The 1999 Institute of Medicine report raised public awareness of the frequency and cost of adverse drug events in medicine. In response, in November 2000 a coalition of healthcare purchasers announced the formation of the Leapfrog Group, an organization dedicated to making “great leaps forward” in the safety and quality of health care in America. Their first target—computerized physician order entry (CPOE)—was selected specifically for its potential to reduce harm to patients from medications. The Leapfrog inpatient CPOE standard included a requirement that the organization operating CPOE should demonstrate via a test that their inpatient CPOE system can alert physicians to at least 50% of common serious prescribing errors. This paper outlines the development of this test which evaluates the ability of implemented CPOE systems to prevent the occurrence of medication errors that have a high likelihood of leading to adverse drug events. A framework was developed to include 12 different categories of CPOE based decision support that could prevent prescribing errors leading to adverse drug events. A scoring system was developed based on the known frequency and severity of adverse drug events. Simulated test patients and accompanying simulated test medication orders were developed to evaluate the ability of a CPOE system to intercept prescribing errors in all 12 decision support categories. The test was validated at a number of inpatient sites using both commercially available and custom developed CPOE systems. A web based application was developed to allow hospitals to self-administer the evaluation.

Adverse drug events (ADEs) are one of the leading categories of iatrogenic patient injury, accounting for 19% of all adverse events in the Harvard Medical Practice Study. More recent studies of the incidence of ADEs indicate that between 6.5% and 20% of patients admitted to hospital in the United States suffer an ADE. In addition to harming many patients, these events are costly; studies by two groups have estimated the attributable cost per ADE at $2000–2500, largely resulting from increased length of stay in hospital.

The Leapfrog Group currently uses a questionnaire to determine whether hospitals are complying with the inpatient CPOE standard. Recognizing the need for expert assistance in developing an effective test for the Leapfrog CPOE standard, the Leapfrog Group, with financial support from the Robert Wood Johnson Foundation and the California HealthCare Foundation, contracted with First Consulting Group. A report summary is available at http://www.chcf.org.

A recent study examining the literature on the effectiveness of clinical decision support systems showed that most of the favorable evaluations of clinical decision support systems have been written by the developers themselves, and that systems evaluated independently do not appear nearly as effective. These findings emphasize the need for, and value of, an independent evaluation process. In addition, reports highlighting the potential for CPOE to introduce significant errors and thereby actually impair patient safety make the need for such certification processes more pressing than ever. The mechanisms by which CPOE can improve the safety, quality, and efficiency of care have been discussed extensively in the literature.

**DEVELOPMENT OF THE METHODOLOGY**

**Overall strategy and principles**

The Leapfrog Group desired that the CPOE evaluation methodology should promote the development and adoption of functions to...
improve safety and quality and to serve as a quality improvement tool for hospitals as well as a method of certification to a standard. The methodology should test for sophisticated leading edge clinical decision support as well as basic commonly available decision support. It should provide feedback to hospitals about their system's clinical decision support capabilities and performance, including excessive alerts which may result in “alert fatigue” causing clinicians to ignore decision support or press for its deactivation, thereby decreasing the overall effectiveness of the system. Systems should also be evaluated for functions that promote efficiency and reduce waste such as duplicate order checking.

**Methodology development**

The CPOE evaluation methodology (fig 1) simulates different clinical scenarios using a wide variety of test patients and orders to evaluate how a hospital’s CPOE system responds to unsafe medication ordering and clinical situations.

The hospital taking the evaluation downloads a list of test patients with various demographic characteristics, medical conditions, and medication regimens and programs them into their CPOE testing environment. At this point, 4 hours are allotted for the user to program the test patients. If they re-log in within the time frame they are allowed to download a series of test orders to be entered against the test patients. The response of the CPOE system to the entered order is then noted and reported through the online evaluation system within a 2 hour time period. At the conclusion of testing the hospital receives an overall score and scores describing performance in specific clinical decision support categories (table 1). This feedback assists the hospital in selecting areas for new implementation of decision support or improvement of their current CPOE system.

Order set development began with an initial set provided by the Institute for Safe Medication Practices previously used by them for evaluating pharmacy information systems. This set was modified extensively to adapt it to the types of decision support appropriate for CPOE as shown by industry experience and literature on the kinds of medication ordering errors most likely to result in ADEs.\(^1\)\(^{-}\)\(^7\) Another set of test patients and orders was developed for pediatrics based on the literature\(^8\) and expert opinion. The resulting master order set consisted of over 130 adult and over 50 pediatric test orders directly ordered or for which laboratory studies must be considered for dosing. The master order set was modified extensively to adapt it to the types of orders that are released publicly at any given time and location, further protecting the content of the test material. In addition, the order set will be periodically reviewed and revised and modified, and new orders and scenarios introduced to maintain the validity and currency of the test.

The scoring system interprets the raw test results reported by hospitals that reflect the relative importance of each type of decision support for prevention of harm to patients. To achieve this, scores need to reflect the elements of both the severity of a potential ADE not intercepted by the system and its likely frequency. Thus, an event that happens rarely but is catastrophic should have a high score attached; an event that happens frequently but is inconsequential should have a low score attached. A key factor in the performance rating of a system is its ability to prevent errors that may result in an ADE.

**Table 1 Clinical decision support categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapeutic duplication</td>
<td>Therapeutic overlap with another new or active order; may be same drug, same drug class, or components of combination products</td>
</tr>
<tr>
<td>Single and cumulative dose limits</td>
<td>Specified dose that exceeds recommended dose ranges; will result in a cumulative dose that exceeds recommended ranges; can also include dose limits for each component of a combination product</td>
</tr>
<tr>
<td>Allergies and cross allergies</td>
<td>Allergy has been documented or allergy to other drug in same category exists</td>
</tr>
<tr>
<td>Contraindicated route of administration</td>
<td>Order specifying a route of administration that is not appropriate for the identified medication</td>
</tr>
<tr>
<td>Drug–drug and drug–food interactions</td>
<td>Results in known dangerous interaction when administered together with a different medication or results in an interaction in combination with a drug or food group</td>
</tr>
<tr>
<td>Contraindications/dose limits based on patient diagnosis</td>
<td>Contraindication based on patient diagnosis or diagnosis affects recommended dosing</td>
</tr>
<tr>
<td>Contraindications/dose limits based on patient age or weight</td>
<td>Contraindication based on laboratory studies or for which laboratory studies must be considered for dosing</td>
</tr>
<tr>
<td>Contraindications/dose limits based on laboratory studies</td>
<td>Contraindication based on the laboratory studies used in the treatment of this patient based on the evidence available</td>
</tr>
<tr>
<td>Contraindications/dose limits based on radiology studies</td>
<td>Contraindication based on radiology studies used in the treatment of this patient based on the evidence available</td>
</tr>
<tr>
<td>Corollary</td>
<td>Intervention that requires an associated or secondary order to meet the standard of care</td>
</tr>
<tr>
<td>Cost of care</td>
<td>Test that duplicates a service within a time frame in which there is typically minimal benefits from repeating the test</td>
</tr>
<tr>
<td>Nuisance</td>
<td>Order with such a slight or inconsequential interaction that clinicians typically ignore the advice/prompt</td>
</tr>
</tbody>
</table>

“on the fly” from the master order set makes it unlikely that a given site will be able to anticipate the specific orders that will be tested, and restricts the proportion of test patients and orders that are released publicly at any given time and location, further protecting the content of the test material. In addition, the order set will be periodically reviewed and revised and modified, and new orders and scenarios introduced to maintain the validity and currency of the test.

The scoring system interprets the raw test results reported by hospitals that reflect the relative importance of each type of decision support for prevention of harm to patients. To achieve this, scores need to reflect the elements of both the severity of a potential ADE not intercepted by the system and its likely frequency. Thus, an event that happens rarely but is catastrophic should have a high score attached; an event that is less severe but likely to happen often might similarly
deserve a high score. The likely frequency of ADEs that would result from specific ordering errors was determined from several large published and unpublished studies performed by automated ADE surveillance, a method superior to voluntary reporting for detecting ADEs.\textsuperscript{2, 19–21} Frequency was scored on a three point scale (most frequent, less frequent, least frequent). Severity determination was based on expert opinion among our advisors and described as life threatening, severe, significant, or not significant. A matrix was designed to determine summary scores from the attributes of severity and frequency based on previous work by Bagian et al.\textsuperscript{22}

Methodology testing and evaluation

We believed that testing the CPOE evaluation methodology at specific hospitals where CPOE systems were implemented was critical to validating the performance of the methodology. The goals of testing were (1) to validate the ability of the methodology to deliver a score reflective of the performance of a particular hospital CPOE system; (2) to ensure that the evaluation could be practically employed; and (3) to refine the methodology accordingly based upon real world use.

To address the goal of evaluating practicality of use it was essential to test the methodology at a number of hospitals that had implemented different vendors’ CPOE systems or built their own. This was important because different CPOE systems use different technical and logical strategies for providing decision support, and user interfaces and workflow vary significantly. Accordingly, we selected six medical centers around the United States as test sites. One hospital used a custom developed system; the others used products from five different commercial vendors.

Sites were provided in advance with the test patients and asked to record the amount of time required to program them into their CPOE test environment. Two of the authors (PK, EW) visited each hospital and entered the full set of test orders into the CPOE system against the preprogrammed test patients according to the evaluation protocol. The system response and time required to enter a representative subset of orders were recorded. At the end of the site visit feedback was provided to site personnel discussing performance of their system as reflected by the methodology.

**REFINEMENT OF EVALUATION TOOL**

Site visit experience revealed that the time required to program the test patients was less than 2 hours and the time needed to enter a representative subset of 30–40 test orders was about 1 hour at all sites. The format for reporting responses to the test orders was reduced to two possible descriptions: “Alert or information received, or order blocked” (for example, decision support intervention of some form) or “Order accepted, no alert or information received” (no decision support intervention).

### CPOE system performance relative to different test order categories

Testing revealed a range of CPOE capabilities across organizations in response to different order types (table 2). None of the systems tested was able to generate alerts or responses to the category “drug–diagnosis interactions”. This was not surprising; this category was included in the methodology to encourage and reward CPOE developers for building this capability into systems in the future. Three of the six test sites had operational alerts or additional information displayed in cases of therapeutic duplication. It is important to note that one center specifically decided not to operate therapeutic duplication decision support after evaluation of performance with such alerts in place; using the drug–drug interaction software provided by the third party vendor yielded too many false positive alerts. Two of the six centers operated dose limit checking on most of the orders tested and two centers on a few selected drugs only. Five of the six systems fired alerts around drug–drug and drug–laboratory interaction alerts; the sixth system checked allergy interactions; the seventh system performed dose checking; the third both. Three of the six systems had operational alerts for single/cumulative dose limits; five sites operated alerts for corollary orders. Our methodology for evaluating CPOE systems included in the system as reflected by the methodology.

Some of the systems tested were able to generate alerts or responses to the category “drug–diagnosis interactions”. This was not surprising; this category was included in the methodology to encourage and reward CPOE developers for building this capability into systems in the future. Three of the six test sites had operational alerts or additional information displayed in cases of therapeutic duplication. It is important to note that one center specifically decided not to operate therapeutic duplication decision support after evaluation of performance with such alerts in place; using the drug–drug interaction software provided by the third party vendor yielded too many false positive alerts. Two of the six centers operated dose limit checking on most of the orders tested and two centers on a few selected drugs only. Five of the six systems fired alerts around drug–drug and drug–laboratory interactions; the sixth system checked allergy interactions; the seventh system performed dose checking; the third both. Three of the six systems had operational alerts for single/cumulative dose limits; five sites operated alerts for corollary orders. Our methodology for evaluating CPOE systems included in the system as reflected by the methodology.

### Finalizing the grading system

An overall grading system was developed based on the severity scoring scheme together with the relative performance of implemented CPOE systems during testing. The grading system in early years of use will emphasize functions that most CPOE systems should be able to accomplish. With time, the requirements to achieve a satisfactory evaluation will become more stringent as hospitals will be expected to progressively strengthen the decision support capabilities of their system.

### Implementation of web based CPOE evaluation methodology

A hospital may access the evaluation program via a similar process used for the Leapfrog Hospital Safety Survey Programs. Hospitals can apply to take either the adult or pediatric evaluation; if both are applicable, they must complete them at separate times. A hospital taking the evaluation will obtain a user identification and security code. There are time constraints associated with each stage of the download and test reporting process, as well as a mandatory lock-out time between attempts to pass the test, to reduce the opportunities for gaming the system.

### DISCUSSION: CERTIFYING ELECTRONIC HEALTH RECORDS

This report describes the development of the first systematic evidence-based methodology for evaluating CPOE systems as actually implemented and operating at hospitals. The methodology focuses on testing decision support functions...
that are needed to intercept dangerous medication orders that can result in serious and/or frequent ADEs, and scores the performance of systems against an evidence-based and expert opinion-based standard. It also tests for certain functionalities that increase ease of use and improve efficiency of care.

Our inability to compare the sensitivity and specificity of our evaluation methodology with a gold standard for testing CPOE systems (as none exists) constitutes a significant limitation of this report. However, we are able to say that the testing methodology detected significant variations in performance among different CPOE implementations during evaluation at site visits, and that the differences detected were consistent with the observations of the evaluators as well as those of the personnel at the sites. Our agreement with the test sites precludes revealing more specifically the details of each system’s performance, but it was clear that important performance differences would have been reflected by the formal scoring system.

The CPOE evaluation methodology will complement the Leapfrog Group’s hospital questionnaire, and its implementation will complete the evaluation component of the initial Leapfrog CPOE standard. We anticipate that the test will be made publicly available by the Leapfrog Group in the next 1–2 years, and it will be required for all hospitals and outpatient clinics wanting to demonstrate compliance with the Leapfrog CPOE Inpatient and the Leapfrog Ambulatory.

While this is the first test developed to certify electronic health record applications in actual use, it is likely to be followed by tests of other applications. Similar efforts are already underway to certify electronic health record products at the vendor level. A national Certification Commission for Healthcare Information Technology has been created by HIMSS, NAHIT, and AHIMA to accelerate the adoption of information technology for providers and encourage payers/

A certification process will provide a clear definition of product capabilities and compatibilities. It will also ensure interoperability of these products with emerging local and national health information infrastructure. Hopefully, this certification process will reduce the risk of investment in information technology for providers and encourage payers/purchasers to offer incentives for investment.

**Authors’ affiliations**
P M Kilbridge, Duke University Health System, Durham, North Carolina 27710, USA

E M Welebob, First Consulting Group and the eHealth Initiative, 45 Hayden Avenue, Suite 2600, Lexington, MA 02421, USA

D C Classen, University of Utah School of Medicine and First Consulting Group, Salt Lake City, Utah 84103, USA

Competing interests: none.

**REFERENCES**
