A Metrics Taxonomy and Reporting Strategy for Rule-Based Alerts

Michael Krall, MD, MS; Alexander Gerace

ABSTRACT

Context: Because institutions rely on rule-based alerts as an important component of their safety and quality strategies, they should determine whether the alerts achieve the expected benefit.

Objective: To develop and to test a method of reporting outcome metrics for rule-based electronic health record alerts on a large scale.

Methods: We empirically developed an action-oriented alerts taxonomy according to structure, actions, and implicit intended process outcomes using a set of 333 rule-based alerts at Kaiser Permanente Northwest. Next we developed a method for producing metrics reports for alert classes. Finally, we applied this method to alert taxa.

Main Outcome Measures: Outcome measures were the successful development of a rule-based alerts taxonomy and the demonstration of its application in a reporting strategy.

Results: We identified 9 major and 17 overall classes of alerts. We developed a specific metric approach for 5 of these classes, including the 3 most numerous ones in our institution, accounting for 224 (67%) of our alerts. Some alert classes do not readily lend themselves to this approach.

Conclusions: We developed a taxonomy for rule-based alerts and demonstrated its application in developing outcome metrics reports on a large scale. This information allows tuning or retiring alerts and may inform the need to develop complementary or alternative approaches to address organizational imperatives. A method that assigns alerts to classes each amenable to a particular reporting strategy could reduce the difficulty of producing metrics reports.

INTRODUCTION

Rule-based alerting within electronic health records (EHRs) is a common approach to addressing safety, quality, and workflow issues in health care. Most mature EHR installations have alerts of this type and some, including ours, have hundreds. Stage 2 of the Center for Medicare and Medicaid Services Meaningful Use incentive program requires that eligible providers and hospitals implement at least five clinical decision-support interventions, most of which will be rule-based alerts. Because institutions look to clinical decision support and rule-based alerts as an important component of their return-on-investment strategy for their EHR investment, it is imperative to

know whether the alerts are not only "working" (firing as designed) but are achieving intended benefits.[1] This understanding is fundamental to improving clinical decision support and may suggest the necessity for considering additional or alternative strategies to achieve strategic goals. A recent publication of the American Health Lawyers Association on minimizing EHR-related safety events specifically called out alert metrics beyond simply override rate as an important determinant of safety in these systems.[2]

Knowledge engineers and builders of alerts typically believe, often with limited or no data, that the ones they deploy are functioning well, are well received by their target audience, and are achieving

their intended goals. Tools within EHRs to evaluate these beliefs are generally very limited. Alert structures can be complex, and achieving an understanding of their performance characteristics and effectiveness can prove difficult. To develop a detailed understanding of the functioning and outcomes of an alert usually requires creating an individualized report. There are few reports of systematic approaches to monitoring alerts.[3] We set out to develop one.

Objective

Our goal was to develop and test a framework for reporting performance metrics for rule-based EHR alerts at scale. Our intention was to achieve alert outcome reports that extend beyond basic metrics and include process outcomes. As a preliminary step we set out to develop a new action-oriented classification system or taxonomy of alerts on the basis of their structure, especially intended and optimized to facilitate outcomes metrics. Existing taxonomies would not suffice because they do not readily map to a measurement framework.[4-8] We sought to generate metrics in sets rather than one by one to reduce the time and expense of developing and maintaining performance metrics for alerts.

Background

In a proposed five-level evaluation hierarchy of rule-based alert performance, "firing rates" are at the lowest and most commonly reported level. Rates by role such as nurse or physician, specialty such as family medicine or general surgery, site of care such as inpatient or ambulatory, department or unit such as intensive care or urgent care, time of day

Michael A Krall, MD, MS, is a part-time Physician Lead for Clinical Decision Support for The Permanente Federation. He retired from Northwest Permanente in 2014 after 30 years practicing Family Medicine and after 20 years engaged in clinical informatics. Dr Krall was the Director of Clinical Decision Support and Knowledge Management for Northwest Permanente. E-mail: michael.krall@kp.org. **Alexander Gerace** is an Information Analyst for Care Delivery Analytics for Northwest Permanente. E-mail: alexander.x.gerace@kp.org.

or week, and similar factors can extend this metric. The second level is "acceptance" and "override" rates. Proximate "action taken" by users in response to the alert is the next highest level. The fourth level is a measure of intermediate process goals or outcomes achievement. Finally, the highest level is a measure of patient health goals or outcomes or clinician goals or outcomes achievement. Each successive level in the evaluation hierarchy is more difficult to create. In this report we chose to focus on the fourth level, process goals and outcomes.

For most alerts, knowing that they fire frequently or infrequently and have a high or low acceptance rate yields only a limited understanding of whether their explicit recommendations were followed or if they were effective in achieving their intended goals. Was the safe practice followed or the unsafe practice averted? Was the evidence-based treatment prescribed? Was the cost-effective choice made, the recommended documentation performed, the dose modified, the procedure ordered or performed? These are "level 4" questions in the hierarchy introduced above. Ultimately, was care improved and the desired health or clinician outcome achieved? These are "level 5" questions. For most alerts, firing, override, and proximate action-taken rates do not answer these higher-order questions because action on the alert may not be the same as action on the triggering event.

In EHRs, alerts may be built with a number of available user actions. In the Epic 2012 and earlier versions (Madison, WI) EHR, for example, standard controls include "accept" and "cancel" buttons that take these actions on the alert. However, there may be additional actions such as "open an order set," "create an order," or "navigate" to the Problem, Allergy, or Medication List activities to perform an operation, and such actions are intended to address the triggering condition. It is possible for users to "cancel" an alert yet still perform the recommended operation, such as order a laboratory procedure. Alternatively it is possible to "accept" the alert but fail to perform the operation. For example, the user can open an order set attached to an alert and choose

not to sign the recommended order, or can navigate to an allergy or medication list and decide not to perform the recommended update. Thus reports that simply count these actions may generate inaccurate and misleading information about alert effectiveness and desired outcomes. In the end, what we want to know is whether the recommended operation was performed, regardless of whether it was done as a direct result of the alert.

A reporting strategy that disengages the outcome from a direct action of the alert creates the additional complexity of determining the optimal time frame for measurement. The measurement interval could be defined as immediately proximate to the alert firing (or viewing, in the case of alerts that do not pop up), as the less stringent "any time within the encounter" definition, or as another predefined interval (such as within eight hours of firing). In the first instance, "credit" is assigned to the alert only if the recommended action occurred as a direct result of or immediately following the alert. An order or action fulfilling the recommendation that was placed later in the same encounter would not count as alert "success." In the second instance, any fulfilling order or action that was taken within the same encounter would count, even though it is not certain that the action taken was directly caused by the alert. In the inpatient setting, where the entire stay might be considered the encounter, this definition could cause difficulty in interpretation. Here the third approach, crediting actions that occurred within a specified time window, such as within an eight-hour shift, might be more meaningful. Conceptually, this requires an approach to metrics similar to an intention-to-treat analysis of a randomized control study.

Complete assessment of the appropriateness of a specific alert, the user's response to that alert, or level five patient or clinician outcomes requires a broader analysis to include patient, user, and environmental factors. Because an automated batch processing of alerts cannot achieve that level of investigation, a more realistic goal is to develop a report that would allow categorizing alerts into those appearing to

be "high," "low," and "intermediate" in performance. An institution might then, for example, choose to direct its efforts at improving, eliminating, or better understanding those alerts with "low" performance. To do so might require a more in-depth assessment, possibly including medical chart reviews. McCoy et al[9] published a framework for evaluating alert appropriateness.

To develop a detailed model of alert metrics, it is necessary to have a common understanding of the structure or "anatomy" of an alert. There have been a number of clinical decision support and alert taxonomies developed for different purposes.[4-8] In 2007, Wright et al[7] at Partners Healthcare System developed a taxonomy for rule-based decision support. They included four functional components they termed *triggers, input data, interventions,* and *offered choices.* Later, the National Quality Forum Clinical Decision Support Expert Panel[8] proposed a modification of this classification. They replaced "offered choices" with "action steps." Triggers are the user actions that can invoke an alert. These actions might include, for example, entering or signing an order, opening a chart or a specific screen, or entering documentation. Input data are the elements in the record that might modify the alert performance. Input data might be information about the patient (eg, age, gender, diagnoses, preferences),

---

National Quality Forum Taxonomy: Selected EpicCare Examples

Triggers
  Open chart, open order entry
  Enter order, enter diagnosis
  File flow sheet

Input data
  Age, gender, comorbidities
  Provider type, role
  Encounter type

Interventions
  Display message in Navigator
  Send in-basket message
  Add modifier
  Pop-up message

Action steps
  Accept or Cancel alert
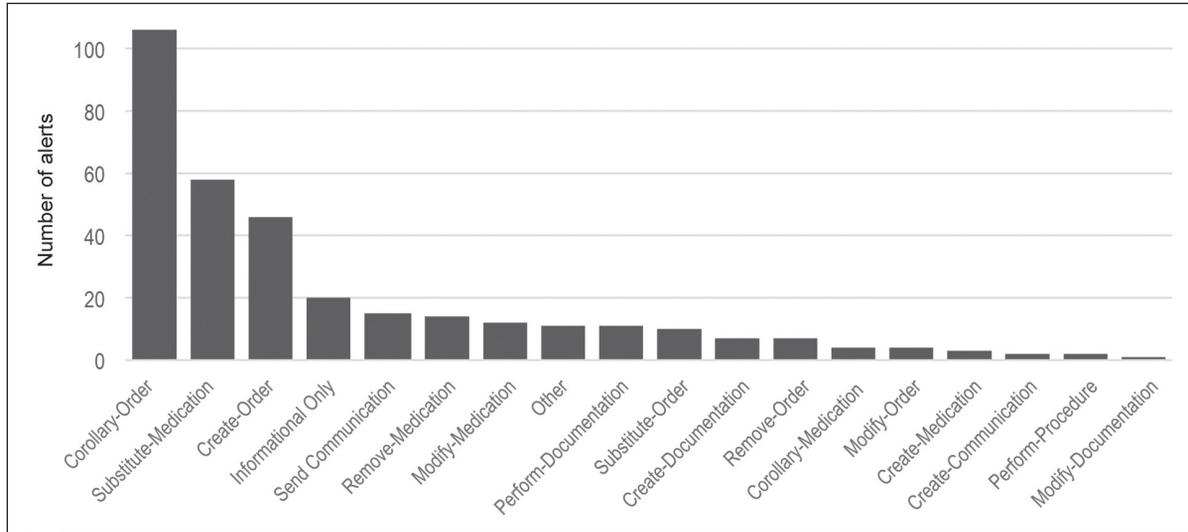 Enter order
  Open order set

Figure 1. Distribution of alerts at Kaiser Permanente Northwest by class, 2013.

about the user (eg, specialty, role, experience, preferences), or about the environment (eg, time of day or week, unit, department, or setting). Interventions refer to computer-human interface actions and the manner in which the alert is displayed, such as via a pop-up message. In some cases the intervention might happen without user awareness, such as if the alert action is to set a modifier or to trigger an asynchronous alert (occurring at a remote time or place). Finally, action steps are recommended or permitted actions that a user can take as a direct result of the alert. Accepting, canceling, and overriding (with or without providing a reason) are the most common actions. Many more actions may also be available in alerts, including opening order sets, accepting an order, and navigating to another activity such as the medication, allergy, or problem list. The Sidebar: National Quality Forum Taxonomy: Selected EpicCare Examples illustrates these structural elements with a few examples from the EpicCare EHR. Most other EHR alerts have similar structure and comparable examples.

Combining available triggers, inputs, interventions, and action steps yields a very large number of possible alert forms, particularly since an alert may incorporate more than one element for each functional component. For example, one alert could have an enter-diagnosis and enter-order trigger, multiple data inputs, and both a pop-up display and an in-basket notification. Alerts may also have more than one permissible or recommended action step. This complexity is a major reason why it is difficult to create generalized alert outcome reports. For each alert it is necessary both to define its structure and to declare which actions constitute a desired outcome or "success." Previous taxonomies are not classified by actions taken (such as create or modify an order) and do not attempt to address alert outcomes.

METHODS

This work was developed at Kaiser Permanente Northwest (KPNW), a Region of the Kaiser Permanente Health Care program located in Oregon and Southwest Washington. The Region employs approximately 1000 physicians and cares for approximately 500,000 Health Plan members. KPNW began implementation of EpicCare in 1994, and it was fully implemented in ambulatory clinics by 1997. The inpatient system was implemented in 2008. KPNW owns and operates 2 hospitals and about 40 outpatient clinic sites, most of which are large and multispecialty.

We took an empirical approach to developing the alert metrics taxonomy. Each of 333 active, production rule-based alerts in our system was assigned to a class on the basis of its structure and intended outcome. If an existing class did not adequately encompass the alert in question, a new class was added or an existing class was modified. This process was iteratively repeated until all alerts were assigned and the remaining classes were coherent. Where alerts had characteristics of more than one class, they were assigned on the basis of their apparent primary intention. For example, an alert that recommended substitution of a particular medication for another (a substitution action) but also facilitated documentation of allergy to the recommended choice via a link to the allergy activity (a documentation action) was classified according to the primary intention of substitution. In order to limit project scope, standard, vendor-supplied drug-drug and drug-allergy alerts were excluded from this analysis. This was deemed appropriate for several reasons. In the Epic EHR such alerts are presented with a different utility. Although they do conform to the same four-element National Quality Forum functional taxonomy, they have a different data and reporting structure and are very numerous. Moreover, for these alerts "acceptance" and "override" rates are more useful metrics because the valid action steps are fewer and more straightforward—the user either accepts or overrides the recommendation to avoid the drug-drug, drug-allergy, or drug-condition combination. Admittedly, that analysis would necessarily
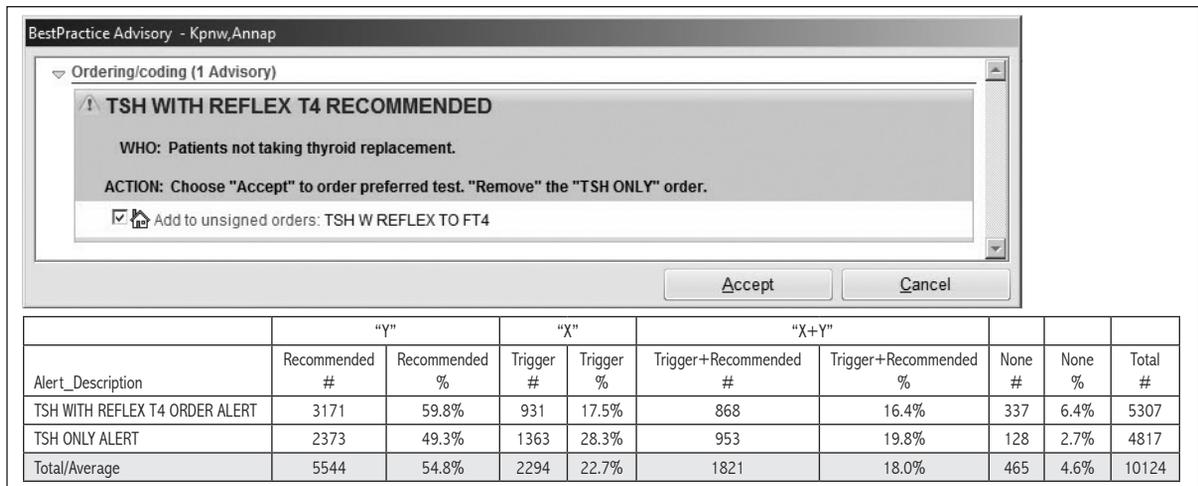
Figure 2. Substitute Order Alert with example report: "X" is the procedure that triggered the best-practice alert, "Y" the recommended substitute order, "X+Y" represents encounters in which both tests were ordered and "none" in which neither test was ordered. © 2014 Epic Systems Corporation. Used with permission.

FT4 = free thyroid; T4 = thyroid hormone; TSH = thyroid stimulating hormone; W = with.

	"Y"		"X"		"X+Y"				
Alert_Description	Recommended #	Recommended %	Trigger #	Trigger %	Trigger+Recommended #	Trigger+Recommended %	None #	None %	Total #
TSH WITH REFLEX T4 ORDER ALERT	3171	59.8%	931	17.5%	868	16.4%	337	6.4%	5307
TSH ONLY ALERT	2373	49.3%	1363	28.3%	953	19.8%	128	2.7%	4817
Total/Average	5544	54.8%	2294	22.7%	1821	18.0%	465	4.6%	10124

be more complex if one considered the possible override reasons that may be required or optionally provided or the alternative actions such as updating medication or allergy lists.

Following development of an initial taxonomy, we turned attention to designing a report creation methodology. Starting with the classes with the largest number of members, we examined a representative alert from each class. For each, we developed a formulaic description of the triggering event and recommended outcome. Then we used the data model and data in Clarity, Epic's analytical database, to create a report comparing initial with final conditions to determine alert outcome. We started with the appropriate trigger table in the reporting data warehouse. For example, we used one table for procedure triggers and another for generic medications. Next, we created a subquery that defines the recommended result, such as a substituted or additional procedure or medication, for each alert in the given class. This recommended result might have been contained, for example, in an order set or in an order attached directly to the alert. We joined this table to the triggered alert by its identifier, the alert locator ID. Next we created a subquery to determine whether the alert trigger and/or the recommended result was ordered or created in a valid time

interval, as discussed above. For ambulatory alerts we defined this interval as within the particular patient encounter, usually an office visit or telephone call. Finally, we created a summary report.

RESULTS

Each alert is unique, owing to its complex structure and specific clinical or operational intention. Nevertheless, through empirical analysis, it appears that there are a manageable number of major structural classes for alert metrics, and these constitute the proposed taxonomy (see Sidebar: Alert Metrics Taxonomy). Each of these classes has a generalized structure that can be expressed in narrative. Substitution alerts, for example, are those in which a clinician orders a test, procedure, or treatment, or enters a diagnosis or finding and the alert recommends a second test, procedure, or finding instead, such as "Replace the order for 'Chest X-ray AP/LAT' with an order for 'Chest X-ray two-views.'" The substitutions are usually of the same type (eg, test for test or diagnosis for diagnosis) but could be of mixed type (eg, medication for procedure). In this case, a trigger "X" causes the alert to recommend substituting an item or action "Y." Corollary-type alerts are similar.[10] In these cases, however, a trigger "X" generates a recommendation of an additional "Y." For example, "In addition

to the order for digoxin, please order a serum potassium and creatinine." Corollary recommendations may be single or multiple ("Y" and "Y^1" and "Y^n"), and like substitution alerts, the recommendations may be of the same or mixed type.

Alert Metrics Taxonomy

Substitution
Substitute-Order
Substitute-Medication
Substitute-Documentation

Corollary
Corollary-Order
Corollary-Medication
Corollary-Documentation

Modify
Modify-Order
Modify-Medication
Modify-Documentation

Create
Create-Order
Create-Medication
Create-Documentation
Create-Communication

Remove
Remove-Order
Remove-Medication

Perform
Perform-Procedure
Perform-Documentation

Informational only

Send communication

Other

The other alert classes follow similar patterns. Modify-type alerts recommend changes to new or existing orders or documentation—for example, "Decrease the dosage of digoxin on the basis of the patient's serum creatinine." Create alerts recommend new orders or documentation, such as "The patient is due for a mammogram. Please sign the attached mammogram order." Remove alerts recommend that existing orders or documentation are canceled—for example, "It appears that the patient

no longer requires a follow-up CT [computed tomography] scan. If this is correct, please cancel this order." Perform alerts recommend the execution of specific actions, such as "Because of the elevated initial blood pressure, please repeat a blood pressure in 5 minutes and document in the chart." Informational Only alerts provide a message without the expectation of a particular action that is captured within the system—for example, "There is a regionwide shortage of influenza vaccine." Alerts might

send communication through a variety of means including in-basket messages or messages to pagers, faxes, phones, or Internet-enabled devices. Because this taxonomy is not exhaustive, there is necessarily an Other category. In our analysis, the alerts that fell into this category were "one-offs" and did not lend themselves to additional characterization at this time. An example of such an alert is a single one that automatically adds a modifier or marker to the patient's chart if he is a man older

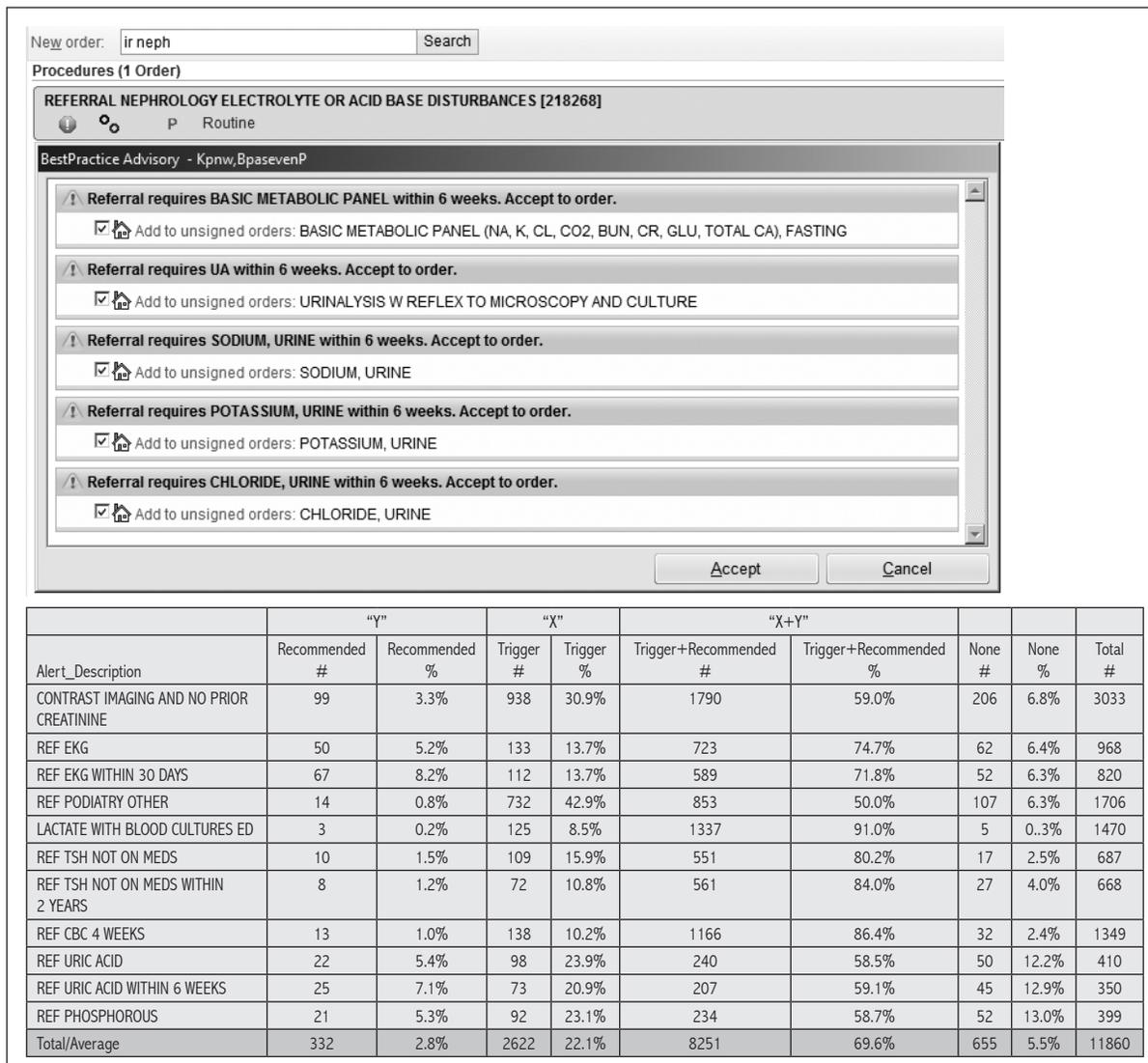| Alert_Description	"Y" Recommended #	"Y" Recommended %	"X" Trigger #	"X" Trigger %	"X+Y" Trigger+Recommended #	"X+Y" Trigger+Recommended %	None #	None %	Total #
CONTRAST IMAGING AND NO PRIOR CREATININE	99	3.3%	938	30.9%	1790	59.0%	206	6.8%	3033
REF EKG	50	5.2%	133	13.7%	723	74.7%	62	6.4%	968
REF EKG WITHIN 30 DAYS	67	8.2%	112	13.7%	589	71.8%	52	6.3%	820
REF PODIATRY OTHER	14	0.8%	732	42.9%	853	50.0%	107	6.3%	1706
LACTATE WITH BLOOD CULTURES ED	3	0.2%	125	8.5%	1337	91.0%	5	0..3%	1470
REF TSH NOT ON MEDS	10	1.5%	109	15.9%	551	80.2%	17	2.5%	687
REF TSH NOT ON MEDS WITHIN 2 YEARS	8	1.2%	72	10.8%	561	84.0%	27	4.0%	668
REF CBC 4 WEEKS	13	1.0%	138	10.2%	1166	86.4%	32	2.4%	1349
REF URIC ACID	22	5.4%	98	23.9%	240	58.5%	50	12.2%	410
REF URIC ACID WITHIN 6 WEEKS	25	7.1%	73	20.9%	207	59.1%	45	12.9%	350
REF PHOSPHOROUS	21	5.3%	92	23.1%	234	58.7%	52	13.0%	399
Total/Average	332	2.8%	2622	22.1%	8251	69.6%	655	5.5%	11860

Figure 3. Corollary-type alerts and report. A referral to nephrology for evaluation of acid-base disturbance generated these alerts for tests recommended before the referral that have not been performed. In the report, "X" is the procedure that triggered the best-practice alert, "Y" the recommended additional order(s); "X+Y" represents encounters in which both tests were ordered: the desired result in this case. © 2014 Epic Systems Corporation. Used with permission.

BUN = blood urea nitrogen; CA = calcium; CBC = complete blood count; CL = chloride; CO2 = carbon dioxide; CR = creatinine; ED = Emergency Department; EKG = electrocardiogram; GLU = glucose; K = potassium; KPNW = Kaiser Permanente Northwest; NA = sodium; REF = referral; TSH = thyroid stimulating hormone; UA = urinalysis.

than age 65 years who has documentation that he is a current or ever-smoker. This marker, in turn, is used by a second alert that recommends screening for abdominal aortic aneurysm. We elected not to create new classes with only single members in our data set. The Sidebar: Alert Metrics Taxonomy also contains subcategories. More granular specification is important because the reporting logic differs for procedures, medications, and documentation.

We assigned each of 333 KPNW active, in-production, rule-based alerts to a class. Some alerts recommend or allow more than one action, such as substitute an order and perform documentation. As noted, for the purposes of report creation, simplification, and this initial stage of development, each was assigned to one class only on the basis of their primary or most impactful recommendation. In our system at this time, 3 of 18 categories accounted for more than 60% of all alerts, with Corollary Order, Substitute Medication, and Create Order contributing 106, 58, and 46 alerts, respectively (Figure 1).

Categorizing alerts by class allowed us to develop a reporting strategy that works for all members of that class, or at least for those members who do not deviate too much from a prototypical class representative. The Substitution Order class offers an example of our approach. In this case, a user orders test or procedure "X" and the alert recommends substituting test or procedure "Y" (Figure 2). Regardless of whether the user accepted or canceled the alert, or opened or did not open an attached order set if present, what we want to know is, at the end of
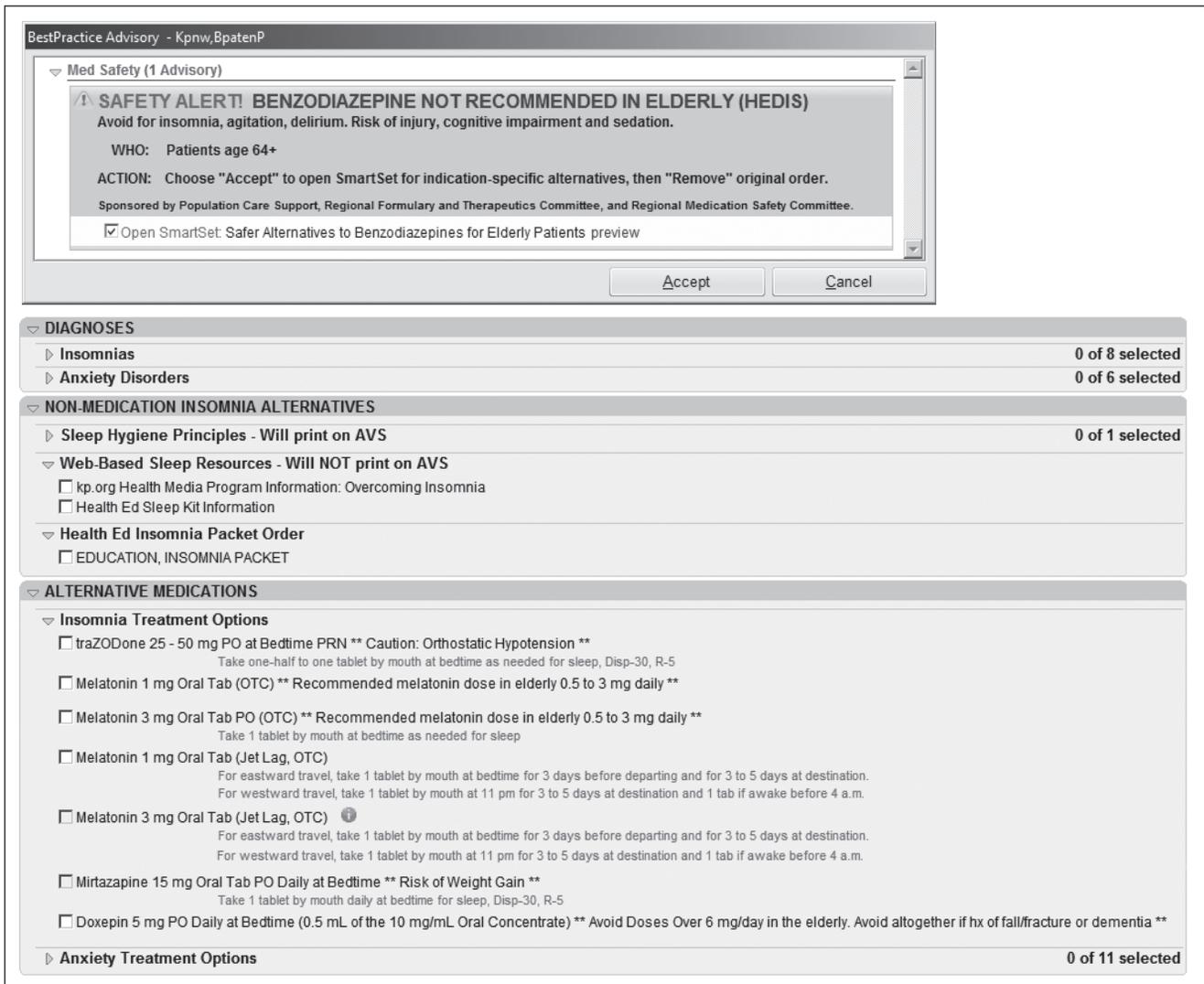

Figure 4. Substitute Medication Alert and attached ambulatory order set. A benzodiazepine drug was ordered in a patient older than age 64, and a substitution is recommended and facilitated via the attached ambulatory order set containing condition-specific recommended alternatives. Note that there are both pharmacologic and nonpharmacologic alternatives. The anxiety treatment options are in an expandable group. Each of these choices is contained in a separate group or item within the database. © 2014 Epic Systems Corporation. Used with permission.

AVS = after-visit summary; Disp-30 = dispense 30; Hx=history of; HEDIS = Healthcare Effectiveness Data and Information Set; OTC = over the counter; PO = by mouth; PRN = when necessary; R-5 = refill 5 (ie, 5 refills).

the measurement period (eg, encounter), which was ordered, "X" or "Y"? (Ordering both is also possible, which would usually not be a desired result, as is ordering neither, which might or might not be desirable depending on the situation.) What we need is a program that can determine for all alerts of this type, which order(s) were present in the "result set" at the end of the encounter. Figure 2 also shows an example of a report created for 2 alerts of the Substitution class. In these 2 alerts, the recommended result occurred 54.8% of the time on average.

Corollary Order alerts provide a second example. In these situations, a user orders test or procedure "X" and the alert recommends one or more additional test or procedure "Y" (Figure 3). In these cases, regardless of whether the user accepted or canceled the alert or performed other actions as a direct result of the alert, we want to know at the end of the encounter what was ordered, "X," "Y," both, or neither? Fortunately, the same approach and logic can be used to generate reports for these alerts. The structure is the same and what differs is the definition of success. In this case, what is desired is "X + Y" and not "X" or "Y" alone. Using this logic, a report for this class of alerts can be

generated (Figure 3). In this example, the recommended result occurred in an average of 69.6% of cases.

The approach to medication-related alerts is similar but often more complex. It is unusual for there to be a one-to-one recommendation in medication substitutions. Usually a class of medications is contraindicated in a population defined by an age range, laboratory findings such as renal insufficiency, a specific genetic marker, or a diagnosis. The recommended substitution might depend on other factors including the initial reason for prescribing the triggering medication. Thus the recommended substitution values might be organized and contained in two or more groups. Furthermore, the valid medication alternatives themselves usually constitute a class rather than a single entity. The reporting program must be able to detect all the valid substitution values. Figure 4 illustrates a Substitute Medication alert. In this case a class of medications is relatively contraindicated in the elderly because of risks of insomnia, delirium, sedation, and cognitive impairment. Members of this medication class are prescribed for a variety of indications, and recommended alternatives depend on these reasons. Such condition-dependent

alternatives can be offered within an order set.

Despite these complexities, it was possible to develop a program that can automatically detect the triggers and recommended substitutions for medication-related alerts, and produce a report (Figure 5). Using pharmaceutical classes and generic drugs facilitates this. We are also able to take advantage of naming conventions in our order set development where, for example, all groups that contain medications begin with the prefix MED. For the 15 alerts in this example, the average rate of prescribing the preferred medication is just 10% (range, 1.5%-54.3%). Because nonpharmacologic therapies may be very appropriate, prescribing "none" (neither the trigger nor the recommended alternative medications) might be considered a positive outcome. Taking the "none" and "Y" results together yields an overall average success rate of 24.8% (range, 9.4%-75.5%) for these alerts.

DISCUSSION

It is difficult to understand or to improve what isn't measured or at least examined. To comprehend which alerts achieve their intended goals, one needs more than only rates of firing,

Alert Description	Order Type	None	"X"	"Y"	"X+Y"	Total
ELDERLY AND BENZODIAZEPINES	SmartSet	9.7%	36.9%	6.5%	46.9%	6989
ELDERLY AND SKELETAL MUSCLE RELAXANTS	SmartSet	19.9%	59.4%	8.2%	12.4%	1723
ELDERLY AND Z DRUG SLEEP AGENTS	SmartSet	9.6%	71.5%	11.6%	7.3%	1545
ELDERLY AND ANTIHISTAMINES/ANTICHOLINERGICS	SmartSet	21.8%	51.5%	14.6%	12.1%	1540
ELDERLY FALL RISK AND ANTIPSYCHOTICS/SLEEP MEDICATIONS	SmartSet	11.6%	63.5%	4.5%	20.5%	978
ELDERLY AND LORAZEPAM OR OXAZEPAM	SmartSet	7.7%	83.0%	1.7%	7.7%	951
DRUG INTRCTN CLOPIDOGREL+PO ESOMEPRAZOLE OR OMEPRAZOLE	SingOrd	5.6%	25.9%	20.2%	48.3%	752
ELDERLY FALL RISK OR DEMENTIA+TCA/PROCHLORPERAZINE	SmartSet	29.7%	49.6%	6.6%	14.1%	708
ELDERLY TERTIARY TRICYCLIC ANTIDEPRESSANTS	SmartSet	45.0%	43.9%	4.2%	6.9%	642
ELDERLY AND PROMETHAZINE	SmartSet	20.0%	49.2%	21.1%	9.8%	551
DRUG INTERACTION WARFARIN+SULFA	SmartSet	21.3%	17.7%	54.3%	6.8%	503
DRUG INTERACTION STATIN+GEMFIBROZIL	SmartSet	36.8%	36.2%	22.7%	4.3%	163
ELDERLY AND BUTALBITAL	SmartSet	25.8%	60.6%	8.3%	5.3%	132
ELDERLY AND DESICCATED THYROID	SmartSet	11.1%	86.1%	2.8%	0.0%	72
ELDERLY AND INDOMETHAZINE	SmartSet	18.5%	76.9%	1.5%	3.1%	65
Total		14.8%	48.2%	10.0%	27.0%	17314

Figure 5. Substitute Medication Alerts with results. "X" is the medication that triggered the alert, "Y" is the recommended substitution medication. "X+Y" represents encounters in which both medications were ordered. This report also indicated whether the medication was ordered via an ambulatory order set or via an attached single order (SingOrd).

PO = by mouth; TCA = tricyclic antidepressant.

acceptance, and override. First, one needs an explicit understanding and articulation of the goal. Next, one needs to determine whether that goal was achieved. To make this a realistic endeavor on a large scale, one needs a systematic approach.

We developed a classification system that allows development of batch reports on alert process outcomes or goals. Our initial approach was empiric and influenced heavily by the existing alerts in our system. We simplified the classification system by assigning alerts to only a single class, even though some alerts have features of more than one class. For these reasons, this taxonomy is certainly not exhaustive. Other classes of alerts no doubt exist in other systems today and will exist within our system in the future. We next created batch reports for some classes of alerts.

> **Alert success results ... may seem low and disappointing ... but likely will not surprise those experienced with alert acceptance and override rates.**

To date, we have developed reports for the Corollary-Order, Corollary-Medication, Create-Order, Substitute-Order, and Substitute-Medication classes of alerts. These represent the top three plus two less-frequent classes of alerts at KPNW, accounting for fully two-thirds of our current alerts. The fourth most common class, Informational Only, is not amenable to reporting of this nature because there is no measurable outcome event. Creating outcomes reports for several other classes, including those related to documentation, may also prove difficult and require special techniques, structured data entry, Concept Unique Identifiers, or natural language processing.

In some systems, certain alerts are designed to allow overrides with rationales, usually selected from a list, with or without the ability to add free text comments. We have very few of these in our system, except in the standard drug-drug and drug-allergy alerting activity. A challenging metrics issue is how to handle such alerts. Examples of overrides include "patient refused," "benefit outweighs risk," "no good alternative," "postponed for medical indications," or "doubtful allergy." When a valid override or postpone reason is selected, this might be counted as alert success or at least not failure, but this assumes that only valid reasons are available for the given alert and that the reasons are selected with fidelity. In practice, some alerts are overridden, often at a very high rate, because the alert is a "false positive" based on incomplete or faulty data, incorrect alert logic, or inaccurate knowledge synthesis. A complete evaluation of the effectiveness of such an alert could thus require examining the selected override or postpone reasons in light of the actual clinical data to determine whether the reasons were applied appropriately.

We provided examples of metrics for three classes of alerts. For the Substitution-Order, Corollary-Order, and Substitution-Medication alerts we measured in this way, the desired outcome was achieved on average 54.8%, 69.6%, and 24.8% of the time, respectively. Running the same report with different alerts, or perhaps even the same alerts after the alert or the alerting environment was modified, would probably yield different numbers. An alert might be modified, for example, by developing a clearer message, making it more specific by incorporating exceptions in the logic, or adding a more useful action step. Examples of modifying the alerting environment would be decreasing the overall number of alerts, providing user education, or changing incentives.

Alert success results such as in our Substitute-Medication report (Figure 5) may seem low and disappointing to some people but likely will not surprise those experienced with alert acceptance and override rates. The value in having such data is first that it allows identification and examination of alerts that appear to be underperforming. Viewing this data also allows a better understanding of actual alert performance more generally and might encourage quality and safety officers, decision-support developers, and leaders to plan more realistic and comprehensive safety and quality strategies, rather than blithely assuming success from alerts alone. Findings such as these suggest that alerts may be part of such a strategy but often will not be sufficient. The main outcomes of this study were the successful development of a rule-based alerts taxonomy and the demonstration of its application in a reporting strategy. The full-scale application of this strategy with detailed outcomes to a corpus of alerts was out of the scope for this article and could itself justify a report. However, even preliminary and partial review of our alerts using this approach resulted in the elimination of several poorly performing alerts and the modification of others.

Our approach decreases the time and effort to produce alert process performance metrics reports, making it more feasible to run them regularly, and this in turn results in a more complete and dynamic representation of alert functioning. Because increased complexity makes measurement more difficult and costly, alert developers may want to add complexity and to use extra options and actions more thoughtfully and only where they add significant value.

This study has some limitations. A single author developed and determined the taxonomy and assigned alerts to the taxa. Although the primary intention and contribution of this work was to develop a reporting framework and approach rather than a broadly applicable and validated taxonomy, it would be useful to do this in the future. Two or more individuals making assignments with a reported interrater agreement would add confidence. We examined alerts in only a single institution using a single EHR, and we know we have not exploited all potential alert actions. There are no doubt alert classes that we did not identify, and we chose not to create an alert class containing only a single member in our data set. Furthermore, we assigned alerts to a single class, although some have actions that might place them in more than one. Alerts assigned to the Other category underline that the current taxonomy is not exhaustive. Further study with additional alerts and at different institutions should result in more classes. The specifics of our reporting approach will be generalizable only to a certain extent because each EHR has its own structure

A Metrics Taxonomy and Reporting Strategy for Rule-Based Alerts

and data model, although most EHRs will have similar components. Finally, this report did not attempt to address the highest level in the evaluation hierarchy of rule-based alerts performance, whether patient or clinician goals or outcomes were achieved.

CONCLUSION

We developed a taxonomy for rule-based alerts and demonstrated its application in developing outcome metrics reports on a large scale.

CLINICAL RELEVANCE STATEMENT

Understanding whether clinical decision alerts achieve their intended function is fundamental to developing effective alerts and realizing the safety, quality, and resource benefits of EHRs. To reach this understanding, individuals and groups charged with developing and maintaining alerts need enhanced tools and reports.

The approach presented here allows nuanced effectiveness reports on a large scale. ❖

References
1. Kuperman GJ, Bobb A, Payne TH, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc 2007 Jan-Feb;14(1):29-40. DOI: http://dx.doi.org/10.1197/jamia.M2170.
2. Belmont E, Chao S, Chestler AL, et al. EHR-related metrics to promote quality of care and patient safety. In: Minimizing EHR-related serious safety events. Washington, DC: American Health Lawyers Association; 2013 May 31. p 3.
3. van der Sijs H, Bouamar R, van Gelder T, Aarts J, Berg M, Vulto A. Functionality test for drug safety alerting in computerized physician order entry systems. Int J Med Inform 2010 Apr;79(4):243-51. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2010.01.005.
4. Wang JK, Shabot MM, Duncan RG, Polaschek JX, Jones DT. A clinical rules taxonomy for the implementation of a computerized physician order entry (CPOE) system. Proc AMIA Symp 2002:860-3.
5. Berlin A, Sorani M, Sim I. A taxonomic description of computer-based clinical decision support systems. J Biomed Inform 2006 Dec;39(6):656-67. DOI: http://dx.doi.org/10.1016/j.jbi.2005.12.003.
6. Wright A, Sittig DF, Ash JS, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. J Am Med Inform Assoc 2011 May 1;18(3):232-42. DOI: http://dx.doi.org/10.1136/amiajnl-2011-000113.
7. Wright A, Goldberg H, Hongsermeier T, Middleton B. A description and functional taxonomy of rule-based decision support at a large integrated delivery network. J Am Med Inform Assoc 2007 Jul-Aug;14(4):489-96. DOI: http://dx.doi.org/10.1197/jamia.M2364.
8. National Quality Forum. Driving quality and performance measurement—a foundation for clinical decision support: a consensus report [Internet]. Washington, DC: National Quality Forum; 2010 Dec [cited 2013 Nov 29]. Available from: www.qualityforum.org/Publications/2010/12/Driving_Quality_and_Performance_Measurement_-_A_Foundation_for_Clinical_Decision_Support.aspx.
9. McCoy AB, Waitman LR, Lewis JB, et al. A framework for evaluating the appropriateness of clinical decision support alerts and responses. J Am Med Inform Assoc 2012 May-Jun;19(3):346-52. DOI: http://dx.doi.org/10.1136/amiajnl-2011-000185.
10. Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. J Am Med Inform Assoc 1997 Sep-Oct;4(5):364-75. DOI: http://dx.doi.org/10.1136/jamia.1997.0040364.

Our Wit Languisheth

Yet when we content ourselves with their discoveries [ie, those of ancient Greece], and calmly believe (which is mere sleepiness) that there is now no more place for new inventions, the spritely edge of our own wit languisheth, and we extinguish the lamp which they lighted to our hands.

— William Harvey, MD, 1578-1657, English physician, first to describe the circulatory system