Paving the COWpath: data-driven design of pediatric order sets

Yiye Zhang,1 Rema Padman,2 James E Levin3

ABSTRACT

Objective Evidence indicates that users incur significant physical and cognitive costs in the use of order sets, a core feature of computerized provider order entry systems. This paper develops data-driven approaches for automating the construction of order sets that match closely with user preferences and workflow while minimizing physical and cognitive workload.

Materials and methods We developed and tested optimization-based models embedded with clustering techniques using physical and cognitive click cost criteria. By judiciously learning from users' actual actions, our methods identify items for constituting order sets that are relevant according to historical ordering data and grouped on the basis of order similarity and ordering time. We evaluated performance of the methods using 47,099 orders from the year 2011 for asthma, appendectomy and pneumonia management in a pediatric inpatient setting.

Results In comparison with existing order sets, those developed using the new approach significantly reduce the physical and cognitive workload associated with usage by 14–52%. This approach is also capable of accommodating variations in clinical conditions that affect order set usage and development.

Discussion There is a critical need to investigate the cognitive complexity imposed on users by complex clinical information systems, and to design their features according to ‘human factors’ best practices. Optimizing order set generation using cognitive cost criteria introduces a new approach that can potentially improve ordering efficiency, reduce unintended variations in order placement, and enhance patient safety.

Conclusions We demonstrate that data-driven methods offer a promising approach for designing order sets that are generalizable, data-driven, condition-based, and up to date with current best practices.

BACKGROUND AND SIGNIFICANCE

As technology assumes an increasingly critical role in healthcare delivery, recent literature indicates that the interaction of technology and humans in healthcare settings warrants a thorough investigation.1 Poor usability of health information technology (IT) causes new types of medical errors that are unique to the technology era, particularly through insufficient customization of the systems to workflow in the care delivery setting, end user information overload, and lack of adequate knowledge about user behaviors.1–4 In this paper, we examine an unintended, adverse consequence of implementing computerized physician order entry (CPOE) systems—the excessive physical and cognitive workload imposed on users.5 This is a significant contributor to poor judgment, inaccuracies, and erroneous actions that may lead to adverse outcomes in the clinical care setting.6 Poor usability stems from the lack of coordination between technology and human practices.6–9 Previous research indicates that knowledge of user behaviors can have a profound impact on achieving health IT use, and incorporating this knowledge into the design of health IT systems is vital for enhancing their usability.1

To address this challenge, we introduce a ‘paving the COWpath’ approach into the design of health IT system components,10–11 where ‘COWpath’ indicates the ordering ‘paths’ taken by busy providers, typically using a ‘computer on wheels’.12 This approach is a design rationale where the system components function by learning from and embracing the best practices performed repeatedly by users, such that they are able to discover users’ ‘desire paths’ and facilitate actual workflow without requiring excessive human intervention.13 We suggest this as a mechanism to combine explicit knowledge14—available from evidence-based guidelines—with the tacit, non-codified knowledge embedded in the user experiences of domain experts resulting from the interactions between processes, culture, and routines.15–17 As medication ordering and management is a critical patient-safety-sensitive electronic health record function,1 we focus on the development of ‘order sets’, a CPOE component that has the potential to make order entry faster, easier, and safer.6

Order sets have the potential to increase the efficiency of order entry.6 There are two ways users can place orders via CPOE. One is a la carte order placement, where users search for an order via a keyword and place a single order found from the search results. When multiple orders are needed, users repeat this process as many times as the number of orders needed. However, by using order sets, users can enter multiple system-suggested orders with as few as one mouse click. Instead of searching by order name, users can search for an order set by disease condition, and be presented with a set of expert-suggested/evidence-based orders, all related to the condition of interest. Hence, order sets also serve as a decision support tool to simplify complex order placement and reduce prescribing variation and errors.18 Traditionally developed by experts in each clinical area, order sets include multiple orders related by clinical purposes that less-experienced/time-constrained users may refer to. Since a majority of potential adverse drug events are the result of errors during prescribing,19 well-designed order sets can increase compliance on order placement based on best practices and improve patient safety by lowering the chances of prescribing variations and errors.

Correspondence to Yiye Zhang, School of Information Systems Management, H John Heinz III College, Carnegie Mellon University, 10 Forbes Ave, Pittsburgh, PA 15213, USA; yiye@andrew.cmu.edu

Received 3 September 2013
Revised 5 February 2014
Accepted 24 February 2014
Published Online First
27 March 2014
Today, the creation of order sets has been widely accepted as a core prerequisite for successful CPOE implementation.\textsuperscript{6} \textsuperscript{20–23}

Despite these benefits, historical data indicate a low rate of usage of order sets, primarily due to lack of order set content to accommodate diverse patient conditions, physicians’ personal preferences and unfamiliarity with order sets, and, most importantly, inconsistency of order set content with current best practices.\textsuperscript{24–27} Instead, we observe heavy use of a la carte orders to compensate for the limitations of order sets, generating a COWpath that provides evidence of evolving care delivery needs and adapted to users’ workflows, but may be highly inefficient. Indeed, owing to the time-, manual labor-, and knowledge-intensive order set development process, the growth, modification, and maintenance of order sets are often not kept up to date with changing provider skills and interests, quality and clinical resource management initiatives, new types of services, regulatory requirements, and an evolving standard of care.\textsuperscript{24–27} Consequently, there may be a significant mismatch between CPOE component capability and provider needs that harms treatment quality and exposes patients to potential medical errors.

Order set usage has been examined in a number of previous studies to identify usability issues,\textsuperscript{28–31} and ambulatory and laboratory corollary order sets have been created using data-mining techniques.\textsuperscript{30,32} For example, Hulse et al\textsuperscript{33} created an automated, knowledge-based order set feedback system that analyzes order set usage relative to the original template and suggests modifications based on actual usage. In a study by Avansino and Leu,\textsuperscript{34} systematically designed order sets were shown to provide a reduction in cognitive workload, but there was no improvement in terms of mouse clicks or time spent on order sets. In this paper, we demonstrate that potential reductions in both physical and cognitive workload are possible through order set optimization. We summarize our research to automate the entire order set design process using machine learning and optimization techniques while incorporating physical and cognitive workload as major development criteria.\textsuperscript{24–27}

The machine learning methods enable us to learn the tacit knowledge embedded in the COWpath and combine them with the evidence built into the current order sets, while the optimization criteria allow us to accommodate human factors in the design process. We further evaluate our methods using CPOE data for inpatient asthma (a chronic medical condition), appendectomy (an acute surgical condition), and pneumonia (an acute medical condition). This approach has the potential to be more broadly generalizable to other conditions and care delivery settings.

OBJECTIVE
We describe the complexity of order set usage, analyze its effect on users and patient safety, and develop order sets that reduce physical and cognitive workload imposed on users. Specifically, we propose an order set development method by leveraging historical ordering data from the inpatient setting. As shown in figure 1, at an abstract level, using only a la carte orders at one extreme or a single, comprehensive order set that includes all possible orders at the other extreme will both incur very high physical and cognitive workload. Between these two, we hypothesize that there must be a ‘satisficing’ number of order sets, each containing an ‘acceptable’ number of order items, such that users can place orders safely with much lower physical and cognitive workload.\textsuperscript{15} Therefore, our methods model the order set development problem as an optimization process embedded with machine learning techniques to identify orders that are clinically and temporally relevant.

Physical and cognitive workloads are measured in terms of mouse clicks. We define mouse click cost (MCC) as the number of mouse clicks users apply to complete ordering tasks, and cognitive click cost (CCC) as MCC multiplied by weights indicating the complexity of specific ordering tasks.\textsuperscript{24,27} Using historical order data for inpatient management of asthma, appendectomy, and pneumonia, we demonstrate the method’s performance on different clinical conditions that may affect the automated order set development process differentially. Rather than merely extracting popularly used orders to develop order sets, we take into consideration each condition’s characteristics, as well as usability issues, to create order sets that are data-driven, condition-based, and up to date with the current best practices.

MATERIALS AND METHODS
Data
Data for this study were obtained from the Children’s Hospital of Pittsburgh (CHP) of the University of Pittsburgh Medical Center (UPMC). As the first pediatric hospital in the USA to reach Healthcare Information and Management Systems Society (HIMSS) Analytics Stage 7 electronic medical record implementation status, CHP has been all-digital with a single integrated eRecord (Cerner Millenium) since 2009. With over 13 000 inpatient visits in 2011, a patient at CHP was hospitalized, on average, for 5.5 days, during which time 36 unique individuals entered 846 order actions. Over 10 million order actions are stored in the clinical data warehouse, which is an exact copy of the production eRecord.\textsuperscript{23} There are 1559 departmental order sets in the eRecord system covering 17 978 orders, designed in 2002 by CHP clinicians, mainly for medical admission and surgery. The order set template has not been updated since its creation in 2002, and all analyses performed for the study were based on the same template. All patient-identifiable health information was removed to create a deidentified dataset for this study. This study was designated as exempt by the University of Pittsburgh Institutional Review Board.

Important variables extracted from the data warehouse include the order, order action time, care set flag, deidentified patient ID, diagnosis, severity, age at admission, gender, and length of stay in hours. For this study, order action time includes time from 20 h before admission to 24 h after admission, as
most order set usage is concentrated on the first day of admission. Data from 20 h before admission often come from patients’ emergency department visits immediately before they are admitted to the inpatient setting. Orders are from the same encounter as the admission, but the data do not exclude orders that are placed for other conditions when the patient has other comorbidities. Diagnosis and severity are captured by drug description and severity of illness using all patient refined diagnosis-related groups (APR-DRGs).36

For this study, we extracted the top 26 APR-DRGs, including 4331 cases that required a total of 490 049 orders that were placed in calendar year 2011. From this set, 292 asthma cases with 20 076 orders and 40 order sets, 105 appendectomy cases with 13 280 orders and 51 order sets, and 179 pneumonia cases with 13 743 orders and 69 order sets were selected for evaluating the methods developed in this study. Patient age is categorized into age groups following American Academy of Pediatrics (AAP) age groups: 1 =neonate (<30 days); 2=infancy (≥30 days and <1 year); 3=early childhood (≥1 and <5 years); 4=late childhood (≥5 and <13 years); 5=adolescence (≥13 and ≤18 years); 6=adult (≥18 years). A care set flag identifies order set items against a la carte items. A care set flag of 0 refers to an a la carte item, 1 an order set parent, 2 an order set, and 3 an order set nested in another order set. In this study, we include only orders that have a care set flag of 0, 2 and 3.

The size of order sets ranges from two to more than 50 unique items. Each item in order sets can be defaulted ‘ON’ or ‘OFF’ according to clinical relevance and frequency of use. An order item can be included in multiple order sets when necessary. Items with a checked checkbox are default ON items, and those with blank checkboxes are default OFF items. Figure 2 is a screen shot that shows part of the ‘Admission Orders Pneumonia’ order set. Unlike a a la carte order placement, where users need to apply a mouse click every time to select an individual order, default ON items are automatically selected when an order set is chosen. With additional clicks, users can add default OFF items to the selection or deselect default ON items from the order placement. Ordering through order sets is not mandatory, with a few exceptions.

Method

Click costs

As shown in figure 2, MCC counts the exact number of mouse clicks that users have to apply for each order entry. One unit of MCC is incurred when a user places an a la carte order, selects an order set, deselects a default ON order set item, or selects a defaulted OFF order set item. Furthermore, given that these tasks require different levels of cognitive workload, we multiply MCC by a different cognitive cost coefficient for each task to calculate CCC. Assuming that these coefficients are independent of the patient, order set, time interval, and physician (this assumption can be easily relaxed in the models but incur significant information gathering overhead), we applied two approaches to derive the coefficients for CCC to represent the cognitive workload that accompanies each mouse click in order placement. One was provided by a clinical and medical informatics expert, and the other was estimated from a survey of 15 users including physicians and nurses. Cognitive-click-cost-expert-estimate (CCCE) coefficients are made assuming that users typically trust the order set contents. Hence CCCE attributes high cost coefficients to actions that disagree with the default setting, such as deselecting ON items or selecting OFF items. Conversely, cognitive-click-cost-survey-estimate (CCCS) coefficients reveal that, while deselecting ON items and selecting OFF items requires high costs, almost the same amount of cognitive workload is required of users to ensure that default settings are suitable for patients, leading to small differences in the four coefficients. In this study, MCC and CCC focus on clicks that are related to ordering only and do not include other activities while ordering such as up or down clicks, and do not account for incomplete orders and rework. Table 1 summarizes the cost coefficients derived for MCC and CCC.

Coverage rate

We introduce a term called coverage rate to measure the goodness of order sets. Coverage rate of an order set c used by p=1, ..., N patients is defined as

\[
\frac{1}{N} \sum_{p=1}^{N} \frac{i_{p,c}}{i_c}
\]

where \(i_{p,c}\) is the number of unique items given to patient p from order set c, and \(i_c\) is the total number of unique items originally in order set c. Generally, orders from multiple order sets are given to a single patient during one order entry e. Given S order sets and N patients, the average coverage rate (ACR) during one order entry is

\[
\frac{1}{S} \sum_{c=1}^{S} \left( \frac{1}{M} \sum_{e=1}^{M} \left( \frac{1}{N} \sum_{p=1}^{N} \frac{i_{p,c}}{i_c} \right) \right)
\]

Then, the average overall coverage rate across S order sets, M order entries, and N patients is

\[
\frac{1}{S} \sum_{c=1}^{S} \left( \frac{1}{M} \sum_{e=1}^{M} \left( \frac{1}{N} \sum_{p=1}^{N} \frac{i_{p,c}}{i_c} \right) \right)
\]

where \(i_{p,c}\) is the number of unique items given to patient p from order set c during order entry e. High coverage rate for an order set suggests that users can use the order set with few further modifications on average.

Relative risk

We attempt to place order items that tend to be ordered together for a single patient during an order entry session into one order set. In order to find clinically relevant order items, we apply a similarity measure called relative risk (RR), commonly used in biomedical informatics to look for similarity between diseases.37 Given a pair of orders x and y, let \(U_{xy}\) be the number of unique patients who had both orders, \(V_x\) be the number of times order x was placed for patients, \(V_y\) be the number of times order y was placed for patients, and N be the total number of patients. Relative risk, \(R_{xy}\), for this order pair is defined as \(U_{xy}/V_x V_y\). Given i number of orders, each order has i–1 number of RR measures when paired with each of the rest of the orders. Bisecting K-means clustering is applied to the i by i matrix of RR to identify similar orders that should belong to the same order sets. Bisecting K-means clustering is known to perform clustering faster than K-means clustering.38

Iterative algorithm

Since order placement is driven by patients’ conditions, we develop order sets by extracting order items most suitable for constituting order sets according to different time intervals. This is modeled as a two-stage optimization problem with MCC and CCC as the objective function, respectively, while satisfying
constraints on time and order set content. Before the optimization process, default settings for orders are updated such that an item is defaulted ON if more than 80% of the study patients have used it at least once, and OFF if otherwise, thus judiciously learning from the COWpath. For each clinical condition, the algorithm iteratively extracts orders present in the data in increasing time intervals, starting with a 2 h interval. In each iteration, the bisecting K-means algorithm clusters items that often co-occur in patients’ order lists into the same order sets. Combinations of order sets with the lowest MCC or CCC are kept as the best solutions. Finally, the iteration ends when MCC and CCC are compared across time intervals of different lengths to select the optimal time interval. The end point of the previous time interval becomes the starting point for the next time interval (figure 3). Each item is part of no more than one order set within one time interval. However, an item can be included in multiple order sets across different time intervals. This process is repeated from 20 h before admission until 24 h after admission.

To computationally evaluate the methods and illustrate the optimization results, the data are divided into training and test sets. While training sets include patients who had only the particular APR_DRG diagnosis as their final diagnosis upon discharge, test sets evaluate the performance of the machine-derived order sets using data on patients who exhibited comorbidities so that the robustness of the order sets can be determined in a realistic manner. For example, the asthma training set includes patients who had a diagnosis of ‘asthma, unspecified’, while its testing set includes patients who had comorbidities such as ‘contact dermatitis and other eczema, unspecified cause’ at some point during their hospital stay. Similarly, the appendectomy training set includes a sample of patients with ‘acute appendicitis’ and ‘appendicitis, unqualified’, and the pneumonia training set includes patients with ‘pneumonia, organism unspecified.’ In addition, the severity of the condition was controlled to reduce the noise in the optimization process. We use the severity group with the largest sample size in each condition for reporting the computational results. Based on the assignment rule discussed above, we have 84 asthma patients for training and 208 for testing, 93 appendectomy patients for training and 57 for testing, and 96 pneumonia patients for training and 83 for testing.

Clustering and optimization were performed using an integrated computational platform that we developed (figure 4). Users can select clinical conditions and severity level and optimize order sets based on MCC, CCCE, and CCCS criteria, and time intervals of interest. Currently available clinical conditions are asthma, appendectomy, and pneumonia, although new conditions can be readily added. Resulting orders and their respective order set assignments, MCC and CCC values, and parameters used for each optimization session on the platform are stored in a secure MySQL database.

### RESULTS

**Descriptive statistics**

Table 2 shows order set usage associated with asthma, appendectomy, and pneumonia management. ‘Number of unique order sets used’ is the total number of unique order sets used across all the patients with each condition. On the other hand, ‘Number of times order set opened’ is the total number of times users chose order sets to place orders. ‘Number of order set orders’ is the number of orders placed from order sets, as opposed to a la carte. ‘Number of clicks to use order set’ is the number of mouse clicks that users applied to place order set orders. ‘Clicks saved by order sets’ is the difference between the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>MCC and CCC coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCC</td>
</tr>
<tr>
<td></td>
<td>Select</td>
</tr>
<tr>
<td>Order set/a la carte</td>
<td>1</td>
</tr>
<tr>
<td>Default ON</td>
<td>0</td>
</tr>
<tr>
<td>Default OFF</td>
<td>1</td>
</tr>
</tbody>
</table>

CCCE, cognitive-click-cost-expert-estimate; CCCS, cognitive-click-cost-survey-estimate; MCC, mouse click cost.
number of order set orders and the number of clicks to place order set orders. For example, the actual number of order set orders needed for pneumonia patients was 8935, but users had to make 9827 mouse clicks to complete order placement from 2662 order sets, adding 10% higher physical workload compared with a la carte order placement, where users simply needed to make 8935 mouse clicks. Yet, had the 2662 order sets been a perfect fit for patients, users would have completed the order placement process with just 2662 clicks, saving 70% of the physical workload. While it is highly unlikely that order sets will match every patient’s requirements perfectly, and hence will always require some sort of modification, this extreme example demonstrates the potentially significant impact that improved usability of order sets may have on treatment efficiency.

**Optimization results**

Table 3 shows a summary of optimization results under three objective functions: MCC, CCCE, and CCCS. Reduction in clicks per patient is the difference in the total number of clicks users have to apply when using current and machine-derived order sets per patient. The new approach potentially reduced both the physical and cognitive workload of order placement by 14–52% in the study sample. For example, pneumonia patients in the training set needed 73 orders on average, which can be placed with 67 clicks on average if machine-derived order sets were used, instead of 79 clicks on average if current order sets were used. The number of new time intervals is the number of optimal time intervals identified by the algorithm, so that the order set content reflects the clinical needs during the particular time interval. There is no time interval for the current order sets. In the redesigned solution, the number of optimal time intervals ranges from 15 to 20, most of which are 2 h intervals. This suggests that ordering patterns tend to change every 2 h.

To make the comparison of the number of order sets and their size easier, figure 5 shows the number of distinct order sets and the total number of items covered by these order set content computational platform. (A) step 1; (B) step 2; (C) step 3.

![Iterative process for order set timing and content](image)

**Figure 3**

Iterative process for order set timing and content. CCC, cognitive click cost; MCC, mouse click cost.

![Order set content computational platform](image)

**Figure 4**

Order set content computational platform. (A) step 1; (B) step 2; (C) step 3.

---

during the inpatient stay supports this explanation: 4.7 for patients with pneumonia compared to 2.4 for asthma and 2.1 for appendectomy. Clinical input combined with optimization may help achieve greater improvement, especially for complex conditions such as pneumonia.

Sample case
A sample case demonstrates the changes made to order sets based on the three objective function criteria. Orders administered for a patient during one ordering action is shown in Table 4 with the current and new order set assignment and default settings. Twelve items ordered for this patient currently come from four order sets (C1–C4). After optimization, these 12 items were reassigned to different order sets (M1–M4, E1–E6, S1–S4), or designated as a la carte items (A) based on MCC, CCCE, and CCCS, respectively. The status quo MCC is 15, CCCE is 20.3, and CCCS is 76. After optimization, MCC is 13 (15.4% lower), CCCE is 12.9 (36.4% lower), and CCCS is 23.1 (69.6% lower). Our algorithm also shortened the length of order sets and turned on OFF items that are often used, and vice versa. For example, we found that more than 80% of the study patients had used ‘initial pulse oximetry continuous’—an OFF item currently—at least once. Hence, our algorithm grouped ‘subsequent pulse oximetry continuous’ and ‘initial pulse oximetry continuous’ into one order set, and defaulted both items to be ON. Also, instead of having an order set such as C1, which contains 63 items currently, our algorithm broke it down into smaller order sets to ease the cognitive burden of reviewing a long list of orders, many of which may not be selected for most patients.

DISCUSSION
Other approaches to order set design
Order sets are intended to assist users in performing their order placement tasks efficiently and effectively. Current approaches rely on continuous and consistent manual updating to accommodate new knowledge, making it a challenge to adapt to changing workflows and patient requirements. We believe a combination of order set development and order set modification methods can better address the evolving care-delivery needs of diverse groups of patients in distinct care-delivery environments. In a prior, preliminary study, we investigated five models of order set usage: (1) status quo; (2) add a la carte orders to current order sets; (3) change default setting of current order sets; (4) combination of (2) and (3); (5) add order sets created from a la carte items using clustering to (4). Results indicated that model (5) had the lowest MCC and CCC, suggesting that combining experts’ consensus with evidence from data is a promising method for increasing usability while maintaining clinical validity. This method preserves experts’ consensus by learning ordering patterns from current order sets while incorporating evidence from actual workflows such that rarely used orders are deleted and frequently used a la carte orders are added to order sets. When scientific knowledge and practice-based evidence both increase at a rapid pace, a complete redesign of the order sets may be necessary at periodic intervals, which can utilize the COWpath approaches developed in this paper.

Limitations
Order sets produced by machine-derived techniques suffer from the disadvantage that they will not capture clinically meaningful

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Order set usage of the selected diagnosis groups in 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis group</td>
<td>Patients (n)</td>
</tr>
<tr>
<td>Asthma</td>
<td>415</td>
</tr>
<tr>
<td>Appendectomy</td>
<td>353</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>357</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Computational results for asthma, appendectomy, and pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis group</td>
<td>Reduction in clicks per patient (n (%))</td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
</tr>
<tr>
<td>MCC</td>
<td>9 (14)</td>
</tr>
<tr>
<td>CCCE</td>
<td>14 (18)</td>
</tr>
<tr>
<td>CCCS</td>
<td>72 (52)</td>
</tr>
<tr>
<td>Appendectomy</td>
<td></td>
</tr>
<tr>
<td>MCC</td>
<td>19 (19%)</td>
</tr>
<tr>
<td>CCCE</td>
<td>26 (21%)</td>
</tr>
<tr>
<td>CCCS</td>
<td>36 (20%)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td></td>
</tr>
<tr>
<td>MCC</td>
<td>12 (15%)</td>
</tr>
<tr>
<td>CCCE</td>
<td>18 (19%)</td>
</tr>
<tr>
<td>CCCS</td>
<td>30 (19%)</td>
</tr>
</tbody>
</table>

* p<0.05.
†84 patients for asthma; 93 patients for appendectomy; 96 patients for pneumonia.
‡208 patients for asthma; 57 patients for appendectomy; 83 patients for pneumonia.
ACR, average coverage rate; CCCE, cognitive-click-cost-expert-estimate; CCCS, cognitive-click-cost-survey-estimate; MCC, mouse click cost.
but rare events. Therefore, it is important to have experts evaluate order sets and, if necessary, embed known scientific evidence and guidelines into them. In addition, the detailed nature of treatment options and possibility of many other complications for each condition play an important role in the success of order set development. Large variations in order placement patterns associated with varied and multiple complications remains the biggest challenge, as illustrated by pneumonia orders. Changing certain parameters in the optimization method, applying different clustering methods according to condition type, or heuristic approaches for convergence may lead to even better results such as improved coverage rates. Moreover, the retrospective data used to build order sets in this paper are more or less influenced by the current order sets, since users are initially trained to place orders using the current order sets. Users may have learned to avoid some flaws in the current system, and incorporate workarounds that are not the best practices.

Provider variability is another factor that may affect the data. It is not uncommon for orders to be placed by interns or residents who may be biased because of lack of knowledge or experience, or both. Hence, the models and methods proposed in this study are a first step towards automating the redesign of order sets on a periodic basis. The results of our methods must be reviewed for clinical appropriateness based on evidence or by an institutional clinical effectiveness group. In addition, since our study uses only 1 year’s worth of data, from 2011, during which the order set template also remained unchanged, it is likely that biases due to provider variability may have limited impact in this study. A future study may explore variations in usage of order sets across provider groups and across years to classify changes as those based on evidence, poor design, or bias.

Future steps
Testing the machine-derived order sets by actual users in realistic settings is a vital next step, which can be conducted first as a pilot study with interested users and in a simulated environment. Clinical review of the order sets by experts needs to be conducted before actual clinical use of machine-derived order sets. Also, since machine-derived order sets learn from actual practices, regular content update and revision should be followed to keep order sets up to date with changes in the workflow and maintain clinical validity. Updates of the order set template can follow the cycle described in this paper, starting with relearning of the workflow using the COWpath approach, followed by expert evaluation.

CONCLUSION
This paper focuses on a critical, unintended consequence of IT in healthcare delivery, the excessive physical and cognitive workload on end users due to poor usability in the context of using order sets within CPOE. We propose a ‘paving the COWpath’ approach by modeling a two-stage optimization method embedded with a machine learning technique to develop data-driven order sets that used retrospective order data from the pediatric inpatient setting. An evaluation of the new approach provides evidence of a successful reduction in physical and cognitive workloads as well as an increase in order set coverage for diverse patient conditions. The results indicate that automated
development is time efficient and able to identify commonly co-located orders that are also supported by evidence. In this paper, the optimization focused only on reducing physical and cognitive workload, but it is possible to add other development- related criteria as well. Also, while our methods were built on the basis of CHP orders, the data-driven approach is generalizable for different diagnosis groups and workflows, with minor modifications. Hence, we believe that it can be a powerful method to use as a starting point for informed order set review and redesign and, in some instances, to replace the traditional manual development. The easy implementation of our methods can facilitate ongoing update of order set content by healthcare organizations and allow treatment quality to stay up to date with the evolving care delivery standards. We anticipate that data-driven order set design will facilitate order sets to be developed on the basis of current best practices, potentially leading to increased order set acceptance, ordering efficiency, and improved patient safety.

Correction notice This article has been corrected since it was published Online First.

Acknowledgements Dr James E Levin passed away on 11 February 2013. We are greatly indebted to his vision, contributions and support that made this study possible.

Contributors JEL initiated the study and obtained data from the Children’s Hospital of Pittsburgh of UPMC. Study design and implementation were conducted jointly by YZ, RP, and JEL. YZ performed the data analysis and drafted the paper, which was revised and confirmed carefully by RP and JEL. RP revision of the paper mainly focused on the analytical components, whereas JEL reviewed the clinical aspects of the paper. Data interpretation was performed jointly by the three authors.

Competition of interests None.

Ethics approval This study was designated as exempt by the University of Pittsburgh Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data are the property of the Children’s Hospital of Pittsburgh of UPMC.

REFERENCES

37. R Core Team. R: A Language and Environment for Statistical Computing. 2.15.1 edn, Vienna, Austria: R Foundation for Statistical Computing; 2012..