Final Progress Report

Preventing Wrong-Drug and Wrong-Patient Errors with Indication Alerts in CPOE Systems

R01HS024945

Agency for Healthcare Research and Quality

September 30, 2016 — September 29, 2022

VIA EMAIL: grantfpr@ahrq.hhs.gov

Project Officer: Edwin Lomotan

Grants Management Specialist: Brian Campbell

PI: Bruce L. Lambert, PhD

Department of Communication Studies

Northwestern University

Key Personnel:

Jason Adelman, MD

Dulal Bhaumik, PhD

Yelstin Fernandes

William Galanter, MD, PhD

Anne Grauer, MD

Robert Green

Connie Larson, PharmD

David Vawdrey

Yuyang Yang

Structured Abstract

Purpose: Assess how indication alerts influence the likelihood of (a) adding a patient problem and (b) the self-detection of wrong-drug and wrong-patient errors.

Scope: Medication errors, incomplete problem lists, and order abandonments are prevalent across health systems. This research spanned inpatient facilities at Columbia/NYPresbyterian (NYP) with 86 study medications and Northwestern Medicine (NM) covering both inpatient and outpatient facilities with 206 study medications.

Methods: Indication alerts were applied to two Electronic Health Records (EHRs) during pre- and post-intervention stages. These alerts triggered when a study drug was prescribed without a corresponding patient problem, prompting the clinician to address the gap. NYP's alerts were non-disruptive, appearing only during inpatient orders outside of order sets. NM's alerts interrupted the ordering process, targeting only attending physicians in select facilities. Both settings utilized rules to identify medication and patient errors, monitored problem additions based on alerts, and validated problem accuracy. An interrupted time series segmented-regression analysis determined the intervention's effects.

Results: After pooling data, indication alerts led to problem placements rising from 19,873 to 65,325 per 100,000 weekly orders. No significant change was observed in wrong-patient retract-and-reorder events, while wrong-drug retract-and-reorder events declined weekly post-intervention. Wrong-patient abandon-and-reorder incidents increased from 156 to 276 per 100,000 orders, and wrong-drug abandon-and-reorder rose from 280 to 373. In summary, alerts effectively increased problem placement rates, modestly elevated abandonments, and slightly impacted retractions.

Key Words: medication safety, wrong-drug, wrong-patient, indication, clinical decision support, retract-and-reorder, near miss.

PURPOSE

The proposed project builds on our prior work and addresses important gaps in knowledge about how to prevent wrong-drug and wrong-patient errors and how to improve the completeness of problem lists in EHRs. Our specific aims were:

AIM 1. At one health system in Chicago and one in New York City, using two commercial EHR systems, implement a set of 30-50 indication alerts for medications vulnerable to look-alike and sound-alike errors.

AIM 2. Using an interrupted time series study design, quantify the effect of indication alerts on (a) the *combined* rate of self-intercepted *wrong-drug* and *wrong-patient* computerized provider order entry (CPOE) errors and (b) on the rate of each type of error viewed separately. We hypothesize that indication alerts will *increase the combined rate* of self-intercepted wrong-drug and wrong-patient errors by roughly 25%, from 158 to 196 events per 100,000 orders, and will *increase the self-interception rate of each type when viewed separately*, as measured by an increase in the sum of abandon-and-reorder and retract-and-reorder events.

AIM 3. Assess the impact of indication alerts on the probability of adding new diagnoses to the problem list during clinical encounters that include CPOE. We hypothesize that indication alerts will *double* the likelihood that a problem is placed on the problem list during encounters that include CPOE, with new problems being placed during 12% of pre-intervention orders and 25% of post-intervention orders.

SCOPE

Background and Context

COVID-19 Pandemic. This project was originally funded in September 2016. In March 2020, the COVID-19 pandemic struck, hitting New York City with an extreme public health emergency. As the data will show, especially at NYP East Campus, the disruption to inpatient care caused by COVID is evident.

Settings. The original application proposed to do the study at University of Illinois Chicago (UIC) Health and NYP, but midway through the study, UIC announced they were transitioning from the Cerner EHR to the Epic EHR and that there would be a code freeze in Cerner, effectively making it impossible for us to continue at UIC. Fortunately, NM had already made the transition to EPIC a year or two prior to UIC's code freeze, and we were able to negotiate with NM to move the UIC part of the project to NM, although at the cost of perhaps a 18-month delay to get the alerts up at NM. (They had already been implemented in Cerner at UIC.) NYP also transitioned from Allscripts to Epic during the project, which prematurely ended all observations at NYP hospitals as Epic went live at each hospital.

Participants

NewYork Presbyterian (NYP). At NYP, the participants were patients and providers in the inpatient setting at three hospitals (NewYork-Presbyterian Columbia University Irving Medical Center (‘West’), NewYork-Presbyterian Weill Cornell Medical Center (‘East), and NewYork-Presbyterian Queens (‘Queens’)). Prescribers and patients were included if they ordered or received a study medication that was not selected from an order set.

Northwestern Medicine (NM). At NM, participants were attending physicians (and their patients) in the inpatient and outpatient settings at all the hospitals and clinics in the health system. (All NM locations can be found at https://www.nm.org/locations). Prescribers and patients were included if they ordered or received a study medication that was not selected from an order set. To minimize nuisance, orders from the emergency department and urgent care, as well as refill orders, were excluded.

Incidence and Prevalence

Wrong Drug Errors. The prevalence and rate of wrong-drug errors vary depending on the measurement method, the clinical context, the availability of CPOE, and what stage of the process is being examined. The literature provides several estimates. In terms of the proportion of *errors* (not *orders*) that are wrong drug, Pham found 11% of 13,932 emergency department (ED) medication errors were wrong-drug.[1] A study of errors reported to MEDMARX found 9.7% of errors in the ICU and 12% of all med errors were wrong-drug.[2] An inpatient study of dispensing errors found roughly 11% of all errors detected by direct observation were wrong-drug, and such errors were the most common cause of ADEs (accounting for 36% of the 1159 ADEs detected).[3] The same study showed wrong-drug errors were present in 0.39% (554 of the 140,755) of all doses. Of these, 25% were undetected by pharmacists. Thus 0.1% of inpatient doses resulted patients getting the wrong-drug (25%*0.39%). Direct observation of 645 doses in an ICU detected 12 wrong-drug errors (1.9%).[4]

In terms of order entry errors, a 2014 emergency department study, 11 of 3136 orders were for the wrong drug (0.35%). Chart review of 1582 orders over one month in a Saudi Arabian hospital found 10 wrong-drug errors (0.63%).[5] Pharmacist review of 33,012 inpatient orders in a U.K. hospital found 39 such errors (0.1%).[6] A similar study in an Iranian nephrology ward found 17 wrong-drug errors in 818 orders (2%).[7] Forty-one (41) wrong-drug errors were found in 2380 pediatric prescriptions (1.7%).[8] An Australian study of Cerner and MedChart EMRs found 2 wrong-drug errors per 100 admissions in MedChart and 8.5 per 100 admissions in Cerner.[9] If one assumes 20 med orders per admission, this would translate to rates of 0.1% for MedChart and 0.42% for Cerner. Our own work detected intercepted wrong-drug errors in 176 of 127,458 alerted orders (0.14%).[10]

Wrong Patient Errors. One known risk of CPOE is that a prescriber may accidentally enter orders in the wrong-patient’s chart.[11-19] Paper charts afforded prescribers visual cues that oriented them to patient’s identity, including the thickness of the chart, the handwriting, and the patient’s problem list. These cues are missing or less salient in the electronic chart, increasing the opportunity for wrong-chart errors. Use of patient lists may cause "pick-list" or other user-interface driven errors, and wrong-patient selection may be facilitated by system features that allow clinicians to find charts rapidly, such as the ability to select from a list of recently opened charts, or to have multiple patient charts open simultaneously.[11,12,15] Wilcox estimated that 0.3% to 0.5% of clinical notes were placed in the wrong-patient’s electronic chart.[20] Pham found that 4.4% of

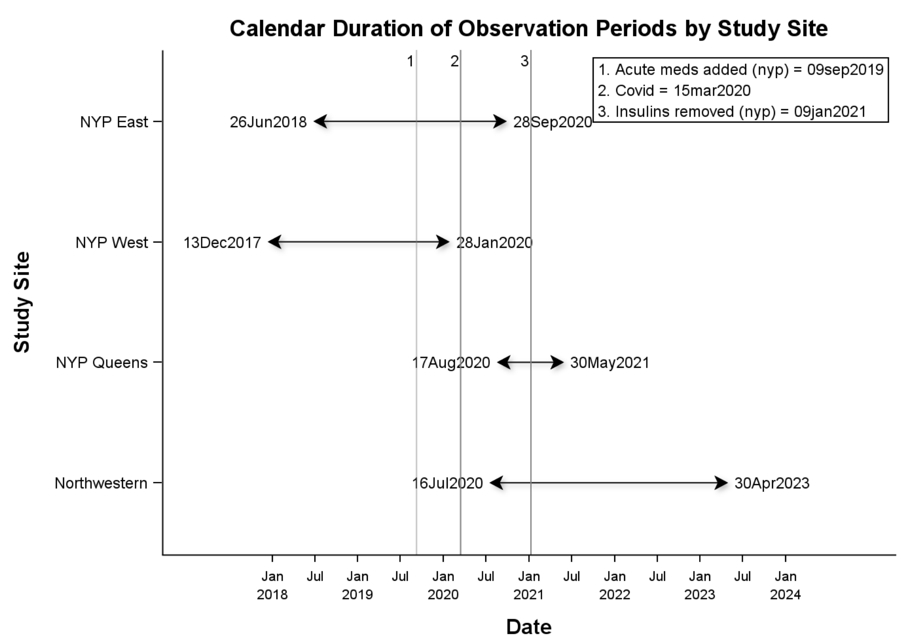
13,932 reported medication errors in the ED were wrong-patient errors.[1] A study of the MEDMARX database identified 4.3% of ICU and 4.6% of all medication errors were wrong-patient.[2] Adelman et al. found 5246 wrong-patient orders at one hospital in 2009 alone (14 per day), and estimated that in about 1 per 1700 medication orders clinicians realized that they had placed the order on the wrong-patient and cancelled it prior to medication administration.[17] The same study evaluated the severity of the errors and rated 166 life-threatening, 359 serious, and 1274 clinically significant (out of 9 million orders).[17] These errors can be fatal.[21] Identification confirmation screens have been reasonably successful in reducing the rate of wrong-patient errors, but additional interface innovations, such as indication alerts, could reduce the rate even further.[15,17-19]

METHODS

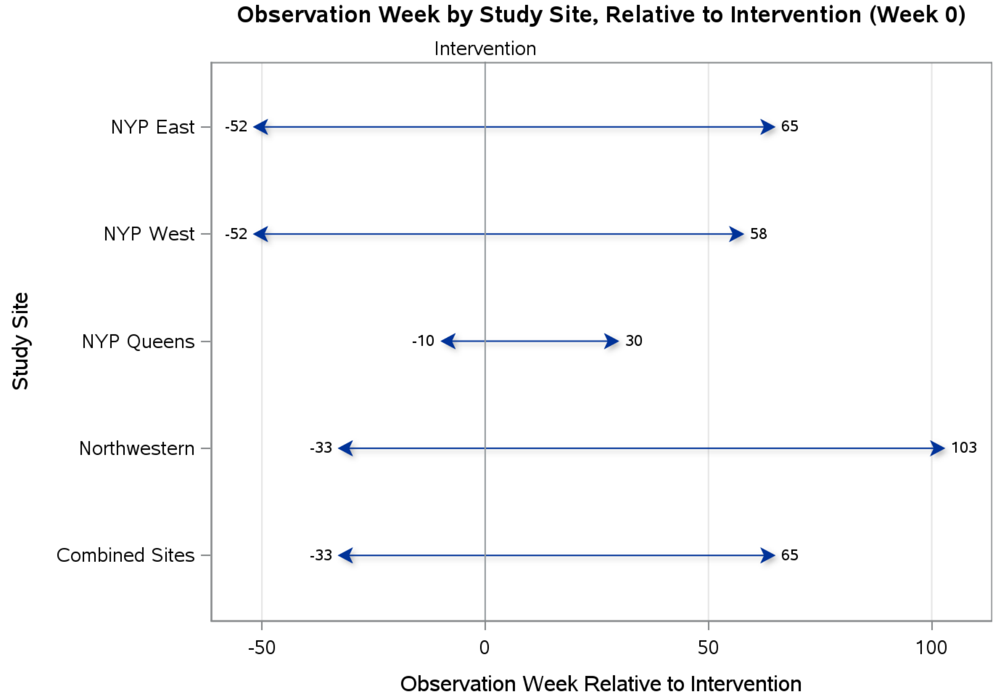
Design. We used a before-and-after, interrupted time series design to quantify the effects of indication alerts on our 5 outcomes (wrong-drug and wrong-patient retract and reorder events; wrong-drug and wrong-patient abandon-and-reorder events; and problems placed).

Intervention. The intervention involved the implementation of indication alerts in two different health systems using two different EHRs (AllScripts at Columbia/NYP and Epic at NM). The original proposal included UIC Health as the Chicago site, but UIC switched from Cerner to Epic and cancelled our project as a result. (See explanation in Settings section above.). At Northwestern, the pre-intervention period was from 7/16/2020 to 3/6/21. There was an initial, pilot go-live period between 3/7/21 and 3/16/21 that was turned off to implement an exclusion for the emergency department and urgent care (based on user feedback). The full intervention period began 5/3/21.

At NYP, there were three participating hospitals (East Campus, West Campus, and Queens Campus) with slightly different start and end dates, but overall observation began on 12/13/17 and ended on 5/30/21. Observation at NYP ended prematurely when each hospital transitioned from Allscripts to Epic. See timelines below for details. The first timeline shows the total observation periods at each site and includes certain important events that took place during observation.



The second timeline shows the observation periods centered on the intervention date at each site, with times shown in weeks before and after the intervention week, represented as week zero. The main analysis was done on combined data that began 33 weeks before the intervention and ended 65 weeks after.



Development of Indication Alerts

This section was taken, with minor edits, from our 2022 paper on indication alerts and problem lists.[22] The development of the indication alerts started with the clinical logic that was developed previously.[6-9] The relationships between indications and medications were updated by pharmacists and physicians on our research team. These disease-medication relationships, or "rules," were then provided to the implementation teams at both institutions. Each site chose which rules to implement based on local medication use and formularies. Both sites made local modifications, and some additional rules were developed at each site.

At NYP, the indication alerts were built in Allscripts, and in contrast to the interruptive alert previously described[6-10] when a clinician selected a medication with no corresponding problem in the patient's problem list, a new required field appeared in the CPOE interface. After clicking on the field, a dialog box appeared and gave the clinician the option to select from ICD-10 codes to be added to the patient's problem list corresponding to the set of indications for the selected medication. Once the clinician signed the order, the selected indication was placed on the problem list. If none of the listed ICD-10 code choices offered was appropriate, the clinician could click "not applicable" and had the choice to enter a free text indication for the medication. The alert's required fields had to be completed before the clinician could complete the order.

At NM, the indication alert was built as an interruptive Best Practice Advisory (BPA) in Epic. To minimize potential nuisance and to accommodate to local preferences, the alert did not trigger during emergency department or urgent care visits. Similarly, to maximize specificity, if the patient had ever been prescribed the medication previously in the EHR, the alert did not fire.

The problem list must be populated with SNOMED nomenclature as required by Meaningful Use Stage II.[11] At NYP, when a study medication order was attempted, the Medical Logic Module (MLM) reviewed the patient's coded problem list (termed Health Issues in Allscripts) and determined if the patient had an appropriate ICD-10 coded health issue in their record. If the MLM determined an appropriate ICD-10 coded health issue was not present, a mandatory field with an associated dialog box would display prompting the clinician to select an appropriate ICD-10 code. When the clinician selected one of the displayed ICD-10 codes and submitted the order, a health issue with that ICD-10 code was created. The ICD-10 codes were mapped to SNOMED concepts and the SNOMED concepts displayed alongside the ICD-10 code in the Health Issue section of the patient's medical record. At NM, the selected problem to be added was chosen from a set of clinical terms (source: Intelligent Medical Objects, Chicago IL) embedded in the EHR. The selected diagnosis clinical term prompted by the BPA was mapped both to a SNOMED concept for use in the problem list as required by Meaningful Use Stage II, and to an ICD10-code for billing if the diagnosis was subsequently moved to the visit/billing diagnosis list.

NYP built indication alerts for 88 medications while NM built indication alerts for 206 medications; 69 of these medications were used for alerts at both sites. The total combined number of unique medication indication alerts was 225. The medications, site usage, therapeutic categories, excluding codes (i.e., codes which if present in the problem list prevented the alert from triggering), and corresponding suggested diagnoses are included in Appendix 1 of the published paper.[22] Also included are any age or sex logic that were criteria for the alert to trigger. At NYP, problems were accurate according to clinician chart review 88% of the time, and 90% of the time at NM. Full details of the quality of the resulting problem lists have previously been published.[22]

Data Collection and Outcome Measurement

We collected data from the electronic data warehouses at the respective systems. This presented challenges because, as we discovered, the data available in the production EHR at the time of ordering (when the alert logic runs) is not identical to the data available in the EHR. For example, data were stored in different tables, sometimes with different variable names. Some data available in the production system were not available directly from the EHR, such as precise datetimes, and the status of the problem list at each moment in time.

This is critical because of the way we measure outcomes. For the comparison between the pre-intervention period and the post-intervention period to be valid, outcomes in each period must be measured in the same way. If they are not, then any differences between pre and post could be due to measurement differences rather than to the effects of the intervention. The goal of measurement is to identify every order for a study drug, determine whether an alert fired (or would have fired in the pre-intervention period), and then to determine (a) whether a problem was placed or (b) the subsequent pattern of actions fit any of our 4 main outcomes (WP-RAR, WD-RAR, WP-AAR, or WD-RAR). In the post-intervention period, it is straightforward to know when alerts occurred because the EDW provides a log of every alert that fires. But in the pre-intervention period, alerts did not fire, and hence there was no alert log. To address this, we had to write software that would simulate the alert logic using only EDW data as input.

Analysis Plan

Outcome Definition and Measurement

There were 5 individual outcomes: problems placed, WP-RAR, WD-RAR, WP-AAR, and WD-AAR. For the abandonment rates and problem placement, the denominator was the number of *initiated, eligible, alertable* study drug orders. We define eligible, alertable order below. For the retraction rates, the denominator was the number of *completed, eligible, alertable* study drug orders. The temporal unit of analysis was the week. We had initially planned on using months, but the alert volume was high enough to let us use weeks instead.

Types of Medication Orders

Our analysis distinguished between several types of medication orders. The first distinction pertains to whether the order is for a study drug. The second distinction refers to whether the order was initiated only or initiated and completed. We say *initiated* because not all orders are completed and signed. Some are abandoned (i.e., never signed), and abandonment is one of our outcomes. To capture abandonment, it is critical to detect as soon as a drug is selected in the CPOE system. We refer to that selection action as 'order initiation.' An order that is initiated but not signed is categorized as abandoned. An order that is signed is categorized as completed. The third distinction refers to whether an order was *alertable,* where *alertable* is defined as meeting the criteria for an indication alert (i.e., an order for a drug in the absence of a matching indication in the problem list). An alertable order would produce an actual alert in the post-intervention period but not in the pre-intervention period. The final distinction concerns whether a study drug order was *eligible* for an alert. To accommodate local constraints and preferences, several types of orders were excluded. At NYP we excluded all outpatient orders and all orders in order sets. At NM, we excluded all orders in order sets and all orders from urgent care clinics and emergency departments, and also refill orders.

Selecting the Numerator and Denominator

Each of the main outcomes is a rate or a proportion, so the definition of each outcome requires the specification of the numerator and denominator.

Numerator. The numerator in each case is the number of events in each period (e.g., per week). The specific definition of each event is given below, but once each is defined, counting them is simple.

Denominator. The decision about which denominator to choose reflects choices about what is being measured and about the desired sensitivity of measurement. In the original proposal, we stated we would use the total number of initiated study drug orders, but there are several alternatives: (1) all initiated drug orders (regardless of whether they were study drugs); (2) all initiated orders for study drugs; and (3) all initiated orders for study drugs that were alertable, (4) all completed study drug orders that were alertable, and (5) all completed study drug orders regardless of alert status.

1. **All initiated drug orders**. This denominator provides the least sensitivity. Since only a subset of all medication orders are study drugs, and alerts did not fire on non-study drugs, then the effect of our alerts would have to be massive to be detected against the background of so many unalerted orders.
2. **All initiated study drug orders**. This is what we said we would use in the original proposal. But it will also be somewhat insensitive because many study drug orders are not alertable because the patient already has the matching problem on their problem list.
3. **All alertable initiated study drug orders.** This is the purest and most sensitive measure of the effect of the alerts since it only measures outcomes after an alert fired.
4. **All alertable completed study drug orders.** This is like what has been used in measurement of WP-RAR in past studies.
5. **All completed study drug order.** This is what we used for RAR outcomes because technical malfunctions at NYP-W prevented us from having valid data on study drug initiations or simulated alert rates.

Order initiations were detected using a silent alert that fired when a study drug was selected in the CPOE interface. Due to a technical error at NYP West Campus, soon after our observation period began, this alert was accidentally set to activate only on orders in order sets, making it impossible to capture order initiations and hence to measure abandonment at NYP West campus. To be able to include West Campus data on RAR events, we used all *completed* study drug orders for those outcomes, regardless of alert outcome. For all other outcomes, we used *initiated alertable* study drug orders.

Measurement Tools for Self-Intercepted Errors. We will combine two different methods to the measure the rate of self-intercepted errors: the retract-and-reorder method (RAR)[17] and the abandon-and-reorder method (AAR).[10,23] We describe them below, along with our approach to implementation and deployment.

The Retract-and-Reorder Tool. The RAR, developed by Dr. Jason Adelman (a co-investigator at NYP on this project), was developed as a method for detecting wrong-patient errors.[17,23] In that scenario, an intercepted error was defined by the following sequence of events: a prescriber orders Action A, retracts it within 10 minutes of the initial order, and then the same prescriber orders Action A on a different patient within 10 minutes of the retraction. Orders are not counted as RAR events if the same canceled action is ordered initial patient within 24 hours of the original order. When the tool is used to detect intercepted wrong-patient errors, we will refer to it as the WP-RAR tool. When it is modified slightly to detect wrong-drug errors, we will refer to it as the WD-RAR tool. The WP-RAR tool has been officially endorsed as a quality measure by the National Quality Forum (NQF Measure #2723).[24]

The Abandon-and-Reorder Tool. Galanter, Lambert and colleagues used a modified version of the RAR tool to measure the rate of intercepted wrong-drug and wrong-patient errors.[10,23] This version of the tool detects orders that are abandoned *before* they are signed or retracted, so we refer to it as the *abandon*-and-reorder (AAR) tool. The tool uses the EMR to capture when a medication order is started (a drug is selected) but not completed. This is the "abandon" step in the AAR process. The wrong- drug version of the tool, which we will refer to as WD-AAR, then captures when a similar drug is ordered for the same patient by the same provider within 10 minutes of the original order. The wrong-patient version of the tool (WP-AAR) detects when a medication order is started, not completed, and then the same medication is reordered on a different patient by the same prescriber within 10 minutes of the original order.

Intercepted Wrong-patient (WP) Error Rate. The number of intercepted wrong-patient errors per period was measured using the WP-RAR and WP-AAR tools. The WP-RAR has been validated[17] and is now an NQF-approved quality measure.[24] The intercepted *wrong-patient error rate* per period is given as the sum of the WP-RAR and WP-AAR events occurring at both health systems, divided by the total number of alerted orders for study medications, also summed across both health systems.

Eq. 1. *Intercepted Wrong-Patient Abandon-and-Reorder* = $\frac{(WP\text{-}AAR_{NM} + WP\text{-}AAR_{NYP})}{Total\ Number\ of Initiated\ Alertable\ Orders\ for\ Study\ Drugs}$

Eq. 2. *Intercepted Wrong-Patient Retract-and-Reorder* = $\frac{(WP\text{-}RAR_{NM} + WP\text{-}RAR_{NYP})}{Total\ Number\ of\ Completed\ Alertable\ Orders\ for\ Study\ Drugs}$

Wrong-Patient Definition. WP-RAR was measured as described above. For a WP-*RAR* event to be confirmed as an error, the following criteria must be met:

1. An order was *completed* for Drug A for a given patient.
2. Same prescriber completed an order for Drug A on different patient within 10 minutes of initial order.
3. Drug A is not ordered for the original patient by any prescriber within 24 hours of initial order.

Because we had to use different denominators for abandonment and retraction events, we separated wrong-patient errors into wrong-patient abandonments and wrong-patient retractions.

2.C.3.3.4. Intercepted Wrong-Drug (WD) Error Rate. The *number* of intercepted wrong-drug errors in a given period is defined as the sum total of WD-AAR events and WD-RAR events arising from CPOE orders for study drugs across both systems. The intercepted wrong-drug error *rate* is given as the sum of WD-AAR and WD-RAR (at both health systems) divided by the total number of alertable orders for the study drugs (again, summed across both systems).

Eq. 3. *Intercepted Wrong-Drug Abandon-and-Reorder* = $\frac{(WD\text{-}AAR_{NM} + WD\text{-}AAR_{NYP})}{Total\ Number\ of Initiated\ Alertable\ Orders\ for\ Study\ Drugs}$

Eq. 3. *Intercepted Wrong-Drug Retract-and-Reorder* = $\frac{(WD\text{-}RAR_{NM} + WD\text{-}RAR_{NYP})}{Total\ Number\ of\ Completed\ Alertable\ Orders\ for\ Study\ Drugs}$

2.C.3.3.5. Wrong Drug Definition. For a WD-AAR event to be confirmed as an error, all of the following criteria must be met, with steps (1)–(4) done by computer program. Orders made during medication reconciliation will be excluded because they proved to be a frequent source of false positives in our earlier study.[10] Steps 1-4 constitute the WD-AAR event itself:

1. An order was *started but not completed* for Drug A.
2. Same prescriber completed an order for Drug B on the same patient within 10 minutes of initial order.
3. The names of Drug A and Drug B had a Bisim and Editex similarity score exceeding 0.40. (Bisim and Editex are validated numerical measures of look-alike and sound-alike similarity.) [10,25,26]
4. Drug A is not ordered for the same patient by any prescriber within 24 hours of initial order.

For a WD-RAR event to be confirmed as an error, all of the following criteria must be met, with steps (1)–(4) done by computer program. Orders made during medication reconciliation will be excluded.[10] Steps 1-4 constitute the WD-RAR event itself:

1. An order was *completed* for Drug A.
2. Same prescriber canceled order for Drug A and ordered Drug B for same patient within 10 minutes of the initial order.
3. The names of Drug A and Drug B had a Bisim and Editex similarity score exceeding 0.40.[10,25,26]
4. Drug A is not ordered for the same patient by any prescriber within 24 hours of initial order.

For all retract-and-reorder outcomes, we used a combination of NM and NYP East, West, and Queens campuses. For abandonment outcomes, we had to exclude NYP-West because the silent alert that measured order initiation malfunctioned.

Results

Participants

Northwestern Medicine. NM provided data from 33 weeks prior to the intervention until 103 weeks after, but to align with the large NYP sites (East and West campus), we only used NM data from week -33 to week 65. Demographic and other descriptive data are given only for this period. At NM, n=172,744 (simulated) alerts fired on orders for n=132,189 unique patients, of whom 56% were women and 44% men, with mean age of 53.7 (s.d. = 19.8, mode = 64, median = 55). Medication orders were written by n=2748 unique prescribers. Prescriber gender was missing for 1833 individuals, but among those for whom we had gender, 54% were men and 46% women. Internal medicine was the most common clinical location (18%) followed by primary care (12%), cardiology (x%), neurology (6%), dermatology (5%), and ophthalmology (4%), and. The most common encounter types were hospital encounter (45%), office visit (37%), orders only (8%), and telephone (6%). Prescribers' mean age was 43.1 (s.d.=10.9, median=41, mode=34). Alerts fired on 188 different medications, the top 5 of which were: albuterol (7.6%), cholecalciferol (5.8%), ergocalciferol (5.6%), atorvastatin (5.5%), ferrous sulfate (5.0%), and lorazepam (3.9%).

Columbia/NYP. NYP provided data from 52 weeks prior to the intervention and up to 65 weeks after (though these periods differed across sites, see the observation week timeline on page 5 above). To align with NM, we only used data from week -33 to week 65 post. At NYP-E and NYPQ, where we had complete data, n=213,465 (simulated) alerts fired. Data at NYP were aggregated at the daily level and we did not have order-level data to tabulate drug frequency or demographic data.

Alert Characteristics

Accuracy of Simulated Alert Software. As noted above, alerts did not fire in the pre-intervention periods (which were examined retrospectively), and differences between real-time EHR data and EDW data made it impossible for us to perfectly simulate the alert in the pre-intervention period. Here we describe the accuracy of the simulated alert.

Northwestern Medicine. The figure below gives the confusion matrix for the simulated alert software used at NM. Sensitivity, specificity, and positive predictive value are defined in the conventional manner. Note that the confusion matrix in the table below could only be computed for study drug orders in the post-intervention period when we had actual alert log data to serve as the gold standard. These data were calculated on the full NM dataset, not just that from -33 to 65 weeks.

The F-measure is defined as

$$TP/(TP + \frac{FP + FN}{2})$$

Table 1. Performance characteristics of simulated alert code at NM

Measure	Value
Total Unique Alerts	225
Actual Positives	161,119
Actual Negatives	3,879,209
Total Eligible Study Drug Orders	4,094,525

Measure	Value
Accuracy	0.986
Sensitivity (Recall)	0.988
Specificity	0.986
Positive Predictive Value (Precision)	0.746
Negative Predictive Value	0.999
F-Score	0.850
Actual Alert Rate/Yield	0.039

Confusion Matrix:

	Predicted Positive	**Predicted Negative**
Actual Positive	159,167 (TP)	1,952 (FN)
Actual Negative	54,197 (FP)	3,879,209 (TN)

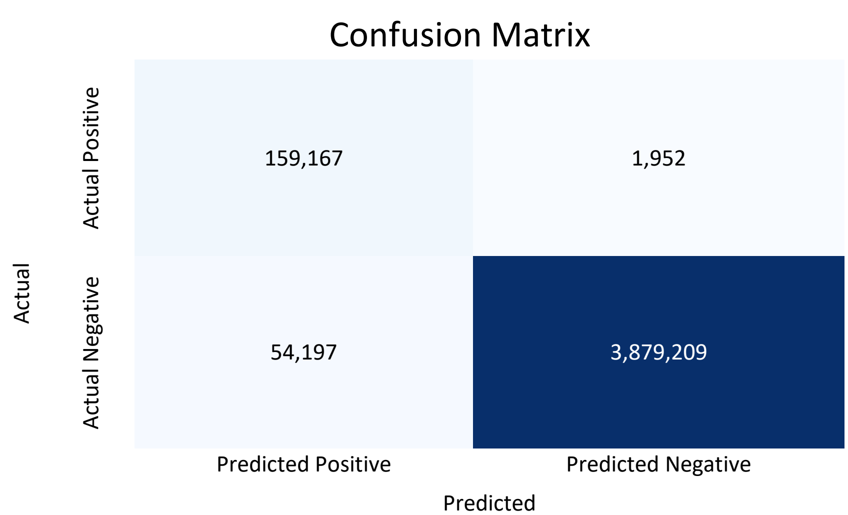


Figure 1. Confusion matrix for simulated alert code at NM.

Columbia/New York Presbyterian. Data unavailable at time of report.

Alert Volume and Rate. We define alert volume as the number of alerts that fired overall and per period (pre- vs. post-intervention). Alert rate is defined as the ratio of alerts to total eligible study drug initiations. Each eligible study drug initiation was an opportunity for an alert, so it is used at the denominator in the alert rate. Note that since actual alert logs were available only for the post-intervention period (i.e., no alerts existed in the pre-intervention period), actual alert volume and rate are only available for the post-intervention period. Volume and rate of simulated alerts are available for both periods and for the overall period.

Across all sites and over the entire observation period, there were a total of 3,318,854 completed study drug orders. During the post-intervention period, when actual alerts could be directly observed, there were 291,452 alerts (alert rate per completed order = 12.68%). Table 2 gives the number of study drug initiations, completions, actual and simulated alert volumes and rates, by study site and intervention period. The fist panel of the table (2A) includes NM, NYP-E, and NYP-Q. NYP-W, which experienced technical problems that prevented the collection of order initiations and simulated alert data, is described in more limited fashion in Table 2B.

Table 2A. Study drug use, actual and simulated alert volume, and alert rate by study site and intervention period

Site	Period	Study Drug Initiations	Study Drug Completions	Abandonment Rate	Actual		Simulated		
					Alert Volume	Alert Rate	Alert Volume	Alert Rate (per Completed Order)	Alert Rate (Per Initiated Order)
Northwestern	Pre	882,907	793,400	10.14%	.	.	56,829	7.16%	6.44%
	Post	1,946,169	1,747,156	10.23%	84,436	4.83%	115,915	6.63%	5.96%
	Overall	2,829,076	2,540,556	10.20%			172,744	6.80%	6.11%
NYP-East	Pre	108,272	97,160	10.26%	.	.	63,397	65.25%	58.55%
	Post	303,359	267,923	11.68%	73,674	27.50%	111,441	41.59%	36.74%
	Overall	411,631	365,083	11.31%			174,838	47.89%	42.47%
NYP-Queens	Pre	20,948	18,565	11.38%	.	.	12,186	65.64%	58.17%
	Post	67,802	59,785	11.82%	21,284	35.60%	26,441	44.23%	39.00%
	Overall	88,750	78,350	11.72%			38,627	49.30%	43.52%
Combined	Pre	1,012,127	909,125	10.18%	.	.	132,412	14.56%	13.08%
	Post	2,317,330	2,074,864	10.46%	235,423	11.35%	253,797	12.23%	10.95%
	Overall	3,329,457	2,983,989	10.38%	.		386,209	12.94%	11.60%

Table 2B. NYP-W

NYP-West	Pre	.	110,457		.	.
	Post	.	224,408		56,029	24.97%
	Overall	.	334,865			.
Combined			3,318,854		291,452	12.68%

Note. N=33 weeks of pre-intervention observation and n=65 weeks of post-intervention observation. Simulated alerts did not function properly, and study drug initiations were not captured properly at NYP-West. Actual alerts did not fire during the pre-intervention period at any study site. Pre-intervention alert counts were all simulated by running the alert logic on retrospective EDW data.

Main Outcomes

For each of the analyses, the segmented regression model tests three main effects: (1) the effect of time elapsed form the start to the end of the entire observation period; (2) the effect of the intervention itself; and (3) the effect of time elapsed after the intervention.[27] The time effect measures any secular trend in the outcome regardless of the intervention. The intervention effect measures the immediate effect of the intervention without regard to time or time after the intervention. The time after the intervention effect measures the effect of time in the post intervention period only (e.g., it is set to 0 during the pre-intervention, and then it is a sequential count of weeks elapsed since the intervention.

Wrong-Patient Retract and Reorder (WP-RAR). Figure 2 shows the WP-RAR rate before and after the intervention (represented by the vertical line). The effect of time was not significant (b=0.02, t=0.17 p=0.86). The effect of the intervention was not significant (b=0.39, t=0.12, p=0.90). The effect of time after the intervention was not significant (b=0.002, t=0.15, p<0.99). Table 3 gives the parameter estimates for the segmented regression analysis. The overall fit of the model was R^2=0.01, F(3, 95)=0.45, p=0.72.

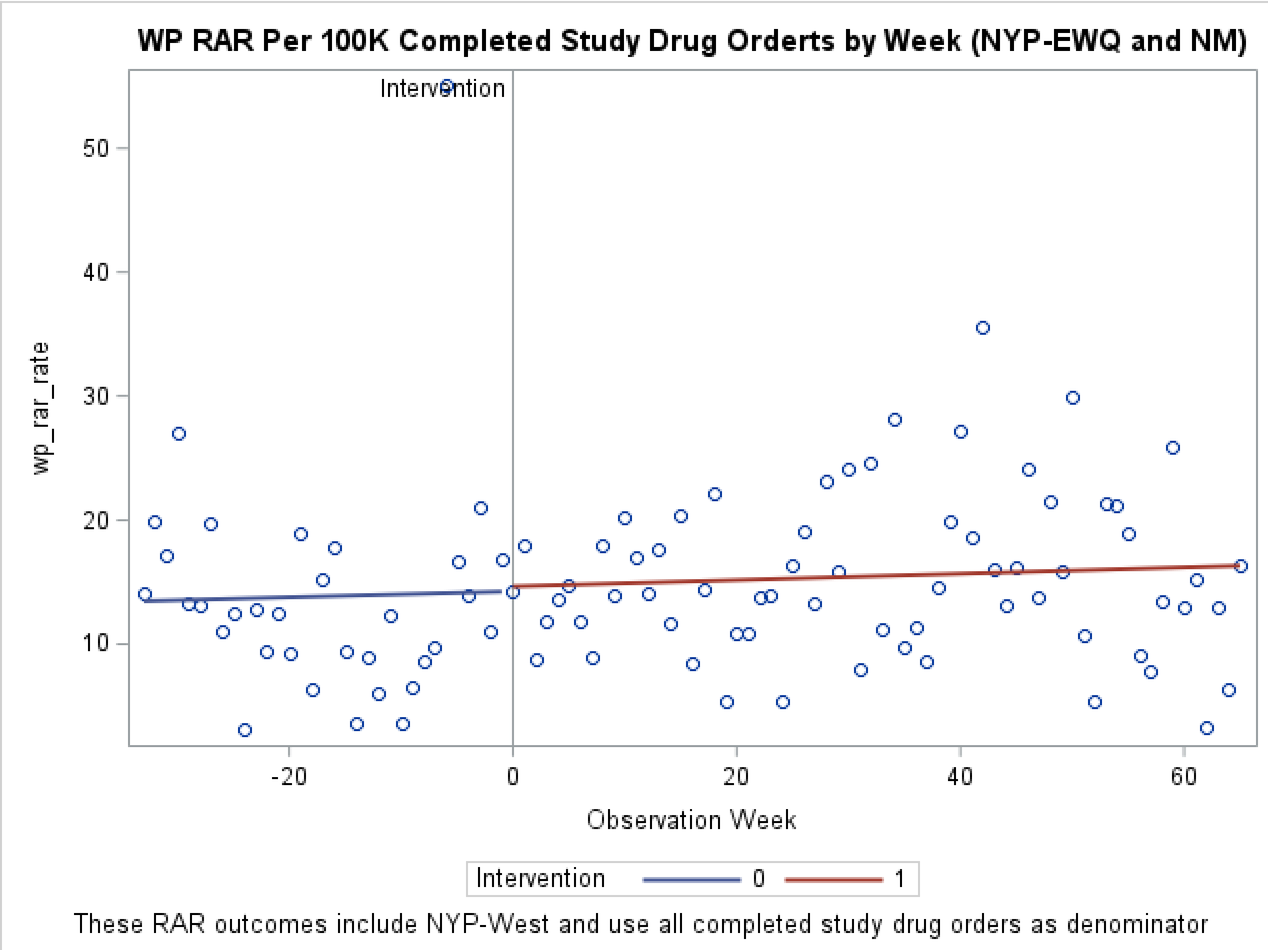


Figure 2. WP-RAR events per 100k completed study drug orders.

Table 3. Parameter estimates from regression of all WP-RAR rates on time, intervention, and time after intervention.

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	12.99827	5.12072	2.54	0.0128
time	Elapsed Time (Weeks)	1	0.02385	0.13751	0.17	0.8627
tai	Time After Intervention (Weeks)	1	0.00182	0.14585	0.01	0.9901
int	Intervention	1	0.39488	3.17231	0.12	0.9012

Wrong-Drug Retract and Reorder (WD-RAR). Figure 3 shows the WD-RAR rate, before and after the intervention (represented by the vertical line). The effect of time was significant (b=0.18, t=1.99, p=0.05). The effect of the intervention was not significant (b=-1.11, t=-0.54, p=0.59). The effect of time after the intervention was significant (b=-0.21, t=-2.24 p=0.03). Table 4 gives the parameter estimates for the segmented regression analysis. The overall fit of the model was poor, R^2=0.02, F(3, 92)=1.82, p<0.15.

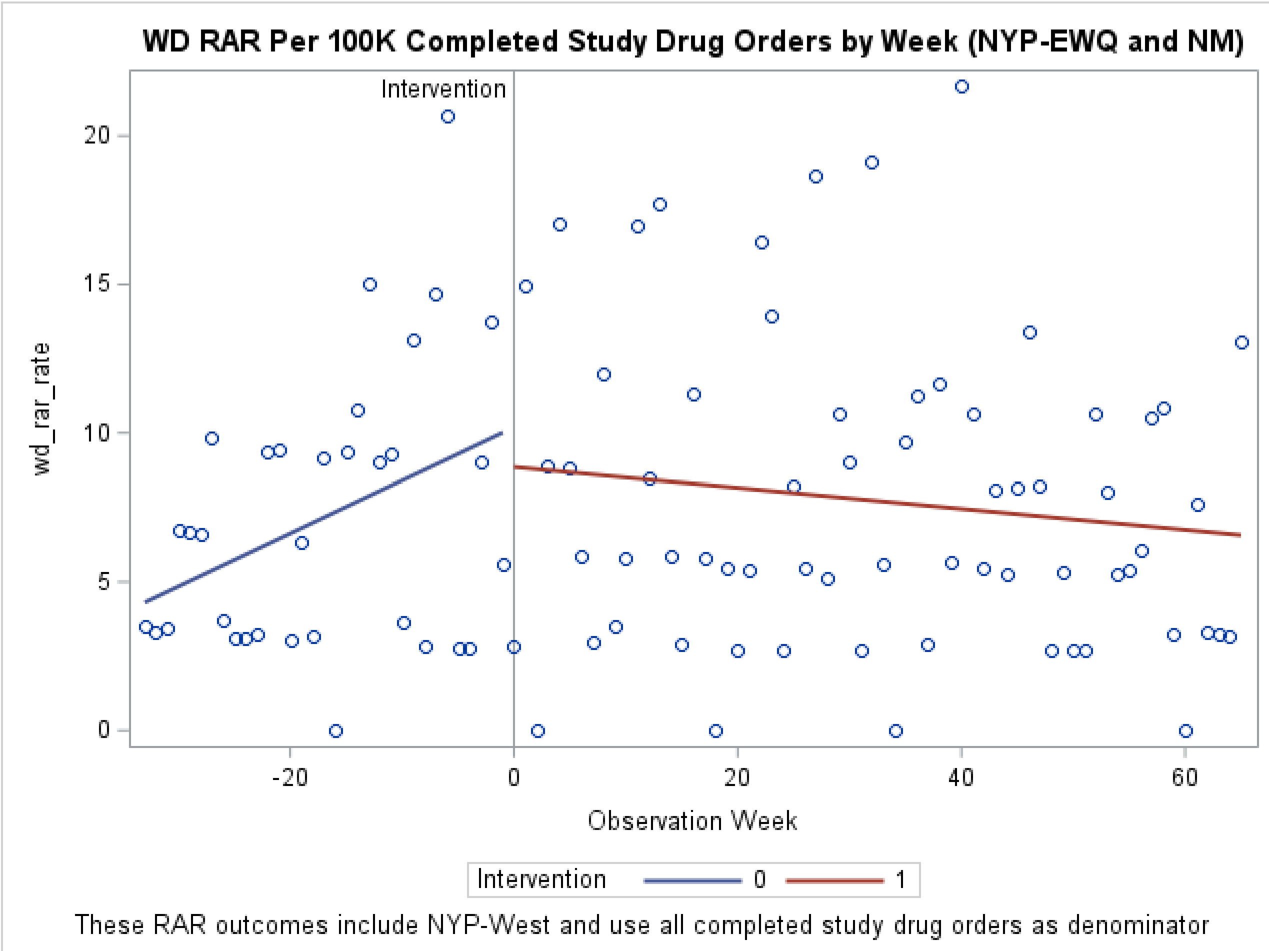


Figure 3. Wrong-drug retract and reorder events per 100k completed study drug orders.

Table 4. Parameter estimates from regression of wrong-drug retract-and-reorder rate on time, intervention, and time after intervention

Label	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.75472	3.33708	0.23	0.8216
Elapsed Time (Weeks)	0.17801	0.08962	1.99	0.0499
Time After Intervention (Weeks)	-0.21315	0.09505	-2.24	0.0273
Intervention	-1.11657	2.06734	-0.54	0.5904

Wrong-Patient Abandon and Reorder (WP-AAR). Figure 4 shows the relationship between the WP-AAR and the intervention. The effect of time was not significant (b=0.18, t=0.11, p=0.91). The effect of the intervention was significant (b=78.1, t=2.03, p<.0.04). The effect of time after the intervention was not significant (b=0.97, t=0.55, p=0.59). Table 5 gives the parameter estimates for the segmented regression analysis. The overall fit of the model was moderate, R^2=0.28, F(3, 95)=13.9, p<.0001.

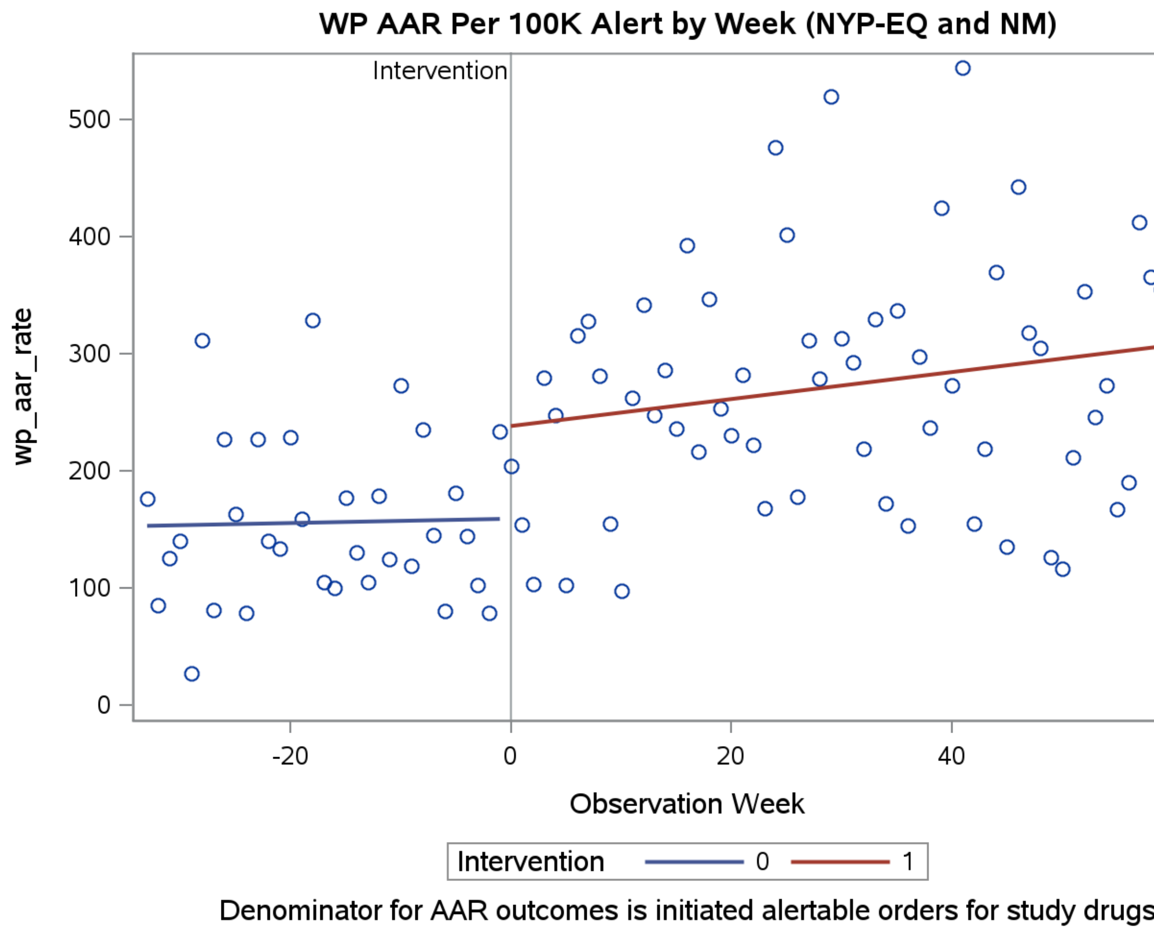


Figure 4. Wrong-Patient Abandon and Reorder per 100k initiated alertable study drug orders.

Table 5. Parameter estimates from regression of wrong-patient abandon and reorder rate on time, intervention, and time after intervention.

Label	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	149.52129	62.05246	2.41	0.0179
Elapsed Time (Weeks)	0.18423	1.66638	0.11	0.9122
Time After Intervention (Weeks)	0.96779	1.76739	0.55	0.5853
Intervention	78.17867	38.44184	2.03	0.0448

Wrong-Drug Abandon and Reorder (WD-AAR). Figure 5 shows the relationship between the WD-AAR and the intervention. The effect of time was not significant (b=-2.89, t=-1.47, p=0.15). The effect of the intervention was significant (b=107.45, t=2.37, p<.0.02). The effect of time after the intervention was not significant (b=3.83, t=1.83, p<0.07). Table 6 gives the parameter estimates for the segmented regression analysis. The overall fit of the model was moderate, R^2=0.15, F(3, 95)=6.78, p=.0003.

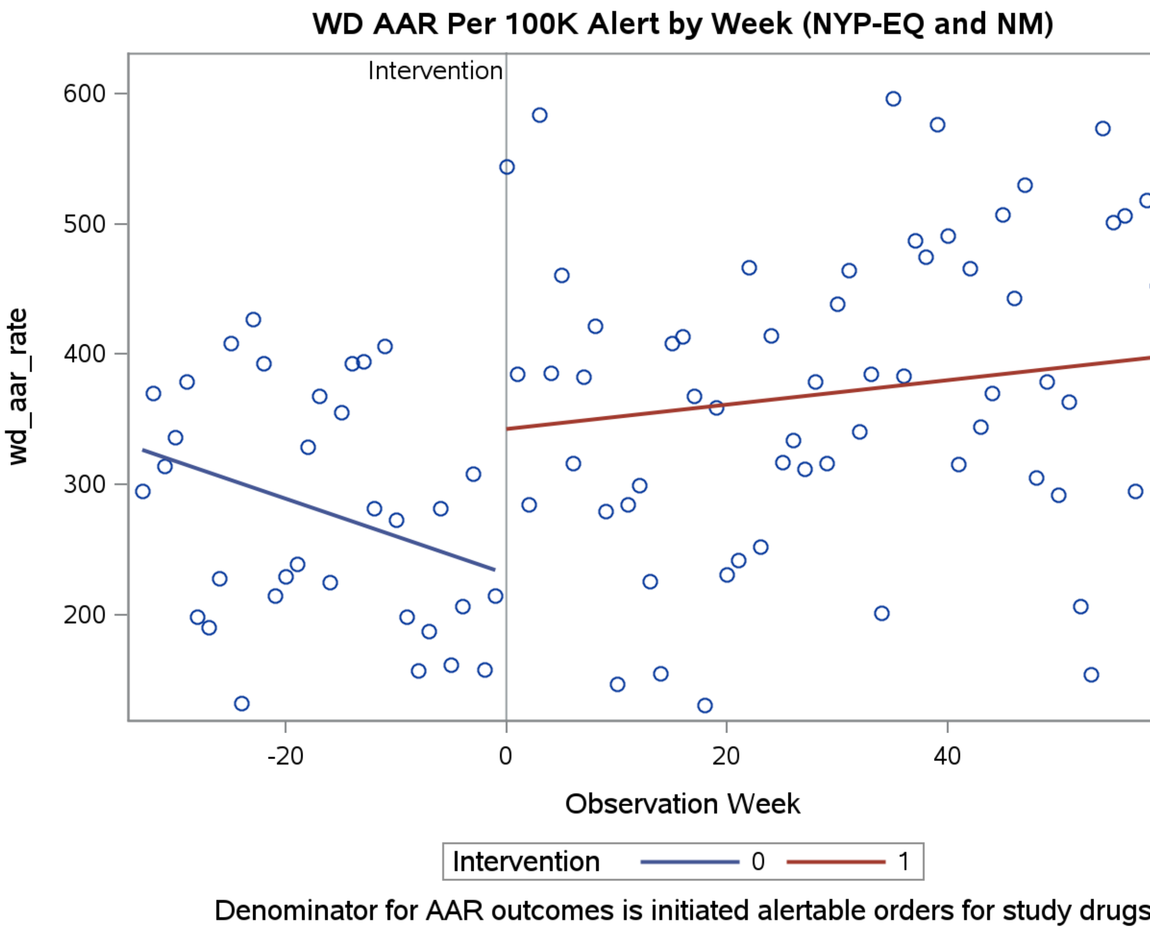
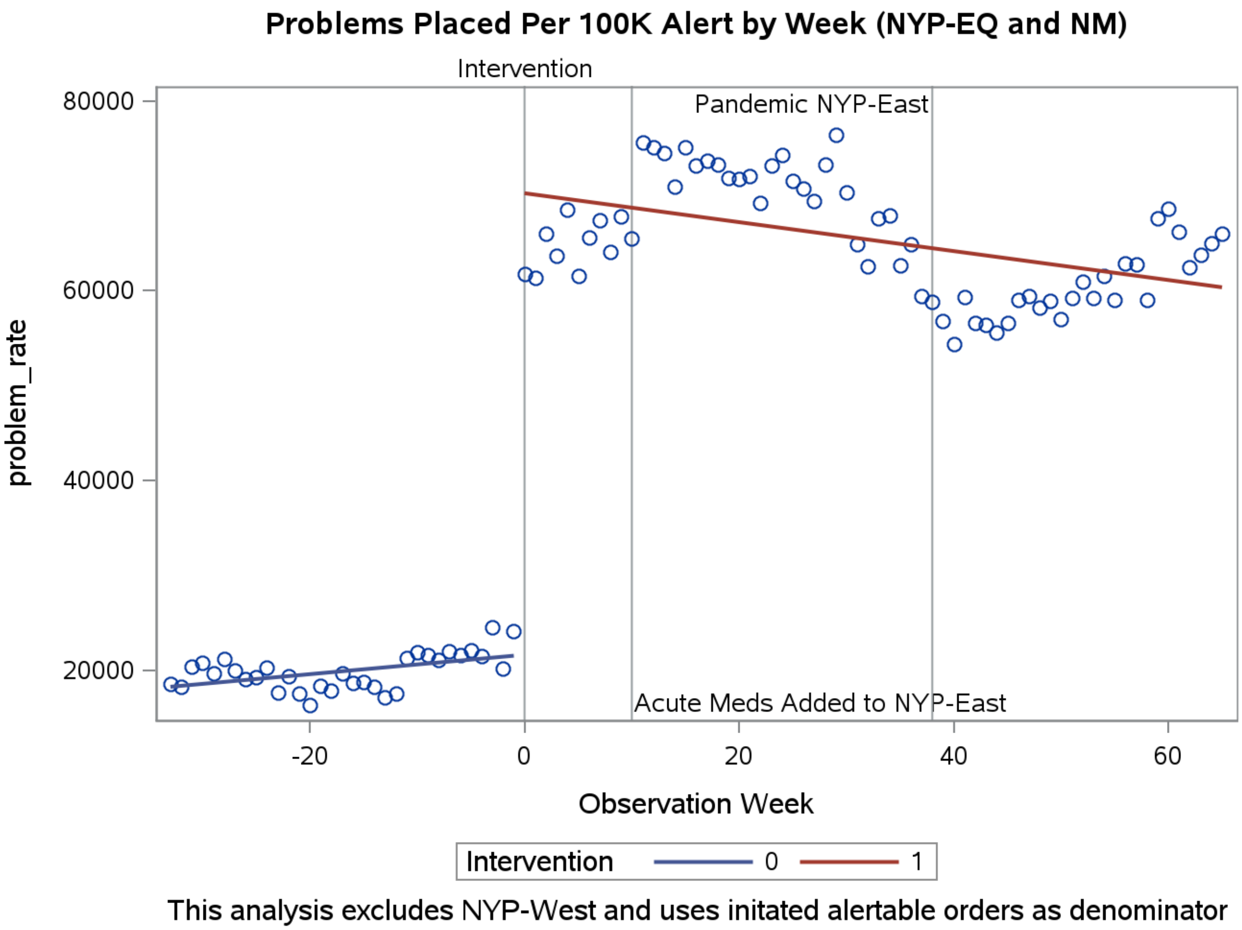


Figure 5. Wrong-Drug Abandon and Reorder per 100k initiated alertable study drug orders.

Table 6. Parameter estimates from regression of wrong-drug abandon and reorder rate on time, intervention, and time after intervention.

Label	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	384.14500	73.20175	5.25	<.0001
Elapsed Time (Weeks)	-2.88711	1.96578	-1.47	0.1452
Time After Intervention (Weeks)	3.82532	2.08495	1.83	0.0697
Intervention	107.44855	45.34888	2.37	0.0198

Problem Placement. Figure 6 shows the relationship between the rate of problem placement (the number of problems placed after alerts divided by the number of initiated alertable study drug orders), before and after the intervention (represented by the vertical line). The effect of time was not significant (b=102.33, t=1.78 p=0.08). The effect of the intervention was significant (b=42,893, t=28.91, p<.0001). The effect of time after the intervention was not significant (b=-28.401, t=-0.37 p=0.71). Table 7 gives the parameter estimates for the segmented regression analysis. The overall fit of the model was R^2=0.98, F(3, 93)=956.5, p<.0001.



the

Figure 6. Problems placed per 100k initiated alertable study drug orders.

Table 7. Parameter estimates from regression of problem placement rate on time, intervention, and time after intervention.

Label	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	16189	2135.24962	7.58	<.0001
Elapsed Time (Weeks)	102.33466	57.34065	1.78	0.0776
Time After Intervention (Weeks)	-28.40187	76.69353	-0.37	0.7120
Intervention	42893	1483.92869	28.91	<.0001
Acute Meds Added to NYP-East	4916.23060	1484.10275	3.31	0.0013
Pandemic Starts at NYP-East	-12598	1655.27656	-7.61	<.0001

Discussion

Problem Placement

The intervention was, as expected, very effective at increasing the rate of problem placement. This is not surprising because it was the original purpose of these alerts. They prompt the prescriber to add a problem when a medication is ordered in the absence of a justifying problem on the problem list. The intervention more than tripled the rate of problem placement, from just under 20,000 per 100,000 orders to more than 65,000 per

100,000 orders. To the extent that a health system values complete and accurate problem lists, indication alerts during medication ordering are an effective strategy. The strategy is limited in only working for medications with narrow sets of indications.

Wrong Patient Retract and Reorder (WP-RAR)

The intervention had no effect on WP-RAR events. We predicted a small decrease, but it did not occur. We predicted that RAR events would decrease because potential RAR events would have been caught prior to signing the order and manifest as AAR events instead. Although we did see the increase in AAR events (described below), we did not see any corresponding increase in RAR events. It may be that indication alerts only have a measurable impact, or at least their primary impact, on abandonment rather than retraction events.

Wrong Drug Retract and Reorder (WD-RAR)

Indication alerts had no immediate effect on WD-RAR events, but these events did decline slowly as time passed after the intervention. We predicted these events would decrease in frequency as abandonment events increased, although we thought this would occur immediately rather than slowly over time. When combined with the null effect of indication alerts on WP-RAR events, it supports the idea that these alerts function primarily to increase the rate of abandon-and-reorder events, with only small or non-existent effects on retract-and-reorder events. It may be that abandonments and retractions have different causes and that an intervention that affects one may not affect the other.

Wrong Patient Abandon and Reorder (WP-AAR)

The intervention increased WP-AAR from 156 to 276 per 100k orders (a 77% relative increase). This supports our hypothesis that indication alerts will increase abandonment by increasing situation awareness about the patient-drug-indication relationship at the moment of ordering. We interpret this as evidence of patient safety improvement because each abandonment is a self-intercepted medication error.

Wrong Drug Abandon and Reorder (WD-AAR)

Indication alerts also increase wrong drug abandon and reorder events, from 280 to 373 per 100k orders (a 33% relative increase). This is also consistent with our hypothesis and supports the idea that the same causal mechanism is at work (i.e., increased situation awareness).

Additional Analyses and Future Plans

One weakness of the interrupted time series analysis is that it aggregates events across orders and days when it tallies event totals per week. This prevents us from studying the factors that vary at the level of the individual order, such as patient and prescriber demographics, clock time (i.e., shift), clinical location, etc. We are now attempting to obtain all the data from our study sites at the level of the individual order. This will allow us to build logistic regression models of each outcome (AAR and RAR events and problem placement) with the standard interrupted time series covariates (e.g., time, intervention, and time after intervention) while also controlling for and examining the effect of these other covariates. We will also be able to account for clustering within prescribers by using mixed effects regression models.

Conclusions

Indication alerts produce large increases in the rate of problem placement, and at the same time they increase the rate of abandon-and-reorder events, with only small or non-existent effects on retract-and-reorder events. We posit that the alerts work by orienting the prescriber to the patient's clinical problems, thereby increasing situation awareness at the very moment of prescribing. The increased situation awareness enhances the ability to self-intercept errors, and hence self-interceptions increase. We do not have direct evidence of prescribers' mental states, but clinical experience with these alerts suggests this as the most plausible mechanism, and the fact that they worked as intended in this study adds to our confidence that this is the correct model of what is happening at the human-computer interface when these alerts fire. It is not trivial to implement these alerts. Even if we share our clinical logic and code, most health systems will want to customize update the alerts to suit local preferences, increasing the implementation burden. Also, the alerts cause some nuisance, requiring extra mouse clicks to order common medications. Decision makers will have to weigh the costs (e.g., implementation burden and nuisance) against the benefits (e.g., more complete problem lists and more self-intercepted errors). Because it is relatively rare for a single kind of clinical decision support to have the dual benefits of improved clinical documentation and increased error interception, we recommend that health systems with the necessary capabilities seriously consider implementing indication alerts.

Significance

Limitations. Readers should interpret these results considering several limitations. We studied only two health systems, and although NM included some community hospitals, both systems consist primarily of large, urban, academic medical centers which are, in many ways, not representative of many hospitals in the US. At NM, to minimize nuisance, alerts did not fire on refill medication orders or on orders in order sets or in urgent care or the emergency department. And they only fired for attending physicians. At Columbia/NYP, alerts only fired on inpatient medication orders that were not part of order sets. In addition, the alerts were non-interruptive at NYP and interruptive at NM. The list of study drugs was limited to those that had only a few, well-defined indications and that the health systems deemed to be useful, resulting in n=206 study drugs at NM and n=87 at Columbia/NYP. The results reported here may not generalize to other drugs, especially to those with many indications. To measure outcomes in precisely the same manner in both the pre- and post-intervention periods (since alerts did not exist in the pre-intervention period, there was no pre-intervention alert log), we wrote specialized software to simulate the alert logic and ran that simulation code on data extracted from each system's electronic data warehouse (EDW). Differences between production data and EDW data tables prevented us from precisely simulating the data available to the alert logic at the actual time of order entry, leading to some measurement error when comparing actual alert logs from the post-intervention period to the simulated alert logs generated by our software. While the measurement error introduced by discrepancies between production data and EDW data tables may have impacted the accuracy of our simulated alert logs, it is crucial to note that this error was consistent across both the pre- and post-intervention periods. Therefore, although this error may have introduced some degree of uncertainty into our results, it should not have biased our estimates of the intervention's effect. Nevertheless, the presence of this error may have attenuated our estimates of the relationships between variables and could have potentially reduced the statistical power of our analyses, making it more challenging to detect true effects. Furthermore, the error may have affected the validity of some of our statistical assumptions. Future research should aim to minimize such measurement errors to enhance the accuracy and reliability of the findings. Also, the presence of an abandon-and-reorder or retract-and-reorder event is far from a perfect predictor of a clinician-confirmed error. Whereas Adelman and colleagues report WP-RAR to have a good positive predictive value for identifying real errors (roughly 80%), the same is not true for WD-RAR, WP-AAR, and WD-AAR whose PPV for confirmed error is low (on the order of 25%) for WP-AAR, and yet undetermined for wrong-drug retraction and abandonment.

Impact on Priority Populations. We have not analyzed our results at the individual patient level, so we cannot comment on any differential effects our interventions may have had as a function of race, gender, ethnicity, or any other characteristic that defines AHRQ priority populations. However, our results indicate that the intervention had beneficial effects on problem placement and on the self-interception of medication errors. Since the project was done at two large, multihospital urban health systems that serve highly diverse patient populations, the benefits of our interventions would have accrued to members of priority populations living in the catchment areas of these health systems.

Implications

Indication alerts during CPOE are a highly effective method for increasing the placement of accurate problems when drugs with narrow indications are prescribed. They also increase the rate of abandon-and-reorder events, which we take to be a kind of self-intercepted medication error. Their effect on retraction events is much smaller and limited to wrong-drug retractions. Institutions that want to increase problem placement should consider implementing indication alerts, though the benefits must be weighed against the implementation burden, opportunity costs, and nuisance.

List of Publications and Products

Grauer A, Kneifati-Hayek J, Reuland B, Applebaum JR, Adelman JS, Green RA, Lisak-Phillips J, Liebovitz D, Byrd TF, Kansal P, Wilkes C, Falck S, Larson C, Shilka J, VanDril E, Schiff GD, Galanter WL, Lambert BL. Indication alerts to improve problem list documentation. J Am Med Inform Assoc. 2022 Apr 13;29(5):909-917. doi: 10.1093/jamia/ocab285. PMID: 34957491; PMCID: PMC9006708.

References

1. Pham JC, Story JL, Hicks RW, et al. National study on the frequency, types, causes, and consequences of voluntarily reported emergency department medication errors. *J Emerg Med.* 2011;40(5):485-492.
2. Latif A, Rawat N, Pustavoitau A, Pronovost PJ, Pham JC. National study on the distribution, causes, and consequences of voluntarily reported medication errors between the ICU and non-ICU settings. *Crit Care Med.* 2013;41(2):389-398.
3. Cina JL, Gandhi TK, Churchill W, et al. How many hospital pharmacy medication dispensing errors go undetected? *Joint Commission Journal on Quality and Patient Safety.* 2006;32(2):73-80.
4. Kopp BJ, Erstad BL, Allen ME, Theodorou AA, Priestley G. Medication errors and adverse drug events in an intensive care unit: direct observation approach for detection. *Crit Care Med.* 2006;34(2):415-425.
5. Al-Dhawailie AA. Inpatient prescribing errors and pharmacist intervention at a teaching hospital in Saudi Arabia. *Saudi Pharm J.* 2011;19(3):193-196.
6. Tully MP, Buchan IE. Prescribing errors during hospital inpatient care: factors influencing identification by pharmacists. *Pharm World Sci.* 2009;31(6):682-688.
7. Vessal G. Detection of prescription errors by a unit-based clinical pharmacist in a nephrology ward. *Pharm World Sci.* 2010;32(1):59-65.
8. Al-Jeraisy MI, Alanazi MQ, Abolfotouh MA. Medication prescribing errors in a pediatric inpatient tertiary care setting in Saudi Arabia. *BMC Res Notes.* 2011;4:294.
9. Westbrook JI, Baysari MT, Li L, Burke R, Richardson KL, Day RO. The safety of electronic prescribing: manifestations, mechanisms, and rates of system-related errors associated with two commercial systems in hospitals. *J Am Med Inform Assoc.* 2013;20(6):1159-1167.
10. Galanter WL, Bryson ML, Falck S, et al. Indication alerts intercept drug name confusion errors during computerized entry of medication orders. *PLoS ONE.* 2014;9(7):e101977.
11. Koppel R, Metlay JP, Cohen A, et al. Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors. *JAMA.* 2005;293(10):1197-1203.
12. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc.* 2006;13(5):547-556.
13. Gurwitz JH, Field TS, Harrold LR, et al.e ambulatory setting. *Jama.* Mar 5 2003;289(9):1107-1116.
14. Ash JS, Gorman PN, Hersh WR, Lavelle M, Poulsen SB. Perceptions of house officers who use physician order entry. *Proceedings / AMIA Annual Symposium.* 1999:471-475.
15. Sopan A, Plaisant C, Powsner S, Shneiderman B. Reducing Wrong Patient Selection Errors: Exploring the Design Space of User Interface Techniques. *AMIA Annual Symposium Proceedings.* 2014;2014:1056-1065.
16. Adelman J, Aschner J, Schechter C, et al. Use of Temporary Names for Newborns and Associated Risks. *Pediatrics.* 2015;136(2):327-333.
17. Adelman JS, Kalkut GE, Schechter CB, et al. Understanding and preventing wrong-patient electronic orders: a randomized controlled trial. *J Am Med Inform Assoc.* 2013;20(2):305-310.
18. Hyman D, Laire M, Redmond D, Kaplan DW. The use of patient pictures and verification screens to reduce computerized provider order entry errors. *Pediatrics.* 2012;130(1):e211-219.
19. Green RA, Hripcsak G, Salmasian H, et al. Intercepting wrong-patient orders in a computerized provider order entry system. *Ann Emerg Med.* 2015;65(6):679-686 e671.
20. Wilcox AB, Chen Y-H, Hripcsak G. Minimizing electronic health record patient-note mismatches. *Journal of the American Medical Informatics Association : JAMIA.* 2011;18(4):511-514.
21. Institute for safe Medication Practices. Oops, sorry, wrong patient! A patient verification process is needed everywhere, not just at the bedside. .

http://www.ismp.org/Newsletters/acutecare/articles/20110310.asp Published 2011. Accessed January 18, 2016.
22. Grauer A, Kneifati-Hayek J, Reuland B, et al. Indication alerts to improve problem list documentation. *J Am Med Inform Assoc.* 2022;29(5):909-917.
23. Galanter W, Falck S, Burns M, Laragh M, Lambert BL. Indication-based prescribing prevents wrong-patient medication errors in computerized provider order entry (CPOE). *J Am Med Inform Assoc.* 2013;20(3):477-481.
24. National Quality Forum. Patient Safety 2015, draft report for comment. http://www.qualityforum.org/Projects/n-r/Patient_Safety_Measures_2015/Draft_Report_for_Comment.aspx. Published 2015. Accessed January 18, 2016.
25. Kondrak G, Dorr B, J. Automatic identification of confusable drug names. *Artif Intell Med.* 2006;36:29-42.
26. Lambert BL, Lin S-J, Gandhi SK, Chang K-Y. Similarity as a risk factor in drug name confusion errors: The look-alike (orthographic) and sound-alike (phonological) model. *Med Care.* 1999;37(12):1214-1225.
27. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy & Therapeutics.* 2002;27(4):299-309.