Oral Tablet 325 MG). This same entity can also be viewed in more general terms, as a brand name drug (eg, Tylenol), as a generic chemical name (eg, Acetaminophen), or more generally as a member of a therapeutic class (eg, Non-Narcotic Analgesics). Our original data (discussed further in the section ‘Medication data for model validation’) contained only the branded drug with dose and route. To provide more general classifications of the drug into generic chemical names or therapeutic classes, we used the Center for Disease Control’s Ambulatory Care Drug Database System (<http://www2.cdc.gov/drugs>) to classify each drug–dose–route entity into its respective branded drug code, generic code, and therapeutic class code.

Irrespective of the granularity of the drug entities, we can represent the complete and accurate medication list of all patients in a population as a matrix $M=\{m_{ij}\}$, for patients $i=1,\dots,I$ and drugs $j=1,\dots,J$, and where:

$$m_{ij} = \begin{cases} 1 & \text{if drug } j \text{ occurred in medication list of patient } i \\ 0 & \text{otherwise} \end{cases}$$

Analogously we can represent the medication list is the set of lists l_i , where l_i constitutes the set of drugs j for a given patient i and $J \in l_i$ if and only if $m_{ij}=1$.

Knowledge of a patient’s true medication list l_i is often incomplete. Hence, prescribers observe only a partial list of drugs for a patient, denoted by \tilde{l}_i . The observed partial list may be incomplete for several reasons (figure 1). These include the failure to record a previous prescribing decision or the unintentional or intentional omission of a drug during patient self-report. The actual probability of omitting a given drug from a patient’s list depends on a variety of factors. For instance, over-the-counter drugs or herbal supplements may have a higher probability of being missing than those prescribed by the current provider. On the other hand, omission of drugs may occur with no discernible pattern, that is to say each drug has an equal chance of being omitted.¹⁶

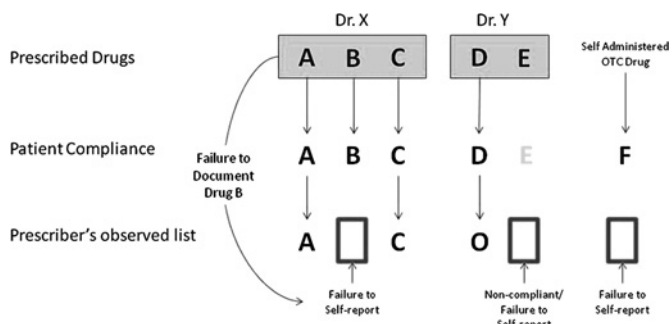


Figure 1 Schematic of how errors may be introduced into a patient’s medication list. OTC, over the counter drug.

Distributions notwithstanding, discrepancies could result in a variety of negative outcomes for the patient, from simple duplication of medications to severe adverse consequences from prescribing drugs that negatively interact with drugs the patient is currently taking but are not recorded in the observed list.

Formal representation of diagnoses and demographic factors

In addition to the patient’s list of medications, we observe two additional patient characteristics which may aid in predicting omitted drugs. The first is the list of coded diagnoses for that patient which were provided to us in the form of three-digit ICD9 codes. We can represent the diagnoses list of all patients in a population as a matrix $D=\{d_{it}\}$, for patients $i=1,\dots,I$ and diagnoses $t=1,\dots,T$, and where:

$$d_{it} = \begin{cases} 1 & \text{if medical record of patient } i \text{ indicates diagnosis } t \\ 0 & \text{otherwise} \end{cases}$$

We also observe two demographic variables, age and sex, represented as a_i and S_i , respectively. The age variable is calculated using the patient’s date of birth, and the sex variable is coded as 1 for males and 0 for females.

COLLABORATIVE FILTERING METHODS

In most applications, the goal of collaborative filtering is to make predictions about products an individual may enjoy based on the aggregate tastes of similar individuals. In our case, we predict whether specific drugs have been omitted from an individual’s medication list based on the known medications of similar individuals and the observed list of medications for that patient. Many computational and statistical methods for collaborative filtering exist, each with its advantages. In this study, we use five methods for ranking the drugs not observed in the partial list.²² In each case, the algorithm assigns a score p_j for each drug j not observed in the partial list. We then sort the drugs in decreasing order based on this score. We assume that the drug with the highest score is the one with the highest probability of being missing from the partial list. Formal descriptions of the standard popular, co-occurrence, and KNN approaches can be found in the review paper by Goldenberg *et al.*²²

Model 1: Drug popularity

The ‘popular’ algorithm considers each drug j not observed in the partial list \tilde{l}_i , counts the number of lists l_{-i} which contain drug j in the training set, and chooses the most commonly occurring drugs. Given a patient i , the score p_j for each drug j is assigned according to the following equation, where $I(x)$ is the indicator function, returning 1 if x is true and 0 otherwise:

$$p_{ij} = \text{score}(ij) = \begin{cases} \sum_{l_{-i}} I(j \in l_{-i}) & j \notin \tilde{l}_i \\ 0 & j \in \tilde{l}_i \end{cases}$$

The popular algorithm can be expected to perform well if there are a small number of very common drugs (eg, aspirin) that occur in many lists. For comparing the performance of the more complex models to follow, excluding the random algorithm, the popular algorithm is considered the baseline.

Model 2: co-occurrence counting

The ‘co-occurrence counting’ algorithm scores each drug j not present in the observed partial list \tilde{l}_i according to the number of times it has co-occurred with drugs g that are observed in the partial list, $g \in \tilde{l}_i$. For each patient i , we calculate the score for each drug as follows:

$$p_{ij} = \text{score}(ij) = \begin{cases} \sum_{g \in \tilde{I}_i} \sum_{L_i} I(g \in L_{-i}) \times I(j \in L_{-i}) & j \notin \tilde{I}_i \\ 0 & j \in \tilde{I}_i \end{cases}$$

The co-occurrence counting algorithm is expected to do well when there is a strong pairwise structure in the prescribing patterns (eg, pairs of drugs are regularly prescribed together). For instance, we may observe a patient taking drugs A, B, and C. The co-occurrence algorithm would assign a score to a candidate drug X by counting the number of co-occurrences of A and X, B and X, and C and X. The candidate drug with the highest total number of co-occurrences would be recommended first, and so on.

Model 3: KNNs

KNN is a memory-based machine learning approach used widely for the purposes of collaborative filtering.²⁵ Given an observed partial list, we find the K training lists $L_1 \dots L_K$ that are closest to it according to a distance metric. Scores for the missing drugs are assigned using a vote of the KNN, where each neighbor votes ‘1’ if the missing drug is present and ‘0’ if not present. In this study, we use the Ochiai Similarity Measure, the binary form of cosine similarity, to compare the observed partial list \tilde{I}_i with each of the lists L_i in the training set. We define a as the number of drugs that are present in both lists, b as the number of drugs present in \tilde{I}_i but not in L_{-1} , and c as the number of drugs present in L_{-1} but not in \tilde{I}_i . The Ochiai similarity measure is defined as:

$$\text{Dist}(\tilde{I}_i, L_i) = \sqrt{\left(\frac{a}{a+b}\right)\left(\frac{a}{a+c}\right)}$$

The nearest-neighbors approach is expected to do well when there are many patients on similar drug regimens. Because these data are relatively sparse, we use a smoothed nearest-neighbors approach, which is a weighted average of base rates and the votes of the nearest neighbors. We specify the smoothed nearest neighbor vote as:

$$p_j = \frac{r_j s + v_j^K}{s + K}$$

In this equation, the parameter s is the strength of the base-rate information, and r_j is the base rate for drug j, calculated as the number of patient drug lists containing drug j divided by the total number of lists I. The smaller the value of s, the less emphasis is placed on the base rates. The term v_j^K is the number of occurrences of drug j, where K indexes the number of the nearest neighbors in list \tilde{I}_i . We use $K=3$ and $s=1$, chosen by cross-validation, using the training samples, in our evaluation discussed below.

Models 4a–d: logistic regression

Logistic regression is a statistical approach for estimating the probability of a binary response given a set of predictors. In the expression below, $p_{ij} = \text{Pr}(m_{ij}=1)$ is the probability that drug j is on patient i’s medication list, and $x_{1,i} \dots x_{K,i}$ are the predictors for the probability of that event.²⁴

$$\text{logit}(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_0 + \beta_1 x_{1,i} \dots \beta_K x_{K,i}$$

In our scenario, the logistic regression equation models the probability that drug j is on patient i’s medication list, given that we observe their binary vector of other drugs $m_{i,-j}$, excluding drug j, the patient’s vector of diagnoses d_{it} , and the patient’s age a_i and sex S_i .

Dealing with high dimensionality of predictors: penalization and regularization

The unknown parameters of the logistic regression model $\beta_0 \dots \beta_K$ are generally estimated by maximum likelihood. In our scenario, these parameters can be estimated using the submatrix $M_{i,-j}$ of drug information excluding drug j, the diagnosis matrix D, and the vectors of demographic information (age and sex). The maximum-likelihood approach estimates the set of coefficients β by satisfying the following criterion, where L is the likelihood function with respect to the given data $\{(x_i, y_i) : i=1, \dots, I\}$: $\hat{\beta} = \arg \max_{\beta} \{L(y; \beta)\}$

However, the traditional maximum likelihood approach poses significant difficulty when the number of predictors J exceeds the number of observations I, or when there are covariates that do not contribute to increasing the predictive accuracy of the model. We can address the limitations of this approach in several ways. One option is to impose a complexity penalty on the coefficients. For our analysis, we impose a penalty on the L_1 norm of the coefficients, which is the sum of the absolute values of the β s. This approach forces many coefficients to 0 and produces automatic variable selection.²¹ To impose the L_1 norm, Park and Hastie²¹ modify the maximum-likelihood criterion with a regularization penalty to: $\hat{\beta}(\lambda) = \arg \min_{\beta} \{-\log\{L(y; \beta)\} + \lambda \|\beta\|_1\}$. This convex optimization problem can be solved using a predictor-corrector method as described in Park and Hastie.²¹

Dealing with high dimensionality of predictors: dimensionality reduction via principal-component analysis

A second approach to dealing with the large number of predictors is through some form of statistical dimensionality reduction. Principal-component analysis (PCA) is one such approach that reduces multidimensional data into lower dimensions for analysis purposes.²³ The PCA approach uses Eigenvalue decomposition to find d orthogonal linear combinations of the original data, which explain the most variance in the data. Here, d is assumed to be less than or equal to the original number of dimensions of the data. The first principal component would explain the most variance in the data, with the second principal component explaining the second most variance, and the last principal component explaining the least.

Model 4a: drugs with regularization

The first model we estimate using the logistic regression framework is one where we model the probability p_{ij} of a given drug j occurring in a patient’s list given the other drugs they are taking $m_{i,-j}$. For this estimation, for each of the J drugs, we use the Park and Hastie²¹ regularization approach to calculate the logistic regression coefficients from the matrix $M_{i,-j}$ of all patients’ drug information excluding drug j. Using the Park and

Table 1 Summary statistics for clinics used in the case study

Statistics	Clinic 01	Clinic 02	Clinic 03
# of Patients	182	153	84
# of Unique drugs	177	140	128
# of Unique therapeutic classes	64	59	56
Median drugs/patient (Q1, Q3)	14 (7, 21)	14 (9, 19)	16 (12, 18)
Median # of drugs prescribed (Q1, Q3)	9 (6, 15)	9 (5, 16)	7 (5, 12)
Median diagnoses/patient (Q1, Q3)	8 (4, 11)	13 (11, 16)	14 (8, 18)
Median age of patients (Q1, Q3)	82 (65, 89)	82 (76, 87)	82 (75, 88)
Proportion of females	51%	50%	72%
Gini coefficient of drug distribution	0.47	0.53	0.43
Gini coefficient of diagnosis distribution	0.87	0.91	0.88

Table 2 Most commonly prescribed drugs for each clinic

Clinic 01 Drug Code—Drug Name	Clinic 02 Drug Code—Drug Name	Clinic 03 Drug Code—Drug Name
32905—Tylenol (101)	10575—Dulcolax (145)	10575—Dulcolax (68)
10575—Dulcolax (100)	12620—Fleet Enema (144)	19375—Magnesium Antacid (65)
19375—Magnesium Antacid (98)	19375—Magnesium Antacid (144)	12620—Fleet Enema (64)
12620—Fleet Enema (96)	98126—Fluvirin (101)	32905—Tylenol (62)
32695—Tubersol (79)	32905—Tylenol (95)	34520—Vitamin C (35)

Hastie²¹ estimation procedure, even if we have fewer observations than the number of predictors, we are still able to estimate the logistic regression coefficients, since many of the coefficients will be forced to zero by the penalty term.

Model 4b: drugs with PCA

Model 4b does not use the matrix of all other drugs $M_{i,-j}$ directly as predictors. Instead, we use the first k principal components of $M_{i,-j}$ which we denote as P_{-j}^k . To estimate this model, we use a Bayesian estimate of the logistic regression parameters.²⁵ We specify Cauchy priors on the constant coefficient centered at 0 and with a scale parameter of 10, and on the other coefficients centered at 0 and with a scale parameter of 2.5. This approach helps prevent complete separation in the data, which is often a problem if there are many more zeros than ones in the data or vice versa. This often seems to be a problem in our analysis, since certain classes of drugs are rarely prescribed, and therefore traditional logistic regression estimation invariably forces the probability of their occurrence to zero.

Model 4c: drugs, diagnoses, and demographics

The third model we estimate using the logistic regression framework is one where we model the probability of a given drug j occurring in patient i 's medication list given their observed diagnoses d_{it} , which other drugs they are taking $m_{i,-j}$, and their age a_i and sex S_i . For this estimation, we again use the Park and Hastie²¹ regularization approach to compute the logistic regression model for each drug J .

Model 4d: dimensionality reduced drugs and diagnoses

The final logistic regression model does not use the matrices of all other drugs $M_{i,-j}$ or diagnoses D , directly as predictors. Instead, we use the first k principal components of $M_{i,-j}$ and D , which we denote as P_{-j}^k and G^k as our predictors respectively, in addition to the two demographic variables. We again use a Bayesian estimate of the logistic regression coefficients by specifying Cauchy priors on the constant coefficient centered at 0 and with a scale parameter of 10. Cauchy priors on the remaining coefficients are centered at 0 with scale parameters set to 2.5.

Model 5: random

Finally, to establish a single baseline for comparison of these methods, we use the random algorithm, which uses no information

from the patient database. For each drug not observed in the partial list, the algorithm assigns a score $p_j = [0,1]$ uniformly at random.

MEDICATION DATA FOR MODEL VALIDATION

To evaluate the effectiveness of the model formulation presented in this paper, we obtained medication data from an online pharmacy that serves long-term care clinics in the eastern USA. Human Subjects approval for conducting the empirical case studies was granted by the Institutional Review Board of our institution, and all patient data were anonymized before being provided to the researchers. Three clinics from a population of 61 were selected as case studies for evaluating the efficacy of the model formulation. Two fundamental points should be made regarding the case studies presented in this article. First, the generalizability of the statistical results is not a purpose of the case studies, but rather a demonstration of the framework's applicability in accurately predicting missing drugs. The performance of any given algorithm should be expected to vary by clinical care setting depending on the distribution of the underlying data. For instance, see Goldenberg *et al*²² for the differential performance of each algorithm depending on the data set used. Thus, any institution applying the framework should thoroughly evaluate the effectiveness of the algorithms on their data before implementation. Second, the three clinics were selected to demonstrate precisely that a one-size-fits-all approach may not be appropriate, and algorithms with different properties may work better in certain settings, while not performing well in other settings. For purposes of brevity, we also conducted an analysis on all clinics available to us using three of the efficient algorithms (popular, co-occurrence and KNN) in an attempt to determine the characteristics that would allow algorithms that use a focal patient's information (co-occurrence and KNN) to perform better than base rates alone (popular).

Basic statistics about patients, drugs, and diagnoses for these three clinics are presented in tables 1–3, respectively. As we can see in table 1, the patient populations for the first two clinics are evenly split between males and females, with a median age at 82 years. The third clinic has a much higher proportion of females at 72%, but an age distribution that is similar to the first two clinics. We also see that the Gini coefficient for the drug distribution is highest for Clinic 2. This suggests that there is

Table 3 Most common diagnoses for each clinic

Clinic 01 Diagnosis Code—Diagnosis	Clinic 02 Diagnosis Code—Diagnosis	Clinic 03 Diagnosis Code—Diagnosis
401—Hypertension (74)	298—Nonorganic Psychoses (126)	401—Hypertension (63)
311—Depression (62)	443—Vascular Disease (106)	311—Depression (45)
414—Heart Disease (44)	401—Hypertension (97)	788—Urinary system symptoms (44)
428—Heart Failure (41)	311—Depression (92)	530—Disease of Esophagus (42)
520—Tooth problems (40)	331—Cerebral degenerations (84)	272—Disease of lipid metabolism (34)

```

Input: Drug matrix M, Diagnoses matrix D, age vector a, sex vector s,  $K=5$ 
Subset data M, D, a, and s into K, approximately equal, random chunks
For  $K=1...5$  {
  Remove  $k^{\text{th}}$  chunk as test data, use remaining chunks as training data
  For  $i$  in Test {
    Estimate Models 1-4d using training data
    Take true list of patient  $i$  and randomly remove one drug, transforming  $l_i \rightarrow \tilde{l}_i$ 
    Use Models 1-4d with  $\tilde{l}_i$  to create an ordered list of  $e_j \notin \tilde{l}_i$ 
    Determine the position of the randomly removed drug in the ordered list of  $j \notin \tilde{l}_i$  for each model
  }
}
    
```

Figure 2 Summary of the algorithm for simulation experiment examining the effectiveness of the algorithms for the automatic 5 detection of omissions in medication lists.

more inequality in the prescribing of drugs in Clinic 2 versus Clinic 1 or 3, namely that there are fewer drugs that make up most of the prescriptions.

For our analysis, we also removed drugs that were prescribed to fewer than two patients, since they would be difficult to predict based on data-driven methods alone. This resulted in 20 patient exclusions in clinic 1, 0 in clinic 2, and 1 in clinic 3. We find that the most commonly prescribed drugs in all three clinics are quite similar. Table 2 presents the top five drugs for each clinic. For the most part, these are quite similar across the clinics, with Tylenol, Dulcolax, Fleet Enema, and a magnesium antacid being highly prescribed in all clinics.

Analysis of the patient diagnoses, in the form of ICD9 codes, also indicates relative uniformity in the types of diagnoses present in these three clinics. The most common diagnoses across the three clinics are presented in table 3. These include hypertension, depression, heart disease, and vascular disease.

SIMULATION EXPERIMENT AND CROSS-VALIDATION

We use a cross-validation approach to test how well each of the collaborative filtering methods described earlier performs in predicting omitted drugs. We begin by randomly dividing these data into $K=5$ segments of approximately the same size.²⁶ In the first iteration of our simulation, we delete the first segment from the data and use it as test data. We construct our training data with the remaining records, and use this training data to compute Models 1–4d. Then, for each patient i in our test data, we randomly remove one drug from their list l_i to construct our observed list \tilde{l}_i . Next, we use the output of Models 1–4d in conjunction with information in \tilde{l}_i to rank the list of drugs $j \notin \tilde{l}_i$ that have not been observed for patient i . We then use the ranked candidate drug list to determine the position of the drug that we removed from the true list l_i to evaluate how well each of the models performed. The higher the rank of the omitted drug, the better the model performed. We repeat the outer loop

an additional four times removing each of the remaining segments to create our test and training samples. Figure 2 provides the pseudo-code for our simulation experiment.

RESULTS

In this section, we present the results of our simulation experiment for the three clinics.

Case study clinic 1

Our experiments compared Models 1–4d described in the previous section. We attempted to predict the correct branded drug that was missing from a patient’s record (without dose and route information). Table 4 summarizes these results for the first of the three clinics we evaluated. Columns headed 1, 10, 25, 50, and 100 present the proportion of patients whose missing drug is ranked at or better than 1, 10, 25, 50, and 100 in the ordered list of candidate drugs generated by each of the algorithms. For instance, the value of 21% for the KNN algorithm in Column labeled ‘1’ indicates that 21% of the time, the KNN algorithm was able to predict the missing drug correctly on the first guess. A value of 42% for the KNN algorithm in Column labeled ‘10’ indicates that the algorithm was able to predict correctly the missing drug within the top 10 guesses, and so on. Asterisks next to the algorithm name indicate that the result was significantly better than the popular algorithm at *** $\alpha=0.01$, ** $\alpha=0.05$, or * $\alpha=0.1$ respectively, using the Friedman Rank-sum test. The Friedman rank-sum test is the recommended test when comparing the predictive quality of competing algorithms in machine-learning tasks.^{27–29} The median and mean columns indicate the median and mean rank of the omitted drugs in the ordered list generated by the algorithms.

The popular algorithm, which used only the base rates, was able to guess the missing drug within the first 10 guesses 31% of the time and required a median of 24.5 and a mean of 47.94 guesses to identify correctly the omitted drug. The modified KNN (with $K=3$ and $s=1$, selected using cross-validation on the training data sets) performed better than the popular algorithm, significantly improving the percentage correct 42% and improving the median and mean number of guesses to 17.5 and 34.48, respectively. The co-occurrence algorithm as well as the logistic regression models 4b–d also performed significantly better than the popular algorithm. This suggests that using information about the patient, particularly about the drugs we observe for them \tilde{l}_i as well as their diagnoses, d_{it} can be useful in predicting missing drugs. We, however, did find evidence that the more computationally intensive logistic regression models 4b performed better than co-occurrence (Friedman test: $p<0.01$, change in median: 3.5), but model 4d did not (Friedman test: $p=0.37$, change in median: 1.99).

We also used these algorithms to predict the therapeutic class of the missing drug, by first predicting the drug and then

Table 4 Results of collaborative filtering simulation for Clinic 1 with prediction of brand drug

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	14	31	51	64	81	47.94	24.5	7	84	175
(2) Co-Occurrence***	18	40	53	71	84	39.27	19.5	4.75	58.2	173
(3) K-Nearest Neighbors***	21	42	58	74	89	34.48	17.5	2	51.2	167
(4a) Logit—Penalized (Drugs)*	22	36	49	69	86	42.10	26.5	2.75	61.8	175
(4b) Logit—PCA (Drugs)***	22	39	56	71	85	40.64	19	2	68.8	168
(4c) Logit—Penalized (D, D, D)	19	37	52	71	86	40.56	25	3	63.2	168
(4d) Logit—PCA (D, D, D)***	16	39	57	69	83	39.73	18	4	67.8	162
(5) Random	1	7	19	32	62	82.29	81	41.8	124	172

Results statistically different from Popular at *** $p<0.01$, ** $p<0.05$, and * $p<0.1$ using the Friedman rank-sum test.

Table 5 Results of collaborative filtering simulation for Clinic 1 with prediction of therapeutic class

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	16	46	79	92	100	16.52	12	3.75	23	60
(2) Co-Occurrence**	21	54	77	95	100	14.64	9	3	21.5	64
(3) K-Nearest Neighbors***	23	53	81	96	100	13.94	8	2	20	58
(4a) Logit—Penalized (Drugs)	23	49	76	96	100	15.52	11	2	25	62
(4b) Logit—PCA (Drugs)***	24	53	81	97	100	14.34	9	2	22	61
(4c) Logit—Penalized (D, D, D)	22	50	79	95	100	15.39	10.5	2	23.2	61
(4d) Logit—PCA (D, D, D)***	19	54	79	98	100	14.34	10	2.75	20	63
(5) Random	1	26	49	91	100	25.95	26.5	10	41	62

Results statistically different from Popular at *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$ using the Friedman rank-sum test.

choosing the corresponding therapeutic class. For instance, if the prediction was ‘Allegra,’ then we generalize the result to the therapeutic class ‘Antihistamines.’

Table 5 presents the results when we generalized the algorithms’ predictions to the therapeutic class. Our results improved for two reasons: (1) the number of therapeutic classes was 64 versus 177 brand name drugs, and (2) one drug from a therapeutic class is often substituted for another. The therapeutic class is therefore easier to predict than the brand name. The KNN algorithm guessed the missing drug class within the first 10 guesses 53% of the time. The KNNs, co-occurrence counting, and logistic regression models 4b and 4d performed significantly better than the popular algorithm.

Case study clinic 2

Tables 6 and 7 summarize these results for the second clinic.

The popular algorithm, which used only the base rates, was able to guess the missing drug on within the top 10 guesses 53% of the time and required a median of 8 and a mean of 27 guesses to correctly guess the omitted drug. This is significantly better than the first clinic (Wilcoxon test: $p < 0.01$, change in median: 9.99). This may have to do with differences in the nature of drug prescribing in Clinic 2 versus Clinic 1, particularly the inequality of the drug distribution. Furthermore, we find that models 3–4c perform significantly better than the popular algorithm, at least at the $p < 0.05$ level. For the second phase of the simulation, we also used these algorithms to predict therapeutic class of the missing drug, by first predicting the drug and then choosing the corresponding therapeutic class. The only algorithm that performed significantly better than the popular algorithm was model 4b, the Logic-PCA (drugs) model at the $p < 0.01$ level, with improvements particularly at the high end of the distribution, reducing the third quartile from 18 to 16 guesses. Overall, the popular algorithm scenario requires a median of four guesses to correctly identify the correct therapeutic class of the drug and a mean of approximately 12 guesses (table 7).

Case study clinic 3

The popular algorithm, which used only the base rates, was able to guess correctly the missing drug within the first 10 guesses 46% of the time and required a median of 13 and a mean of 27 guesses to guess correctly the omitted drug. This, again, is significantly better than the first clinic (Wilcoxon test: $p < 0.01$, change in median: 8.99) and may have to do with differences between clinics in the nature of drug prescribing. We also do not find that any other algorithms perform significantly better than the popular algorithm in this scenario, either for predicting drugs or for predicting therapeutic classes, although we see that the Bayesian logistic regression with the principal components did qualitatively better, reducing the median guesses from 13 to 10.5 (tables 8 and 9).

Combined analysis

We conducted two additional kinds of analyses. First, we conducted an analysis to examine how algorithms would perform if data from clinics 1, 2, and 3 were combined into a single data set. A similar pattern of results appears with the co-occurrence algorithm performing significantly better than the popular algorithm (Friedman test: $p < 0.01$, change in median=5.5). We also find that the two logistic regression models that use the PCA dimensionality reduction, 4b (Friedman test: $p < 0.01$, change in median: 2.499) and 4d (Friedman test: $p < 0.01$, change in median: 6.5), also perform significantly better than the popular algorithm. However, there is evidence that the simple co-occurrence algorithm performs better than the more complex logistic regression model 4d (Friedman test: $p < 0.05$, change in median: (1). A similar pattern of results hold for the therapeutic class analysis; we do not present these results for purposes of brevity.

The detailed analysis of these three clinics suggests that simple algorithms such as popular or co-occurrence may perform as well as or significantly better than more computationally intensive models such as logistic regression. We also conducted supplementary analysis to examine when algorithms such as

Table 6 Results of collaborative filtering simulation for Clinic 2 with prediction of brand drug

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	39	53	66	77	95	27.08	8	1	48.8	134
(2) Co-Occurrence	37	55	65	79	95	26.35	7	1	42	129
(3) K-Nearest Neighbors**	38	53	68	81	96	24.21	6.5	1	42.5	116
(4a) Logit—Penalized (Drugs)***	37	49	60	77	95	28.12	11	1	43.8	135
(4b) Logit—PCA (Drugs)**	38	53	67	80	93	25.94	8	1	40.8	130
(4c) Logit—Penalized (D, D, D)**	39	52	67	79	99	23.75	7.5	1	37	104
(4d) Logit—PCA (D, D, D)	37	54	64	77	93	27.25	7	1	46	128
(5) Random	1	8	17	35	79	67.25	70	39.5	95.5	133

Results statistically different from Popular at *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$ using the Friedman rank-sum test.

Table 7 Results of collaborative filtering simulation for Clinic 2 with prediction of therapeutic class

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	40	61	83	98	100	11.78	4	1	18	57
(2) Co-Occurrence	39	63	86	97	100	11.39	4	1	18.8	57
(3) K-Nearest Neighbors*	39	65	85	99	100	10.75	4	1	16.8	54
(4a) Logit—Penalized (Drugs)	38	61	85	98	100	11.60	4.5	1	19.8	57
(4b) Logit—PCA (Drugs)***	38	67	84	99	100	10.60	4	1	16	56
(4c) Logit—Penalized (D, D, D)	41	65	86	100	100	10.73	5	1	15.8	50
(4a) Logit—PCA (D, D, D)	37	65	85	99	100	10.76	4	1	16.8	56
(5) Random	1	29	65	93	100	21.84	17.5	9	33.8	57

Results statistically different from Popular at ***p<0.01, **p<0.05 and *p<0.1 using the Friedman rank-sum test.

KNN or co-occurrence perform better than the simple popular algorithm using 60 clinics out of the larger sample of 61 (one clinic was excluded because it had fewer than five patients). These results suggest that KNN and co-occurrence tend to work better than popular when there is diversity in the age of the patients, and patients are prescribed approximately equal number of drugs (eg, there are not only a few patients who are taking most of the drugs). This suggests that algorithms that use more information about patients perform better when there is greater heterogeneity in the patient population.

Overall, our results suggest that simple collaborative filtering approaches such as the popular algorithm, co-occurrence counting, and KNN may work as well as or better than more complex models such as logistic regression. Table 10 provides a summary of each algorithm, the sample running times from clinic 1 (run on a MacMini with OS X version 10.4.11 with a 1.83 GHz Intel Core Duo and 1 GB of RAM) and several pros and cons of each approach. Since the logistic regression approaches do not currently have published computational complexity defined, we use running time for onefold on clinic 1 as a metric for comparing between the computational requirements of each algorithm for both training and prediction purposes.²¹ The results lead us to believe that it may be worthwhile to begin implementation of collaborative filtering systems for medication reconciliation with simple algorithms such as popular, co-occurrence, or KNN, as they have performed better in our case studies than the more computationally intensive logistic regression models. Furthermore, the Logit—PCA models may be initially more appropriate when incorporating larger amounts of information since they predict as well as or better than the Logit—Penalized models and require significantly less time for model estimation and training.

DISCUSSION AND FUTURE WORK

Medical data are increasingly being structured, stored, and linked across organizational boundaries. Although this phenomenon will improve efforts to reduce medication errors, better access to information alone cannot fully address the problem. There will

always be occasions where important information is not stored in any available database. Using a collaborative filtering approach, we are able to look beyond what is recorded for a particular patient, using information from many other patients' records to predict omissions and improve the accuracy of each individual patient's medication list.

Contributions

The contributions of this article beyond the earlier work presented in Hasan *et al*²⁰ are threefold. First, we present a general formulation of the verification step in the medication reconciliation process as a collaborative filtering problem which can potentially include information about three kinds of patient characteristics: drugs, diagnoses, and demographics. Second, we describe and implement several algorithms ranging in complexity, from simple models, such as the popular algorithm that uses only base rates, to more complex logistic regression models that use information about a patient's known medication histories, diagnoses, and demographic characteristics to predict the likelihood that drugs are missing from their record. Finally, we evaluate the effectiveness of these algorithms in the context of three long-term care clinics, in detail, and 60 clinics with the computationally efficient algorithms. We find that simple algorithms such as popular perform relatively well compared with more computationally intensive algorithms such as KNN or logistic regression. Algorithms such as co-occurrence counting and KNNs do well in clinics that have a greater diversity of patients in terms of age and where the number of drugs per patient are more equally distributed.

Limitations

One limitation of our current research is that it assumes a fixed set of drugs, drawn only from a limited set of long-term care centers. A second limitation is the implicit assumptions about the accuracy of the training data—including medication, demographic, and diagnosis data. In real-life situations, the training data may also be of uncertain quality. However, addressing this is beyond the scope of the current article. In future work, we plan to evaluate how robust the predictions are

Table 8 Results of collaborative filtering simulation for Clinic 3 with prediction of brand drug

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	31	46	61	80	95	27.06	13	1	42.8	116
(2) Co-Occurrence	31	46	64	78	94	27.54	12.5	1	39.8	124
(3) K-Nearest Neighbors	28	48	65	79	96	25.53	14	1	32.2	109
(4a) Logit—Penalized (Drugs)	30	40	61	79	96	28.16	17.5	1	43	115
(4b) Logit—PCA (Drugs)	26	50	61	78	98	27.05	10.5	1	48.2	108
(4c) Logit—Penalized (D, D, D)	34	49	70	83	99	22.48	12.5	1	31.5	109
(4d) Logit—PCA (D, D, D)	28	43	58	80	99	27.25	16.5	1	44.2	107
(5) Random	0	8	24	44	91	59.43	63	29	90	124

Results statistically different from Popular at ***p <0.01, **p <0.05, and *p <0.1 using the Friedman rank-sum test.

Table 9 Results of collaborative filtering simulation for Clinic 3 with prediction of therapeutic class

Algorithm	1 (%)	10 (%)	25 (%)	50 (%)	100 (%)	Mean	Median	1st quartile	3rd quartile	Max
(1) Popular	31	60	81	100	100	12.64	6	1	18	49
(2) Co-Occurrence	33	61	84	100	100	11.70	6	1	17.8	50
(3) K-Nearest Neighbors	28	60	88	99	100	11.53	5.5	1	20	51
(4a) Logit—Penalized (Drugs)	34	50	86	99	100	12.60	11	1	17.2	51
(4b) Logit—PCA (Drugs)	28	65	80	100	100	12.51	6.5	1	21	48
(4c) Logit — Penalized (D, D, D)	35	59	86	99	100	11.99	7.5	1	17	51
(4d) Logit—PCA (D, D, D)	28	59	79	100	100	12.81	7.5	1	19.2	44
(5) Random	4	30	60	98	100	22.05	18.5	9.75	35.2	54

Results statistically different from Popular at *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$ using the Friedman rank-sum test.

to relaxation of these assumptions. Furthermore, the results of this study, although providing evidence about the usefulness of the collaborative filtering methods, are not necessarily generalizable to other clinical settings. The patients in a long-term care setting are relatively homogenous in terms of their demographics, drug regimens, and their diagnoses. In other care settings, such as an ambulatory clinic or a general hospital, the patient population would be significantly more diverse. In such situations, we expect that the heterogeneity among patients may make the reconciliation task more difficult but also will allow more complex prediction methods to improve upon the results of the popular algorithm. Finally, it should be noted that the data sets are small, and further work should examine how the methods scale to significantly larger data sets. Nevertheless, the methodology and general approach proposed in this study are applicable in many other care-delivery environments.

IMPLICATIONS AND FUTURE WORK

Decision theoretic approach for list ordering

Applying collaborative filtering methods to the observed partial lists produces a list of all drugs $j \notin \tilde{I}_i$, and their respective probabilities of being contained in $I_i - j \notin \tilde{I}_i$. However, just using probabilities to order drugs may ignore information regarding the clinical value that drugs provide to patients, their risks, and other potential costs and benefits. There are three potential ways to order this list of entities, depending on the decision-making task. The first way is to order the list by each drug’s estimated probability of membership in the true list. This approach places primacy on ensuring that all drugs in $I_i - \tilde{I}_i$, irrespective of potential consequence, are predicted correctly. The second approach is to order the list based on the probability weighted by the unconditional expected consequence $r_j = p_j \times E[c_j]$. This approach places emphasis on ensuring that drugs with the highest potential for future harm

are included in the list and could potentially emphasize drugs with high potential benefits. For instance, common drugs such as Accutane, Ovide, and Lariam are commonly prescribed drugs that may pose significant risks, and physicians should know about them when they are making their prescribing decisions.

We anticipate that the final approach for sorting and using the results of the collaborative filtering methods occurs within the context of a specific prescribing decision. In this decision task, the physician is about to prescribe a specific drug p . According to an electronic prescribing system’s drug-interaction database, drug p negatively interacts with the set of drugs in $j \in R_p$, none of which are contained in I_i^- . Thus, we can construct a list of drugs sorted on a measure that is a function of estimated probability p_j weighted by conditional consequence $E[c_j|p]$ as:

$$r_j = p_j \times E[c_j|p] \quad \forall j \in R_p$$

All three tasks are important to keep in mind. However, because the focus of this paper is on accurately estimating the probabilities and scores, we evaluate the collaborative filtering approaches using the probability measure alone.

Decision aids

Currently, the verification step of medication reconciliation relies primarily on information stored in a patient’s medication records, and on patient self-reports. Research suggests that self-reports are flawed, and patients regularly do not report all the drugs they are taking and may often provide incorrect information. However, research also suggests that decision aids can help improve patients’ recall of drugs. The result of our framework, particularly the ordered list of the top k drugs (by probabilities and/or consequence-weighted probabilities) can be used to enhance the verification process. In addition to providing a patient with a blank form where they can enter their

Table 10 Comparison of algorithms used for detection of omissions in medication lists

Algorithm	Running time	Pros	Cons
Popular	T: 0.00 P: 0.046	Fast and uses only basic data about the base rate of drugs being prescribed in clinics	Does not use any information about the patient for whom the prediction is being made
Co-occurrence	T: 0.096 P: 0.029	Fast and uses only basic information about the co-occurrence of pairs of drugs. Uses information about patient for whom the prediction is being made.	Assumes independence of all pairwise co-occurrences
K-Nearest Neighbor	T: 0.00 P: 0.034	Does not require potentially computationally intensive model estimation as does logistic regression. Able to incorporate information about patient data.	Computationally intensive since approach is not model-based. It requires a search of nearest neighbors at every prediction event.
Logit—PCA	T: 11.8 P: 0.050	Model based with clear theoretical interpretations about probability that a given drug is omitted. Is able to incorporate significant amount of patient information.	Computationally intensive for training and may require a significant amount of data before feasible for prediction purposes
Logit—Penalized	T: 241.7 P: 0.11	Model based with clear theoretical interpretations about probability that a given drug is omitted. Is able to incorporate a significant amount of patient information.	Very computationally intensive for training and may require a significant amount of data before feasible for prediction purposes

P, prediction time (s); T, training time (s).

medications, the clinician could also print out a list of these top k drugs and give them to the patients as a decision aid that may help them recall drugs that are likely to be not only missing from their record but also of significant consequence.

As subsequent steps, we hope to test this methodology on data from a diverse set of care delivery settings. Furthermore, the methodology currently uses only data-driven methods, and in two cases (Models 4b and 4d) only nominally Bayesian estimation of logistic regression (with uninformative priors) is used. In the future, we hope to be able to take into account the vast amount of medical evidence and information about clinical protocols within a clinical setting to set informative priors. For instance, order sets are increasingly becoming common in hospitals and other clinical settings. Order sets are a set of commonly prescribed medications, laboratory tests, and procedures which are prescribed to a patient for a condition or a combination of conditions. Knowing which drugs are generally prescribed together can help inform the assignment of priors specifying the relationship between two or more drugs. These priors can subsequently be updated based on available data from the clinical setting. In our current work, we only considered information from within a clinic for predicting missing drugs. In future work, we plan to incorporate more global information from other clinics to inform not only base rates for drugs, but also relationships between drugs which might exist and are globally visible but difficult to observe in individual clinics. Finally, other information such as a patient's allergies, the prescribing preferences of their physicians, and other relevant information from their medical history may also help to predict missing drugs.

Our current experiments suggest that the collaborative filtering approach to medication reconciliation holds promise. We anticipate improvements as additional information is used and as more and diverse clinical settings are evaluated using this methodology. We do not envision a one-size-fits-all solution to this problem, since clinical settings are heterogeneous, and different collaborative filtering approaches may work better in different scenarios. We also hypothesize that a collaborative filtering approach may be beneficial in dealing with other types of discrepancies in medical data, such as laboratory tests, diagnoses, and allergies.

Acknowledgments We thank C. Harle and our three anonymous referees for their very helpful and detailed comments.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Aspden P**, Wolcott J, Bootman L, et al. *Preventing Medication Errors. Quality Chasm Series*. Washington, DC: National Academies Press, 2006:151–220.
2. **Barnsteiner JH**. Medication reconciliation: transfer of medication information across settings—keeping it free from error. *Am J Nurs* 2005;**105**:31–6.
3. **Varkey P**, Cunningham J, Bisping DS. Improving medication reconciliation in the outpatient setting. *Jt Comm J Qual Patient Safe* 2007;**33**:286–92.
4. **Yang JC**, Tomlinson G, Naglie G. Medication lists for elderly patients. *J General Intern Med* 2001;**16**:112–15.
5. **Glintborg B**, Andersen SE, Dalhoff K. Insufficient communication about medication use at the interface between hospital and primary care. *Qual Saf Health Care* 2007;**16**:34–9.
6. **Glintborg B**, Hilleström PR, Olsen LH, et al. Are patients reliable when self-reporting medication use? Validation of structured drug interviews and home visits by drug analysis and prescription data in acutely hospitalized patients. *J Clin Pharmacol* 2007;**47**:1440–9.
7. **Kaboli PJ**, McClimmon BJ, Hoth AB, et al. Assessing the accuracy of computerized medication histories. *Am J Manag Care* 2004;**10**:872–87.
8. **Lindberg M**, Lindberg P, Wikström B. Medication discrepancy: a concordance problem between dialysis patients and caregivers. *Scand J Urol Nephrol* 2007;**41**:546–52.
9. **Seaton TL**, Gergen SS, Reichley RM, et al. Prescription claims in patients hospitalized with heart failure. *AMIA Annu Symp Proc* 2005:1109.
10. **Cornish PL**, Knowles SR, Marchesano R, et al. Unintended medication discrepancies at the time of hospital admission. *Arch Intern Med* 2005;**165**:424–9.
11. **Midlöv P**, Holmdahl L, Eriksson T, et al. Medication report reduces number of medication errors when elderly patients are discharged from hospital. *Pharm World Sci* 2008;**30**:92–8.
12. **Santell JP**. Reconciliation failures lead to medication errors. *Jt Comm J Qual Patient Safe* 2006;**32**:225–9.
13. **Nicol N**. Case study: an interdisciplinary approach to medication error reduction. *Am J Health Syst Pharm* 2007;**64**:17–20.
14. **Crichton E**, Smith D, Demanuele F. Patient recall of medication information. *Ann Pharmacother* 1978;**12**:591–9.
15. **Agrawal A**, Wu W, Khachewatsky I. Evaluation of an electronic medication reconciliation system in inpatient setting in an acute care hospital. *Stud Health Technol Inform* 2007;**129**:1027–31.
16. **Bedell SE**, Jabbour S, Goldberg R, et al. Discrepancies in the use of medications: their extent and predictors in an outpatient practice. *Arch Intern Med* 2000;**160**:2129–34.
17. **West SL**, Savitz DA, Koch G, et al. Recall accuracy for prescription medications: self-report compared with database information. *Am J Epidemiol* 1995;**142**:1103–12.
18. **Breese J**, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)* 1998:43–52.
19. **Mitchell AA**, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986;**123**:670–6.
20. **Hasan S**, Duncan G, Neill D, et al. Towards a collaborative filtering approach to medication reconciliation. *AMIA Annu Symp Proc* 2008:288–92.
21. **Park MY**, Hastie T. L1-regularization path algorithm for generalized linear models. *J R Stat Soc B (Stat Methodol)* 2007;**69**:659–77.
22. **Goldenberg A**, Kubica J, Komarek P, et al. A comparison of statistical and machine learning algorithms on the task of link completion. *Proceedings of the ACM SIGKDD Workshop on Link Analysis for Detecting Complex Behavior*, 2003.
23. **Hastie T**, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2001: 459–75.
24. **Christensen R**. *Log-Linear Models and Logistic Regression*. New York: Springer, 1997:116–67.
25. **Gelman A**, Jakulin A, Pittau MG, et al. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2008;**2**:1360–83.
26. **Wasserman L**. *All of Statistics: A Concise Course in Statistical Inference*. Berlin: Springer, 2004:362–8.
27. **Demsar J**. Statistical comparisons of classifiers over multiple data sets. *J Machine Learn Res* 2006;**7**:1–30.
28. **Friedman M**. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc* 1937;**32**:675–701.
29. **Graczyk M**, Lasota T, Telec Z, et al. Nonparametric statistical analysis of machine learning algorithms for regression problems. In: Setchi R, Jordanov I, Howlett R, et al, eds. *Knowledge-Based and Intelligent Information and Engineering Systems*. Berlin: Springer, 2010:111–20.