The impact of electronic medical records data sources on an adverse drug event quality measure

Michael G Kahn,1,2,3 Daksha Ranade3

ABSTRACT
Objective To examine the impact of billing and clinical data extracted from an electronic medical record system on the calculation of an adverse drug event (ADE) quality measure approved for use in The Joint Commission’s ORYX program, a mandatory national hospital quality reporting system.

Design The Child Health Corporation of America’s “Use of Rescue Agents—ADE Trigger” quality measure uses medication billing data contained in the Pediatric Health Information Systems (PHIS) data warehouse to create The Joint Commission-approved quality measure. Using a similar query, we calculated the quality measure using PHIS plus four data sources extracted from our electronic medical record (EMR) system: medications charged, medication orders placed, medication orders with associated charges (orders charged), and medications administered.

Measurements Inclusion and exclusion criteria were identical for all queries. Denominators and numerators were calculated using the five data sets. The reported quality measure is the ADE rate (numerator/denominator).

Results Significant differences in denominators, numerators, and rates were calculated from different data sources within a single institution’s EMR. Differences were due to both common clinical practices that may be similar across institutions and unique workflow practices not likely to be present at any other institution. The magnitude of the differences would significantly alter the national comparative ranking of our institution compared to other PHIS institutions.

Conclusions More detailed clinical information may result in quality measures that are not comparable across institutions due institution-specific workflow, differences that are exposed using EMR-derived data.

INTRODUCTION
National reporting of clinical quality and patient safety measures has expanded rapidly and shows no signs of abating.1–3 At the same time, a growing body of literature has raised concerns about the accuracy and comparability of frequently reported measures that initially were designed and validated for internal quality monitoring but now are used in national comparative ‘quality report cards’, especially in the pediatric population.4–6 This study adds to that cautionary literature by examining differences in a single adverse drug event (ADE) quality indicator that is available as a national comparative quality measure in The Joint Commission’s mandatory ORYX quality reporting system.

BACKGROUND
Established in 1997, The Joint Commission ORYX program is one of the oldest national comparative quality score card.7 8 Reporting ORYX quality measures has been mandatory for The Joint Commission accreditation since 2002. All ORYX measures are available to the public at no cost via the QualityCheck website (http://www.qualitycheck.org).

The vast majority of national clinical quality indicators are based on administrative data sources that are created primarily to support reimbursement. For example, the Agency for Healthcare Research and Quality hospital quality indicators are widely used measures of potential quality issues specifically designed to use data found in standard discharge administrative records.9 10 There are many well-recognized limitations with the use of administrative data sets compared to comprehensive patient record abstracts for indicators of clinical quality.11–14 But the widespread availability of administrative data to support billing requirements make them the most widely used source of internal and comparative quality indicators appearing in publicly reported scorecards.15 16

Electronic medical records (EMRs) contain detailed clinical data that are not contained in administrative data sets. The availability of more clinically relevant data in electronic queryable format represents a new source of data that can be leveraged without the expense of manual chart abstraction. The American Recovery and Reinvestment Act contains explicit language linking the ‘meaningful EHR user’ to the ability to capture and report clinical quality measures.17 The recent recommendations by the Meaningful Use Workgroup of the Department of Health and Human Services’ Health IT Policy Committee proposed a large array of quality indicators to be generated by EMRs by 2015.18

Trigger tools are a retrospective computer-based adverse event detection methodology developed by Classen and colleagues at LDS Hospital as an alternative strategy to voluntary manual reporting or random manual chart abstractions for assessing potential ADEs.19 20 The trigger tool methodology was adopted and significantly expanded by the Institute for Healthcare Improvement (IHI).21 22 The current IHI web site lists eight sets of trigger tools designed for specific clinical applications along with links to a large library of trigger tool white papers, implementation instructions, and on-line data collection forms.23 The pediatric-specific ADE toolkit was designed jointly by IHI and the Child Health Corporation of America (CHCA), a national collaboration of 48 free-standing pediatric facilities.


Research paper
and is available from the CHCA. Takada published a description of the development and validation of 15 pediatric trigger tools, labeled T1- T20.24

CHCA combined two trigger tools that focused on the use of flumazenil, which reverses the effects of benzodiazepines (Trigger T3), and naloxone, which reverses the effects of narcotics (Trigger T3) into a single quality measure. The combined triggers, called ‘Use of Rescue Agents—ADE Trigger’, were submitted to The Joint Commission as a single ORYX quality measure. The approved definition of the combined quality measure is given in table 1. CHCA defined the quality measure using pharmacy charges only, without any required manual chart reviews, in order to calculate the measure directly using billing data available in the CHCA PHIS national data warehouse (described below). The Joint Commission approved the measure, which is now being reported to the ORYX program by at least seven CHCA institutions, including our institution.

RESEARCH QUESTIONS
Numerous studies have shown significant differences in findings when EMR-derived data rather than administrative data are used.25-28 In this study, we examined the impact of EMR data on the CHCA-defined, ORYX-approved ‘Use of Rescue Agents—ADE Trigger’ quality measure. The basic questions underlying this investigation were:

▶ How does the availability of detailed clinical data alter the results of the quality measure?
▶ If differences in the quality measure are detected, what are the causes of the observed differences?
▶ Are the observed causes likely to be different across institutions and therefore impact the validity of cross-institutional comparisons of the quality measure?
▶ How do the observed differences within a single institution compare to the differences seen across multiple institutions?
▶ How different is the quality measure for the same institution calculated using local data compared to using national data?

METHODS
Human subjects protections
The Colorado Multiple Institutional Review Board approved the protocol for the conduct of this study. CHCA provided the query specifications used to construct the quality measure in the form of an Impromtu (IBM Cognos 7) report written against the CHCA PHIS national database. The PHIS database contains inpatient demographics, diagnostic and procedure codes, and detailed charge data from 48 not-for-profit free-standing children’s hospitals in the USA. Data are subjected to rigorous reliability and validity checks prior to inclusion into the database. PHIS medication charges are based on pharmacy medication charge records provided by participating institutions. Each institution’s charge master codes are mapped manually by CHCA into a single uniform charge coding system. The Impromtu report provided by CHCA uses PHIS-specific pharmacy charge codes for narcotics, benzodiazepines, flumazenil, and naloxone medications.

Institutional-level data were obtained from The Children’s Hospital (TCH), Denver, using a comprehensive commercial electronic medical record system (EPIC Systems) which has implemented physician computer-based order entry, electronic medication administration, and integrated pharmacy orders and billing. Because of the additional detailed level of data available from the EMR, we were able to calculate the quality measure results from four data sources: (1) ‘medication orders placed’ records, (2) ‘medications charged’ records, (3) ‘medication orders with associated charges’ records, and (4) ‘medication administered’ records. We modified the CHCA measure definition (table 1) in the following ways:

▶ The PHIS charge codes for flumazenil, naloxone and the drug classes for benzodiazepines and narcotics were replaced with the corresponding TCH-specific charge codes.
▶ TCH-specific pharmacy order codes for the same drugs and drug classes were mapped to drug codes used in the PHIS query to provide access to TCH orders for these drugs.
▶ TCH-specific drug dispensing codes for the same drugs and drug classes were used to provide access to the TCH electronic medication administration (eMAR) records.

Access to medication orders requires electronic provider order entry. Access to medication charges requires access to detailed pharmacy charge records. ‘Orders with associated charges’ requires merging the previous two data sources. ‘Medications administered’ is perhaps the most clinically relevant because it captures the actual clinical care perspective. Access to these data requires an electronic medication administration record documentation system. Thus three of the EMR-based data sources require sophisticated inpatient electronic systems (Computerized Physician Order Entry (CPOE) and eMAR) that are not available in the majority of US hospitals.29-30

National comparative-level data were obtained from PHIS using the CHCA-supplied Impromptu report. Forty of the 48 PHIS hospitals submit detailed pharmacy charge data to CHCA required to execute the quality measure query.

Hospital encounters rather than unique patients form the underlying population. The denominator defines the population at risk for the quality event. For this quality measure, all inpatient encounters with evidence of exposure to at least one narcotic or benzodiazepine-containing medication are included in the at-risk population. The numerator defines those members of the at-risk population (members of the denominator) who had evidence of the trigger event, which was the administration of either naloxone or flumazenil, indicating a potential adverse drug event. For definitions using orders, the denominator and numerator used order codes queried from the physician’s orderables list. For definitions using charges, the appropriate charge codes from the TCH pharmacy charge master were used. For definitions using medication administrations, the appropriate drug dispensing codes from the pharmacy system were used.

The definition of the quality measure that has been approved by The Joint Commission does not define a temporal sequence or duration between denominator and numerator events. Thus, as defined, the rescue agent could precede the narcotic or benzodiazepine administration or the rescue agent could follow the narcotic or benzodiazepine dose by days or weeks within the same hospital stay. Both temporal scenarios are nonsensical

Table 1 Use of rescue agent: trigger tool/quality measure definitions

<table>
<thead>
<tr>
<th>Trigger short description</th>
<th>Denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reversal of adverse drug event based on use of rescue agents</td>
<td>Patients who received opiates which includes all narcotics as well as narcotic combinations, or benzodiazepines.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>numerator</th>
<th>Inclusions</th>
<th>Exclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients in the denominator who receive naloxone or flumazenil (agent-specific antidotes)</td>
<td>Ages between 30 days and 12 years</td>
<td>Ages less than 30 days or greater than 12 years</td>
</tr>
</tbody>
</table>

clinically for a potential ADE. However, in order to follow the approved definition as closely as possible, the queries we constructed also did not implement any temporal constraints.

**Institutional trigger rate**

Four versions of the ‘Use of Rescue Agent—ADE Trigger’ quality measure’s numerator and denominator events were developed as SQL-based queries against the EPIC electronic medical record system:

1. Events defined using medication order codes (labeled as ‘Orders Placed’ in figure 1).
2. Events defined using medication charge codes (labeled as ‘Medication Charged’).
3. Events defined by intersecting medication orders with medication charges (labeled as ‘Orders Charged’).
4. Events defined using medication administration codes (labeled as ‘Administrations’).

Denominators used the appropriate codes (order, charge or administration drug codes) for narcotics and benzodiazepines. Numerators used the same denominator codes but required the existence of flumazenil if a benzodiazepine was present or naloxone if a narcotic was used. If an encounter met criteria multiple times, it was counted only once.

For all queries, the same encounter-level data was used for the age inclusion and exclusion criteria.

Queries were executed for all inpatient admissions discharged from The Children’s Hospital, Denver Colorado between January 1, 2008 and December 31, 2008 using Business Objects Crystal Reports XI and EPIC System’s Clarity database. Query results were verified by manual chart audits of all numerator cases and a random sample of denominator cases. Only query correctness was validated, not clinical correctness or appropriateness as was done in the Takada study because the quality measure definition as approved by The Joint Commission does not require any manual clinical validation. No random chart audits were performed on cases not included in the denominator.

**National comparative trigger rate**

Queries were executed for all inpatient admissions discharged between January 1, 2008 and December 31, 2008 using IBM Cognos 7 Impromptu and CHCA’s PHIS database using the CHCA supplied Impromptu report that CHCA uses to report the measure to the The Joint Commission ORYX program. The CHCA query uses pharmacy billing records as its data source and therefore most closely matches the ‘medication charged’ data query from our local EMR.

No chart audits validating the correctness of the CHCA-supplied Impromptu report could be performed as patient identifiers are encrypted within the PHIS database and we have no access to patient records at other institutions.

**RESULTS**

From January 1, 2008 to December 31, 2008, 15,662 children were discharged from TCH. Of these, 5178, 4747, 4150, and 4116 discharges met the inclusion and denominator definitions using orders placed, medication charged, orders charged, and medication administrations respectively. It was not the case that each smaller denominator was a proper subset of a larger denominator population. For example, there were encounters in the medication administration denominator that did not have either an order or a charge. Interpreting results raised by these observations are explored in the Discussion section.

Figure 1 plots denominators, numerators and rates (numerator/denominator) calculated using the four EMR data sources for each month (medication orders placed, medication orders with associated charges, medications charged, and medications administered). The four lines in this figure use TCH-only EMR data. Panel A plots the four denominators (population at risk—patients where a narcotic or a benzodiazepine was ordered, charged or administered during an encounter), panel B plots the numerators (denominator encounters where either naloxone or flumazenil was ordered, charged or administered), and panel C plots the quality measure (rate=numerator divided by denominator). Only the rate is reported to The Joint Commission ORYX program as the quality indicator. Table 2 lists the 2008 annual totals for the denominators, numerators, and trigger rates for the four EMR data sources and for the PHIS national database.

In all three measures (denominator, numerator, and rate), ‘medication orders placed’ resulted in a significantly higher number of included encounters, followed by medications charged followed by medication orders charged followed by actual medication administrations. Focusing on the three data sources with some evidence that the patient may have received any of the rescue agents (the two charges and medication administrations-based data), medication charge-based methods identified more encounters billed for narcotics or benzodiazepines than were identified from medication administration records (figure 1, panel A). For the calendar year (table 2), there was a 15% difference (651 encounters) between medication charged and medication administrations.

For the numerator (figure 1, panel B), in all months there were more encounters flagged using medication charge records than were flagged using actual medication administration records. For the calendar year (table 2), there was a 215% relative difference (66/31 ÷ 2.15) between charges and administrations, an absolute difference of 35 encounters. Using relative differences to interpret the numerator population can be misleading due to very small numbers which artificially magnifies small absolute differences (eg, a 215% relative difference but an absolute difference of only 35 encounters). The largest absolute monthly difference was five cases in the medication charges numerator that were not in the medication administration numerator.

The quality measure (figure 1, panel C) followed a similar pattern in the relative position of each data source: the rate using ‘medications charged’ data was approximately the same as the rate using ‘orders charged’ data. Both of these rates were larger than the rate using ‘medication administrations’ data. For the calendar year (table 2), the annual rates were 1.59% for medication charged data (66 numerator events), 1.50% for medications orders charged data (54 numerator events) and 0.75% for medication administrations data (51 numerator events).

Figure 2 plots the 2008 annual quality measure for 40 CHCA hospitals. Monthly measures are not plotted but are available as a data appendix. There was a 40-fold difference between the highest (14.56%) and the lowest (0.56%) annual trigger rate. The TCH annual rate using the PHIS database was 1.62% (figure 2; Institution ID=19).

Table 2 lists the annual number of denominator and numerator encounters plus the annual rate for the four EMR-based data sources and the PHIS national database. Because the CHCA PHIS trigger tool uses charge records to calculate its trigger tool rate, the 1.62% annual rate reported by PHIS is conceptually most similar to the ‘medications charged’ EMR-based rate of 1.39%. These two ‘conceptually similar’ measures show a 16% relative difference (1.62%/1.39%−1.165=16% relative difference) despite both calculations being based on pharmacy charge records.
Figure 1  Denominators (panel A), Numerators (panel B) and Rates (panel C) using four electronic medical record (EMR) data sources. Detailed data used to derive these figures are available as an Excel spreadsheet in a data appendix available from the JAMIA website.
Table 2  Annual (2008) results by local and national data sources

<table>
<thead>
<tr>
<th>Data source</th>
<th>Denominator</th>
<th>Numerator</th>
<th>Trigger Rate = denominator/ numerator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orders placed</td>
<td>5178</td>
<td>411</td>
<td>7.94%</td>
</tr>
<tr>
<td>Medication charged</td>
<td>4747</td>
<td>66</td>
<td>1.39%</td>
</tr>
<tr>
<td>Orders charged</td>
<td>4150</td>
<td>54</td>
<td>1.30%</td>
</tr>
<tr>
<td>Medication administered</td>
<td>4116</td>
<td>31</td>
<td>0.75%</td>
</tr>
<tr>
<td>PHIS (medication charged)</td>
<td>3270</td>
<td>53</td>
<td>1.62%</td>
</tr>
</tbody>
</table>

The first four rows are based on local electronic medical record (EMR) data; the last row is based on the Child Health Corporation of America (CHCA) Pediatric Health Information Systems (PHIS) national data warehouse. The EMR-based ‘medication charged’ data source is conceptually similar to the charge-based PHIS data source.

DISCUSSION

Four different data sources extracted from the same EMR from a single institution using conceptually similar definitions for a medication adverse event trigger tool yielded substantially different results. This finding alone is not surprising. Each of the EMR-based definitions used a different data domain to calculate the quality indication: orders, charges, or medication records. However, all three data domains ‘looked’ for the same unambiguous clinical event: the use of a narcotic or benzodiazepine to identify denominator encounters and the additional use of flumazenil or naloxone to identify numerator encounters. For understanding clinical quality, one of the clinical data sources may be more meaningful than administrative or billing data. For example, although it may seem that medication administrations data is the most meaningful since it is the closest representation of what actually happened to the patient, the orders data source may provide a more meaningful viewpoint of medication ‘near-misses’ where an ordered medication would identify the potential adverse event but the medication was never administered to the patient.

In attempting to understand the source of the observed differences, our analysis has shown that for this quality measure, differences in denominators, numerators, and potential adverse event rates reflect a combination of common clinical practices and TCH-unique administrative, billing, documentation, data capture and workflow practices. A difference based on common clinical practices is predictable and may be similar across multiple institutions. A difference based on TCH-unique workflows is more insidious to detect, more difficult to adjust, and highly unlikely to be similar across institutions. An example of a common practice was the very large number of orders placed compared to orders charged or medications administered being due to PRN (‘as needed’) orders that do not result in the actual administration of medications. The quality measure definition would inappropriately flag these cases using orders data whereas medication administration data would only flag those cases where the PRN medication was actually given. Both orders charged and medication charged data exceeded the medication administration data, giving the appearance that medications were ordered and charged but not given. Workflow analysis revealed that medications are frequently pre-ordered and drawn for potential use by anesthesia during operating room cases. Even if these medications are not used, the patient is charged since the medication must then be discarded after surgery. Differences in medications charged versus orders charged give the appearance that some medications are being charged without corresponding orders. However, this difference is due to a specific anomaly of the EMR deployment and clinical documentation infrastructure at TCH. Computer-based physician ordering (CPOE) has been implemented in all clinical care settings except for the operating room. In addition, medications used by the anesthesiologist during a procedure are not entered into the electronic medication record as is done in all other clinical settings. The anesthesiologists record medication administrations only on a paper-based anesthesia record. Medications used during surgery are entered post-procedure by the pharmacist directly into the pharmacy system, resulting in a non-clinical ‘order’ which is not available in the EMR. Thus, medication charges can appear without any evidence in the eMAR or CPOE in the electronic clinical record of the drug having been ordered or administered, even though they had. These unique features of the TCH CPOE and eMAR deployments and pharmacy-based orders workflow result in no electronic clinical orders and no electronic documented medication administrations but a medication charge record. Correctly interpreting these alternative methods for calculating event rates requires detailed knowledge about institutional documentation workflows and data capture processes, which undermines efforts to create one standard definition that could be applied across disparate organizations for comparative reporting.

Using a national database that calculates the same quality measure using charge records obtained from multiple institutions, we observed differences across institutions that are roughly 40 fold different between the highest and lowest rate (figure 2: 14.56%/0.36%=40.44). In contrast, the largest difference in quality measure among the four EMR-based rates was 10.6-fold (table 2: 7.94%/0.75%=10.6) or 1.83-fold when excluding the outlying ‘orders placed’ rate (table 2: 1.39%/
0.75%–1.83%). In figure 2, over half of the PHIS institutions had comparative rates at 2% or less; 65% of PHIS institutions had comparative rates 3% or less. The differences in rates using EMR data would shift the relative ranking of TCH from the current 19th rank to the seventh rank for the lowest EMR-based rate (0.75%) and the 38th rank for the highest EMR-based rate (7.94%). Thus, wide swings in comparative rankings could be caused by differences in workflow and information technology infrastructure, which in turn, impact on the data that are available in the PHIS database. We expect similar workflow and IT implementation differences with similar impact on reported rates to exist at the other institutions in the PHIS comparison group. None of these institutional features that impact comparative trigger rate rankings is related to actual differences in clinical quality across institutions.

The observed 40-fold difference across the 40 PHIS institutions also raises questions about the comparability of the quality measure even when the same data source is used. The Takada validation study cited a positive predictive value of 12% (95% confidence interval: 3.40 to 28.20) for the T3 naloxone trigger and observed no instances of the T3 flumazenil trigger. Thus, the very wide range observed using the PHIS database also suggests significant issues with the comparability of this measure across institutions.

Even within the same institution, using conceptually similar charge-based data sources from EMR versus national data sources and the same query criteria, we observed differences in the calculated quality measure (table 2). This difference is likely caused by re-coding TCH-specific charge master records into an integrated charge master coding system used by PHIS to harmonize charges across multiple institutions. If differences in rates between local and national data are seen within the same institution, it seems highly likely that similar differences exist within other institutions, leading to another source of variability that also is not related to true differences in clinical quality across institutions.

Data quality is always an issue that could be a risk to the validity of a measure but it is especially difficult when dealing with low frequency events. In these settings, the inclusion or exclusion of a single encounter or event can dramatically change an apparent rate. One method to address this situation is to include error bars around all point estimates. However, none of the widely used measures provide definitions for calculating error bars. Many internal quality measures are used for trending over time. In this setting, the relative performance of an internal-use only measure is less critical. But in the context of comparative quality reporting, which pits one organization’s measure results against another, the relative performance of a measure is enormously important. If the measure’s performance is affected by local clinical practices as seen for this simple quality measure, the use of the measure in a comparative setting is highly problematic.

This study has a number of limitations. The population analyzed was entirely pediatric patients although the quality measure definition originally was developed and applied to adults. Only a single institution was used for the EMR-derived quality measures. One cannot generalize that the differences in rates observed in figure 1 would be observed at other institutions. Although we hypothesize that each institution would have idiosyncratic workflows and IT implementations that would impact observed rates, this conjecture cannot be confirmed without repeating this study at other institutions. Because the data sources require clinical applications that are not widely deployed (CPOE, eMAR, integrated pharmacy), it may be difficult to replicate this study at other locations. The use of the PHIS national database also limits the ability to replicate this study at locations who do not participate in this collaborative although many other national data collaboratives exist that could be used instead. The inability to validate the numerators and denominators obtained from the PHIS query at non-TCH institutions due to institutional privacy restrictions could be an issue although this barrier also exists in the current process used to report this measure to The Joint Commission. Finally, we examined only one quality measure designed to detect only one type of adverse drug event. Other quality measures may not exhibit the same behavior we observed with the CHCA Rescue Agent quality measure. Repeating the basic study design with other measures used in comparative quality report cards remains future work.

CONCLUSION

Electronic medical records can be a very rich source of highly detailed clinical data. This unprecedented level of detail opens new possibilities for defining clinical quality measures in more clinically meaningful ways. However, with this increased level of detail comes new difficulties with ensuring data across institutions are comparable. More detailed data begins to surface the impact of institutional-specific workflows into the data set. As EMRs become more prevalent and pressures to produce comparative clinical quality indicators for public consumption increases, the use of EMR data must be approached with caution. Constant validation and re-validation of measures, using laborious multi-institutional manual chart reviews, must be done when a new source of clinical data becomes available. Detailed understanding of specific components of the EMRs, electronic documentation, and institutional workflows in the interpretation of those data is necessary.

DATA APPENDICES

Monthly and annual denominator, numerator, rates, and plots for figures 1 and 2 are available as data supplements from the JAMIA website. These appendices can also be obtained directly from the corresponding author.

Acknowledgments The insightful comments of the two anonymous JAMIA reviewers are gratefully acknowledged.

Funding MGK is supported in part by Colorado CTSA grant 1 UL1 RR 025780 from NCRR/NIH and the TCH Research Institute. Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


