AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Brief Communication

Developing and maintaining clinical decision support using clinical knowledge and machine learning: the case of order sets

Yiye Zhang,[1] Richard Trepp,[2] Weiguang Wang,[3] Jorge Luna,[4,5] David K. Vawdrey,[4,6] and Victoria Tiase[4,7]

[1]Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, Cornell University, New York, NY, USA, [2]Department of Emergency Medicine, Columbia University Irving Medical Center, Columbia University, New York, NY, USA, [3]Decision, Operations and Information Technologies Department, Robert H. Smith School of Business, University of Maryland, College Park, Maryland, USA, [4]Value Institute NewYork-Presbyterian Hospital, New York, NY, USA, [5]Department of Epidemiology, Columbia University Irving Medical Center, Columbia University, New York, NY, USA, [6]Department of Biomedical Informatics, Columbia University Irving Medical Center, Columbia University, New York, NY, USA, and [7]Department of Information Services, NewYork-Presbyterian Hospital, New York, NY, USA

Corresponding Author: Yiye Zhang, PhD, MS, Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, Cornell University, New York, NY, USA (yiz2014@med.cornell.edu)

Received 15 March 2018; Revised 27 June 2018; Editorial Decision 1 July 2018; Accepted 6 July 2018

ABSTRACT

Development and maintenance of order sets is a knowledge-intensive task for off-the-shelf machine-learning algorithms alone. We hypothesize that integrating clinical knowledge with machine learning can facilitate effective development and maintenance of order sets while promoting best practices in ordering. To this end, we simulated the revision of an "AM Lab Order Set" under 6 revision approaches. Revisions included changes in the order set content or default settings through 1) population statistics, 2) individualized prediction using machine learning, and 3) clinical knowledge. Revision criteria were determined using electronic health record (EHR) data from 2014 to 2015. Each revision's clinical appropriateness, workload from using the order set, and generalizability across time were evaluated using EHR data from 2016 and 2017. Our results suggest a potential order set revision approach that jointly leverages clinical knowledge and machine learning to improve usability while updating contents based on latest clinical knowledge and best practices.

Key words: order set, clinical decision support, computerized provider order entry, mouse clicks

INTRODUCTION

Order sets are a component of computerized provider order entry (CPOE) that present appropriately grouped medical orders to increase the efficiency of ordering.[1] As a clinical decision support (CDS) for order placement, more importantly, order sets serve to reduce variations in the ordering process and encourage compliance with best practices.[2] Orders in an order set are included based on the intended functions and ordering times of their content items.[1] At our institution, each order set is recommended to be reviewed every

2 years. When there is a request to update an order set, a cross-campus group of clinicians including residents, attending physicians, nurses, and Laboratory Medicine staff discuss to reach a consensus request. The discussions are informed by historical usage statistics while also considering resource utilization concerns. The order set is then reviewed by the Order Set Working Group, an interdisciplinary group including staff from Information Technology (IT), Quality and Patient Safety, Nursing, Pharmacy, and Laboratory Medicine for approval. While knowledge content is presumably

satisfactory, this process can be resource intensive and difficult to scale as also reported at other institutions.[2–5] Therefore, an order set is a domain that may be potentially benefited from using statistics and machine learning to assist with scalable and efficient development and maintenance.

Population statistics such as most commonly used orders are useful information to consider in creating and maintaining order sets. However, historical data may contain biased and suboptimal practices such as over-utilization of laboratory orders that should be discarded even when they are commonly observed.[6,7] In addition, population statistics do not apply to all individual cases. To promote personalized CDS for order placement, applications of machine learning for the order set development and maintenance process have been studied in recent years to provide scalability and automation to this task that is commonly knowledge and resource intensive.[2,8–11] For example, an analogy can be made between order placement within CPOE and online retail, for which machine learning algorithms have been used to recommend merchandise to customers based on their historic activities and population statistics.[2,11] In the context of order sets, clinician users of order sets are the customers, and the orders that the clinicians choose for their patients are the merchandises. By similar principle, previous studies have demonstrated a potential for deriving order sets using data extracted from electronic health records (EHRs).[2,8,11,12] However, machine-learning applications on order placement must address a unique challenge, which is the alignment between finding orders that a user wants to place, and orders that are optimal for patient care practices and resource utilization. In retail recommendations, the former alone achieves the goal. However, in order placement, we must ensure that the machine-derived order sets achieve both goals. Otherwise, we will be blindly deriving order sets from historic ordering data that promote potentially common but suboptimal practices.[13] While machine learning has brought much promise for scalable and personalized development of CDS including order sets, how to better integrate it with clinical knowledge remains to be explored further.[14]

OBJECTIVE

In this study, we aimed to investigate common and new methods to revise order sets to identify a best approach. Compared to previous literature, the novelty of this study is 2-fold. First, we present an approach that adds machine learning to the latest clinical knowledge by dynamically determining default settings within a clinician-developed order set at each order placement based on individual patients' previous orders' results. Second, we show through 6 simulated revision approaches that our approach may be the most effective to reduce user workload while promoting best practices. We applied this study to an "AM Lab Order Set," one of the most commonly used order sets across departments in our institution. The 6 approaches include adjusting default settings and order contents by clinical knowledge, population statistics, machine learning, and their combinations. We evaluated the revised AM Lab Order Set's clinical validity, user workload measured by mouse clicks, and generalizability across time. We hypothesize that appropriately integrating clinical knowledge with machine learning on individual statistics may allow us to scale the order set development and maintenance process while promoting best and personalized care practices.

MATERIALS

The AM Lab Order Set is used to order laboratory tests to be collected the next morning and thus include commonly used relevant laboratory orders. Order-related information from September 2014 to October 2015 was extracted from an EHR system (Allscripts Healthcare Solutions, Inc.). The data used to revise the order set included 998 946 order placements for 37 924 patients who had at least 1 order placed via the AM Lab Order Set from 2014 to 2915. In addition, the data also included 3561 a la carte orders placed within 10-minute proximity to an AM Lab Order Set order to capture a la carte orders that could potentially be added to a revised order set.

Relevant variables include de-identified visit ID, order name, order set name (or a la carte), time stamp of order placement, and each order's result categorized by an abnormality code representing values being too high, high, normal, low, or too low. The AM Lab Order Set contained 12 order items shown in Table 1, all unselected by default (ie, default-OFF) in the CPOE user interface. To evaluate the revised order set's generalizability across time under each simulated approach, we also computed the mouse clicks for using them from 1-month ordering data from May 2016 and April 2017, which contain data for 3373 patients and 3237 patients, respectively. The evaluation data contained orders from the AM Lab Order Set, and 1650 and 1630 a la carte orders placed within 10-minute proximity to the order set orders from May 2016 and April 2017. We randomly selected the month to be studied in 2016 and 2017.

METHODS

Following a similar approach as Zhang et al,[2] We simulated the workload of a user by the number of mouse clicks associated with placing orders based on our current EHR interface. For this study, to compute the number of mouse clicks from data, we estimated that 1 click is required to select a default-OFF item or to de-select a default-ON item within an order set. We estimate that opening and order set take 3 mouse clicks. In addition, we estimated that it requires 3 clicks to place an a la carte order to incorporate additional clicks for search, although this is a conservative estimate.

Below we describe the simulated revisions of the AM Lab Order Set, which we note as M0 to M6. Table 2 summarizes the characteristics of each revision. M0 is the AM Lab Order Set in its original form. In M1, we changed the default setting of the order items based on population usage, as shown in Table 1. We switched "basic metabolic panel," "CBC with differential," and "magnesium" to ON, as these 3 orders were used for over 80% of patients who have used the order set at least once.[2] M1 is a revision entirely based on the team of clinicians, including members of the Order Set Working Group, as noted earlier. M3 adds or removes diagnostic orders based on population statistics from 1-year ordering data. Under this model, we added "glucose whole blood meter POC" and "calcium serum," both of which are diagnostic and commonly placed with AM Lab orders. We removed no orders from the original order set, as we did not find orders that were barely used to justify removal. M5 adds 13 orders for 8 kinds of laboratory measurements at through or peak levels if applicable, including "amikacin level," "cyclosporine level," "fibrinogen," "gentamicin level," "lactate dehydrogenase," "tacrolimus level," "tobramycin level," "uric acid," and "vancomycin level." The selection was primarily based on the team's clinical knowledge, practical experience, and recommendations from the Choosing Wisely initiative to promote best practices in Laboratory Medicine.[13]

M2, M4, and M6 all used individual statistics rather than population statistics. Specifically, we applied a logistic regression with order default status as a binary dependent variable, such that default

Table 1. Order items contained in AM Lab Order Set and their population usage statistics. % of unique patients: % of patients who used each order among those who have used AM Lab Order Set at least once during the study period; # Opened: number of times that the order set was opened.

Category	Order	% of unique patients		
		2014-15 (# Opened = 161 233)	May 2016 (# Opened = 12 786)	Apr 2017 (# Opened = 11 270)
Labs	Complete blood count (CBC)	40.3%	33.9%	30.6%
	CBC with differential	86.0%	87.9%	85.9%
	Liver function panel	48.4%	49.6%	52.1%
	Basic metabolic panel	98.7%	98.6%	98.3%
	Phosphorus	63.4%	58.7%	56.2%
	Troponin I	10.5%	10.0%	10.5%
	Type and screen	34.2%	35.3%	37.1%
Coag panel	PT/INR	58.1%	55.7%	55.0%
	Activated partial thromboplastin time (APTT)	54.7%	52.7%	51.9%
Cardiology adult	Magnesium	95.1%	93.1%	91.5%
	ECG 12 lead.	15.1%	15.4%	15.0%
Radiology	Port chest 1 view	8.1%	9.6%	6.2%

Table 2. Description of the simulated revisions. Dynamic: default settings are determined based on individual prediction using logistic regression.

Simulation model	Number of orders	Number of default-ON orders	Revision criteria			
			Default		Order	
			Population statistics	Machine learning	Population statistics	Clinical knowledge
M0	12		Original (no revision)			
M1	12	3	X	–	–	–
M2	12	Dynamic	–	X	–	–
M3	14	0	–	–	X	–
M4	14	Dynamic	–	X	X	–
M5	25	0	–	–	–	X
M6	25	Dynamic	–	X	–	X

settings for each order within the order set would be determined at the point of order entry. Predictive variables included age, gender, orders from the AM Lab Order Set in each model, and results from orders placed. We included orders and results from 2 previous order placements, such that the default settings can be informed by patients' earlier orders and results. The model was trained using a randomly selected 80% of the data from 2014 to 2015. We tested the model by the rest of the 20% of 2014 to 2015 data, May 2016 data, and spring 2017 data. Training and testing were done using Python scikit-learn package.[15] For the logistic regression model, the tolerance for stopping criteria was set to 0.0001. We used L1 regularization with the regularization strength as 1.0. Library LIBLINEAR was adopted for the optimization.[16] M2 and M4 apply the model to the original AM Lab Order Set and a revised AM Lab Order Set in M1, respectively. M6 applies the model to the AM Lab Order Set revised with clinical knowledge in M5, thus combining clinical knowledge with the individual-based machine-learning model.

RESULTS

Figures 1–3 show the comparisons between original and revised order sets in terms of the total user workload per patient for AM Lab
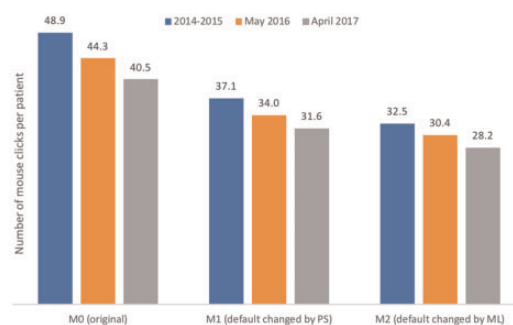

Figure 1. Average number of mouse clicks per patient for all order set orders and a la carte orders within 10-minute proximity in M0, M1, and M2. Horizontal axis indicates model. PS: population statistics; ML: machine learning.

Order Set orders and a la carte orders placed within 10-minute proximity. Figure 1 shows the comparison across M0, M1, and M2. In M1, "basic metabolic panel," "CBC with differential," and "magnesium" were defaulted ON for all patients in the order set based on population statistics. This revision is an example of how data-driven approaches may reinforce suboptimal practices such as the over-ordering of laboratory orders in this case. On the other
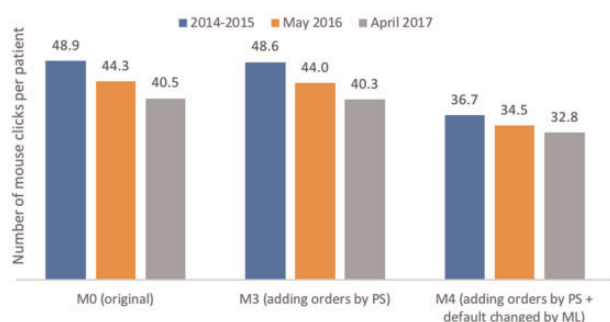
Figure 2. Average number of mouse clicks per patient for all order set orders and a la carte orders within 10-minute proximity in M1, M3, and M4. Horizontal axis indicates model. PS: population statistics; ML: machine learning.
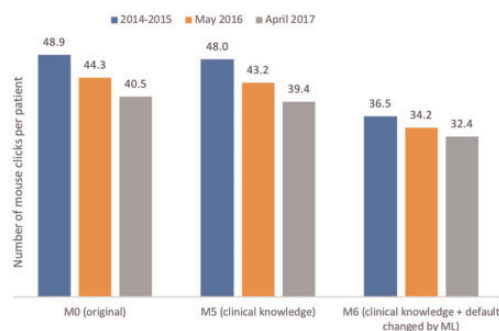

Figure 3. Average number of mouse clicks per patient for all order set orders and a la carte orders within 10-minute proximity in M0, M5, and M6. Horizontal axis indicates model. PS: population statistics; ML: machine learning.

hand, in M2, the defaults were determined based on patients' previous orders and their results, and the user workload is 33.5% and 12.4% less than M0 and M1, respectively. In Figure 2, we see that while the 2 diagnostic orders added in M3 barely changed the number of mouse clicks, further applying machine learning to M3 resulted in 24.9% reduction in the number of mouse clicks. Lastly, Figure 3 shows that M6, where clinical knowledge and machine learning are used to revise, achieved 25.4% reductions in user workload while incorporating updated clinical knowledge and best practices. M5 revised the order set using clinical knowledge only and did not result in significant changes in the user workload. In the May 2016 and April 2017 test data, M6 generated, respectively, 488 and 381 individualized default settings according to previous orders and results, whose validity will be examined in a future study. The differences in mouse clicks per order placement between M0 and M6 are all significant with $P < .001$ using a 1-sided 2-sample paired t test at a significance level of 0.05 for the 2014 to 2015 data, May 2016 data, and April 2017 data, respectively. Detailed user workload per patient is reported in Supplementary Table 1 of Online Appendix.

DISCUSSION

The overarching goal of this study is to understand how we can facilitate the order set development and maintenance process towards a continuous improvement cycle.[3] Our results suggest that combining clinical knowledge with machine learning may be promising not only in significantly reducing user workload, but also might serve to balance usability and best practices. M6 can be generalizable beyond morning laboratory tests and to other order sets by re-training the

machine-learning model with a different patient population and their data. For instance, for a sepsis order set, we can apply our method to the patient cohort who used a sepsis order set to derive a sepsis-order-specific default suggestion model. In our future work, we may explore other models such as recurrent neural networks to aim for higher accuracies in determining default settings. Furthermore, while order sets traditionally are considered to be institutional or departmental CDS, the success of this potential approach may push order sets to be an individualized CDS, although an implementation of individualized default determination may be constrained by the cognitive workload of acknowledging new default settings and the availability of real-time data at every order placement.

The metric used in this study to evaluate the order set focused only on mouse clicks. However, we realize that cognitive factors such as the amount of time used for each order placed should also be considered when appropriate measuring devices or experimental settings are available. We are underway to conduct a survey with order set users to obtain cognitive workload coefficients that allow us to better report the cognitive workload. In addition, future experiments will measure the outcomes associated with these different approaches such as variability in order placements. In this study, the clinical knowledge was obtained by following the discussion of a team of clinical and institution staff who presented their knowledge and understanding of the best practices. Hence, the proposed approaches in this study are still subject to the scalability challenge. Natural language processing of the clinical practice guideline may be explored in future studies to mitigate the challenge.

CONCLUSION

We demonstrate that applying machine learning to dynamically determine the default settings of an order set developed based on clinical knowledge may promote scalable and personalized development and maintenance of order sets.

FUNDING

CONTRIBUTORS

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

REFERENCES

1. Payne TH, Hoey PJ, Nichol P, Lovis C. Preparation and use of preconstructed orders, order sets, and order menus in a computerized provider

order entry system. *J Am Med Inform Assoc* 2003; 10 (4): 322–9. [published Online First: Epub Date].

2. Zhang Y, Padman R, Levin JE. Paving the COWpath: data-driven design of pediatric order sets. *J Am Med Inform Assoc* 2014; 21 (e2): e304–11. [published Online First: Epub Date].

3. McClay JC, Campbell JR, Parker C, Hrabak K, Tu SW, Abarbanel R. Structuring order sets for interoperable distribution. *AMIA Ann Symp Proc* 2006: 549–53.

4. McGreevey JD 3rd. Order sets in electronic health records: principles of good practice. *Chest* 2013; 143 (1): 228–35. [published Online First: Epub Date].

5. Hulse NC, Lee J, Borgeson T. Visualization of order set creation and usage patterns in early implementation phases of an electronic health record. *AMIA Annu Symp Proc* 2016; 2016: 657–66.

6. Melendez-Rosado J, Thompson KM, Cowdell JC, *et al.* Reducing unnecessary testing: an intervention to improve resident ordering practices. *Postgrad Med J* 2017; 93 (1102): 476–9. [published Online First: Epub Date].

7. Rosenbloom ST, Chiu KW, Byrne DW, Talbert DA, Neilson EG, Miller RA. Interventions to regulate ordering of serum magnesium levels: report of an unintended consequence of decision support. *J Am Med Inform Assoc* 2005; 12 (5): 546–53. [published Online First: Epub Date].

8. Chen JH, Goldstein MK, Asch SM, Mackey L, Altman RB. Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets. *J Am Med Inform Assoc* 2016; doi: 10.1093/jamia/ocw136[published Online First: Epub Date]|.

9. Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. *J Biomed Inform* 2015; 53: 73–80. [published Online First: Epub Date].

10. Woods AD, Mulherin DP, Flynn AJ, Stevenson JG, Zimmerman CR, Chaffee BW. Clinical decision support for atypical orders: detection and warning of atypical medication orders submitted to a computerized provider order entry system. *J Am Med Inform Assoc* 2014; 21 (3): 569–73. [published Online First: Epub Date].

11. Chen JH, Podchiyska T, Altman RB. OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records. *J Am Med Inform Assoc* 2016; 23: 339–48.

12. Wright A, Sittig DF. Automated development of order sets and corollary orders by data mining in an ambulatory computerized physician order entry system. *AMIA Ann Symp Proc* 2006; 2006: 819–23.

13. Goldbach P. The choosing wisely campaign. *Health Affairs* 2018; 37 (2): 335. [published Online First: Epub Date].

14. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med* 2017; 376 (26): 2507–9. [published Online First: Epub Date].

15. Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–30.

16. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: a library for large linear classification. *J Mach Learn Res* 2008; 9: 1871–4.