

Improving patient safety via automated laboratory-based adverse event grading

Joyce C Niland, Tracey Stiller, Jennifer Neat, Adina Londrc, Dina Johnson, Susan Pannoni

Department of Information Sciences, City of Hope National Medical Center, Duarte, California, USA

Correspondence to

Dr Joyce C Niland, Department of Information Sciences, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010, USA; jniland@coh.org

Received 27 July 2011
Accepted 2 October 2011
Published Online First
14 November 2011

ABSTRACT

The identification and grading of adverse events (AEs) during the conduct of clinical trials is a labor-intensive and error-prone process. This paper describes and evaluates a software tool developed by City of Hope to automate complex algorithms to assess laboratory results and identify and grade AEs. We compared AEs identified by the automated system with those previously assessed manually, to evaluate missed/misgraded AEs. We also conducted a prospective paired time assessment of automated versus manual AE assessment. We found a substantial improvement in accuracy/completeness with the automated grading tool, which identified an additional 17% of severe grade 3–4 AEs that had been missed/misgraded manually. The automated system also provided an average time saving of 5.5 min per treatment course. With 400 ongoing treatment trials at City of Hope and an average of 1800 laboratory results requiring assessment per study, the implications of these findings for patient safety are enormous.

INTRODUCTION

Patient safety is of major concern during the conduct of clinical trials, where experimental and potentially toxic therapies are evaluated in humans.¹ Complete adverse event (AE) reporting during trial conduct imposes a large burden and presents a major challenge, requiring multiple assessments over time, for every treatment course for each participant.^{2–4} Chart review to assess the presence and severity of AEs is expensive, inefficient, and imperfect.^{5–6} Problems include under-reporting of low grade/recurrent AEs, and inconsistent or incomplete characterization and reporting of high grade AEs.⁷ Without accurate AE reporting, treatments may appear less toxic than they are, potentially endangering patients.⁸

Approximately 30% of more than 100 000 clinical trials registered on the <http://ClinicalTrials.gov/> website involve cancer. To assess AEs in oncology, the National Cancer Institute (NCI) developed the Common Terminology Criteria for Adverse Events (CTCAE),^{8–9} a graduated scale for evaluating the severity of ~350 qualitative and quantitative AEs, from grade '1' (least severe) to '4' (most severe), with grade '5' signifying AE-related death. Approximately 13% of the CTCAE is based on laboratory results, accounting for a significant number of reportable AEs (see figure 1 for examples).

This critical need to accurately and efficiently assess large quantities of laboratory-based AEs provides a prime opportunity to apply automated

decision support to reduce errors in transcription, calculation, and interpretation. However, to date development of such applications is lagging due to barriers such as organizational issues, inadequate design, poor system performance, non-standard terminology/clinical documentation, and lack of demonstrable system value.^{10–13} As Bates *et al* state, 'information technology has been viewed as a commodity, like plumbing, rather than as a strategic resource that is vitally important to the delivery of care.'¹⁴ Herein we report on a strategic decision support tool developed at City of Hope (COH) to improve subject safety, and our evaluation of this tool's utility and value.

As a NCI-funded Comprehensive Cancer Center, COH conducts ~400 clinical trials each year, enrolling over 1500 patients annually. Recognizing the enormous safety challenges created by this volume, in 2005 the COH Department of Information Sciences developed a software tool to automate detection of laboratory-based AEs. This decision support tool instantaneously assesses hundreds of electronic laboratory results to detect any abnormal findings, and grades AE severity according to CTCAE algorithms. While detecting abnormal laboratory results has been an informatics staple for many years,^{15–17} applying decision support to invoke the complex CTCAE algorithms to automatically grade AEs represents a novel application.

COH Clinical Research Associates (CRAs) have assessed over 1 million laboratory results using our automated grading tool to date. Recognizing the potential value to other institutions, COH developed an open source version, the Cancer Automated Lab-based Adverse Event Grading Service (CALAEGS). While experientially we believed this tool greatly enhanced the validity and efficiency of laboratory-based AE grading, a formal evaluation was required to confirm this impression. This paper describes our evaluation of CALAEGS, to our knowledge the first open source tool to assist with the complex task of grading laboratory data to ensure patient safety.

METHODS

CALAEGS intakes electronic laboratory data, and provides grading results through a web-based user interface, web services, and/or a Java API (application programming interface). The user interface allows institutions to customize the system to their specific data source formats and coding. The system is installed behind an institution's firewall to avoid confidentiality issues. Laboratory data can



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jamia.bmj.com/site/about/unlocked.xhtml>

BLOOD/BONE MARROW						
		Grade				
Adverse Event	Short Name	1	2	3	4	5
Hemoglobin	Hemoglobin	<LLN – 10.0 g/dL <LLN – 6.2 mmol/L <LLN – 100 g/L	<10.0 – 8.0 g/dL <6.2 – 4.9 mmol/L <100 – 80g/L	<8.0 – 6.5 g/dL <4.9 – 4.0 mmol/L <80 – 65 g/L	<6.5 g/dL <4.0 mmol/L <65 g/L	Death
Leukocytes (total WBC)	Leukocytes	<LLN – 3000/mm ³ <LLN – 3.0 × 10 ⁹ /L	<3000 – 2000/mm ³ <3.0 – 2.0 × 10 ⁹ /L	<2000 – 1000/mm ³ <2.0 – 1.0 × 10 ⁹ /L	<1000/mm ³ <1.0 × 10 ⁹ /L	Death

METABOLIC/LABORATORY						
		Grade				
Adverse Event	Short Name	1	2	3	4	5
Lipase	Lipase	>ULN – 1.5 × ULN	>1.5 – 2.0 × ULN	>2.0 – 5.0 × ULN	>5.0 × ULN	—
Magnesium, serum-high (hypermagnesemia)	Hypermagnesemia	>ULN – 3.0 mg/dL >ULN – 1.23 mmol/L	—	>3.0 – 8.0 mg/dL >1.23 – 3.30 mmol/L	>8.0 mg/dL >3.30 mmol/L	Death
Magnesium, serum-low (hypomagnesemia)	Hypomagnesemia	<LLN – 1.2 mg/dL <LLN – 0.5 mmol/L	<1.2 – 0.9 mg/dL <0.5 – 0.4 mmol/L	<0.9 – 0.7 mg/dL <0.4 – 0.3 mmol/L	<0.7 mg/dL <0.3 mmol/L	Death

Figure 1 Example of laboratory-based adverse event (AE) grading algorithms for two CTCAE V.3.0 organ systems: blood/bone marrow and metabolic/laboratory. LLN, lower limit of normal; ULN, upper limit of normal; WBC, white blood cell.

be submitted as comma-separated values, Extensible Markup Language (XML), or Health Level Seven (HL7) version 3 messages. Grading results are returned in a machine readable format compatible with the original input format, and as a human-consumable flowsheet rendered via Portable Document Format (PDF) (see figure 2).

CALAEGS incorporates national standards such as the Biomedical Research Information Domain Group (BRIDG) model¹⁸ and Unified Code for Units of Measure (UCUM),¹⁹ and is certified as bronze-level compatible with NCI’s Cancer

Biomedical Informatics Grid (caBIG^R).²⁰ It runs on Java 1.5+ in a J2EE web container (Tomcat 5.0+ and JBoss 4.0.5+) and requires a MySQL 5.0+ database.

CALAEGS assesses 39 laboratory-based AE terms based on NCI CTCAE version 3.0⁹ (refer to table 2). The grading algorithms received thorough testing across several phases, including unit, integration, system, and regression testing. The test approach included a range of conditions, including grade boundaries, simple and complex assessments, and fail conditions. CALAEGS assessments are considered preliminary only, as

The figure consists of two screenshots from the CALAEGS web application. The left screenshot shows the 'Adverse Event Grading Calculator' interface. It includes a 'Lab Test' dropdown set to 'Uric Acid, serum' and a 'Unit Of Measure' dropdown set to 'mg/dL'. The 'Normal Range' is shown as '2.5 - 7.0' (LLN/ULN) and the 'Test Result(mg/dL)' is '10.2'. The 'Grade' is calculated as '4'. A red box highlights the 'CTCAE Term' 'Hyperuricemia' and the 'Grade' '4' in the 'Grading Results' table below. The right screenshot shows a 'Lab-based AE Assessment Report' for Patient Name: Man, Any, Patient ID: ABC123, and Date Range: 9/2/2007-9/2/2007. The report is a table with columns for 'Unique Specimen Collection Date', 'Test Name', 'CTCAE Term', 'Result (reference range)', 'Grade', and 'Attrib'. It lists several lab tests including Calcium, Cholesterol, Creatinine, and Glucose, each with its corresponding CTCAE term and grade.

Figure 2 CALAEGS screenshots showing the entry screen for assessing a single laboratory result, for example, from an outside laboratory with no electronic file available (left), and the flowsheet generated to grade multiple laboratory-based adverse events (AEs) imported from an electronic file (right). CALAEGS, Cancer Automated Lab-based Adverse Event Grading Service.

Table 1 Protocols for comparing manual versus automated laboratory-based adverse event (AE) grading

Study number	Study phase	Number of COH patients graded	Number of laboratory results evaluated
Hematologic protocols			
1	Pilot	8	9775
2	Pilot	5	5117
3	I/II	1	614
Subtotal		14	15 506
Solid tumor protocols			
4	I	13	1382
5	I	3	379
6	I	1	313
7	I	2	150
8	I	2	65
9	II	4	542
10	II	1	266
Subtotal		26	3097
Total		40	18 603

COH, City of Hope.

some laboratory-based AE grades depend on human judgment as well, such as knowledge of additional patient conditions (eg, concomitant life-threatening consequences).

In a paired retrospective study design, we compared the accuracy and completeness of AE data graded manually, prior to the availability of the automated tool, with results reassessed via CALAEGS. We evaluated 10 sequential in-house therapeutic trials of varying size, diagnoses, and phase, from the time frame just prior to implementing our automated grading service, to minimize confounding factors (eg, CRA expertise). These 10 trials encompassed 40 patients and 18 603 laboratory results (table 1).

The 18 603 laboratory results were read into CALAEGS, and the automated results compared with manually graded results recorded in our clinical trials system. Discrepancies were categorized as missed AEs (true AEs that were not identified) or misgraded AEs (AEs with an incorrect numeric grade or direction, ie, hypo- vs hyper-). All discordant results were reviewed by our QA experts to verify that each suspected discrepancy was a true error, eliminating any protocol-specific exceptions (eg, if the study only requires recording the highest grade per course.)

To quantify AE grading efficiency, we conducted a prospective paired evaluation comparing time required for manual versus automated AE grading. In timed sessions, four CRAs graded five patients each from their current protocol portfolio, first manually and then 2–4 weeks later utilizing the CALAEGS tool, yielding 20 paired assessments. The assessment sequence was fixed (manual followed by automated), as if CALAEGS was run first, familiarity with the resulting AEs might have increased CRA efficiency when re-grading AEs manually.

A protocol specifying the design and regulatory processes for this evaluation was approved by the COH Institutional Review Board. The protocol stipulated that the Principal Investigator and biostatistician for studies evaluated were to be notified of any grading discrepancies identified; if any serious consequences were identified, the IRB and appropriate regulatory agencies would be notified as well. Analyses were conducted using SAS software version 9.1 (SAS Institute).

RESULTS

From the 18 603 laboratory results, 643 true AEs were detected. No valid AEs identified manually were missed by the automated system, and review of all 643 AEs by our QA experts verified

Table 2 Comparison of laboratory-based adverse events (AEs) detected by manual versus automated grading method by AE term for 643 true AEs*

AE term	True AEs	AEs correctly detected manually	AEs missed by manual method	AEs misgraded manually
Hematologic laboratory results				
Hemoglobin	47	43	2	2
Leukocytes	47	40	5	2
Neutrophils	29	22	6	1
Platelets	48	46	1	1
PTT*	6	5	1	0
Subtotal	177	156	15	6
Chemistry laboratory results				
Acidosis/alkalosis	5	5	0	0
Alkaline phosphatase	28	26	1	1
ALT*	38	34	1	3
Amylase	1	1	0	0
AST*	58	52	3	3
Bicarbonate serum low	14	10	4	0
Bilirubin	7	6	1	0
Cholesterol	13	10	2	1
Creatine phosphokinase	2	2	0	0
Creatinine	21	19	2	0
GGT*	5	4	1	0
Hypoalbuminemia	39	35	4	0
Hyper/hypocalcemia	44	38	5	1
Hyper/hypoglycemia	40	30	7	3
Hyper/hypokalemia	29	24	3	2
Hyper/hypomagnesemia	34	26	6	2
Hyper/hyponatremia	26	24	2	0
Hypertriglyceridemia	17	14	3	0
Hyperuricemia	7	3	3	1
Hypophosphatemia	32	24	5	3
Lipase	2	2	0	0
Proteinuria	4	2	2	0
Subtotal	466	391	55	20
Total	643	547	70	26
Percent		85	11	4

This table shows the true AEs that were missed, misgraded, or correct; 5 labs that were incorrectly graded manually as an AE, but the true Grade was 0, are not included here. ALT, serum glutamic-pyruvic transaminase; AST, serum glutamic-oxaloacetic transaminase; GGT, gamma glutamyl transferase; PTT, activated partial thromboplastin time.

that the CALAEGS grades were accurate. Therefore, discrepancies between the automated and manual approaches were attributable to errors made during manual grading, found to be inaccurate 15% of the time (96/643, table 2). Seventy laboratory-based AEs (11%) were missed by manual grading, and 26 manually graded AEs (4%) were misgraded (25 understated the condition, one was in the wrong direction).

Of the missed AEs, 86% (60/70) were relatively minor (grade 1–2). However, 22 severe AEs (grade 3–4) missed detection by the manual method, through lack of identification (n=10) or incorrect grading to a lower level (n=12). Out of 130 severe grade 3–4 AEs identified via CALAEGS, 17% were missed/misgraded manually. Overall, 40% of patients evaluated (16) experienced one or more missed/misgraded severe AEs.

Figure 3 shows the direction and magnitude of grading error for 101 missed/misgraded AEs. The majority involved under-reporting; however, in five instances the manually recorded AE grade was higher than the true result (recorded as grade 1, true grade 0). One misgraded AE (see ^{***} in figure 3) was recorded at the appropriate grade, however the direction was incorrect ('hyper' when it was actually 'hypo').

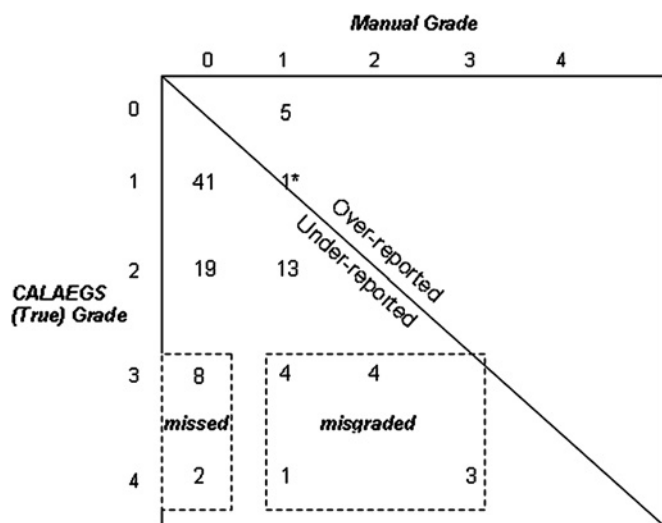


Figure 3 Missed/misgraded adverse events (AEs) by the manual assessment method, against the true grade as detected by CALAEGS; dashed boxes highlight the severe (grade 3, 4) missed/misgraded AEs. *Misgraded because of wrong direction: term incorrectly identified as 'hyper' instead of 'hypo'. CALAEGS, Cancer Automated Lab-based Adverse Event Grading Service.

The prospective timed grading evaluation showed that using CALAEGS led to time savings in 18/20 paired assessments (90%); the average time saved was 5 min 25 s (5:25) per treatment course (95% CI 2:24 to 8:26). For two assessments the decision support tool required slightly more time (10 s and 2 min).

DISCUSSION

Health information exchange systems can substantially impact medical quality and safety through automated decision making and knowledge acquisition tools.²¹ Yet to date the nation's healthcare system has fallen far short in applying new technology safely and appropriately to enhance the translation of new biomedical discoveries into practice.¹¹ Strategies for AE detection that incorporate electronically screened data can cost significantly less per AE detected, an attractive improvement over pure manual review.²²

The high prevalence of AEs has made patient safety a major concern when treating patients with experimental clinical trial agents.¹¹ Identification of AEs is a major challenge, and effective methods for detecting such events are required.^{6, 23} Because laboratory data are computerized, AEs detected through electronic surveillance of laboratory results and their normal ranges are particularly suited for automated decision support.²⁴

A very high overall accuracy level was seen in our evaluation (18 502 correct assessments, 99.5%). Yet the fact remains that 17% of all severe grade 3–4 AEs went undetected by traditional chart review, affecting 40% of patients evaluated. Fortunately, a thorough review of the medical records of these 16 patients showed that no harm occurred, as in each case concurrent medical problems led to appropriate care. However, the potential for patient harm certainly exists if severe AEs go undetected.

Missed/misgraded AEs are concerning not only for patient safety, but for overall scientific validity. In phase I studies, dose escalation is driven by AEs, such that discrepancies can impact study conduct. Comprehensive AE reporting is needed to correctly interpret trial results, and avoid under-representing toxicity burden. Even low grade AE detection is crucial in reporting clinical trials,^{1–6} for example, to uncover pharmacoge-

netic syndromes. While 78% of errors in our evaluation involved grade 1–2 AEs, even these reveal critical toxicity patterns prior to introducing experimental agents into standard care.

Although the time savings was less dramatic than we expected (~5.5 min per treatment course), even this small improvement translates into a potentially large benefit, given the volume of laboratory results per protocol (averaging 1800 per study in our evaluation). With an average of three courses of treatment for 1500 patients accrued annually at COH, even modest efficiency improvements have major impact.

Limitations and future plans

Due to the large number of laboratory results evaluated, it was not possible to directly assess every result for true AEs that might have been missed by both the manual and automated methods. However, we can reasonably infer that such false negatives are highly unlikely based on the testing and validation of the system.

Achieving the optimal specificity of detection systems often still requires some manual review, prompted by the automated decision support.⁶ CALAEGS prompts such a review when additional criteria are required to determine grade (eg, concurrent hospitalization or physiological consequences). Therefore CALAEGS is an aid to, not a replacement of, human judgment.

As with any decision support system, there is a potential danger when changes to the input data or algorithms occur, intentionally or unintentionally. Our domain experts are continually vigilant for any changes in laboratory reporting standards, and rigorous retesting/validation is performed if the algorithms are updated. Recently NCI released CTCAE V.4.0, with many more laboratory-based AEs involving qualitative criteria. Integration of additional data sources regarding patient status is optimal with the advent of CTCAE V.4.0, planned for our next system enhancements. The caBIG program is developing tools to manage AE collection and regulatory/institutional reporting requirements (eg, caAERS); integration of CALAEGS with such tools may facilitate accurate real-time identification of serious AEs that require immediate reporting.

Information technology can not only help detect AEs, but also facilitate more rapid response once an AE occurs.¹¹ Currently, the COH grading system is used as a data collection tool following treatment course completion. We are in the process of deploying the system to conduct nightly surveillance of the past day's laboratory results, to provide caregivers with refined signals indicating worsening patient conditions. Deployment will require an appropriate workflow in clinic, and avoidance of 'alert fatigue' among caregivers.^{25, 26} Adding a configurable rules engine interface to incorporate protocol-specific rules to 'fine tune' the algorithms will provide additional efficiency in future.

Conclusions

Our evaluation demonstrated that CALAEGS improves accuracy, completeness, and efficiency in detecting and grading laboratory-based AEs, facilitating documentation of the full toxicity profile of experimental agents. With the large number of clinical trials performed at centers nationwide, the potential beneficial impact on patient safety, efficient resource usage, and unbiased trial reporting is tremendous.

Acknowledgments We would like to express our gratitude to the QA and CRA teams who assisted with this evaluation: Claudia Aceves, Riza Apuan, Alicia Bogardus, Abigail Guinto, Deron Matsuoka, Beelynda Martinez, Isa Quazi, Amalia Rincon, Susan Hmwwe and Jennifer Simpson. In addition, we thank our City of Hope development team including: Russ Sarbora, Eric Huang, Dave Ko, Karen Rickard, Cindy Stahl, and Dr Doug Stahl.

Competing interests None.

Ethics approval The City of Hope IRB approved this study.

Provenance and peer review Not commissioned.

REFERENCES

1. **Jenkins J**, Hubbard S. History of clinical trials. *Semin Oncol Nurs* 1991;**7**:228–34.
2. **Moher D**, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* 2001;**357**:1191–4.
3. **Martin PJ**, Antin JH, Weisdorf DJ, et al. Reporting of adverse event data in hematopoietic stem cell transplantation clinical trials involving investigational new drugs or devices: a report from the William Guy Forbeck Foundation 2001 focus meeting on clinical trials in hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant* 2002;**8**:295–302.
4. **Goldberg RM**, Sargent DJ, Morton RF, et al. Early detection of toxicity and adjustment of ongoing clinical trials: the history and performance of the North Central Cancer Treatment Group's real-time toxicity monitoring program. *J Clin Oncol* 2002;**20**:4591–6.
5. **Bates DW**, Evans RS, Murff H, et al. Detecting adverse events using information technology. *JAMIA* 2003;**10**:115–28.
6. **Murff HJ**, Patel VL, Hripcsak G, et al. Detecting adverse events for patient safety research: a review of current methodologies. *J of Biomed Inform* 2003;**36**:131–43.
7. **Scharf O**, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol* 2006;**24**:3933–8.
8. **Trotti A**, Colevas AD, Setser A, et al. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol* 2007;**25**:5121–7.
9. **CTEP**. *NCI Adverse Event Grading Criteria, CTCAE v3.0*. http://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/ctcae3.pdf
10. **Bates DW**, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;**10**:523–30.
11. **Corrigan JM**, Kohn LT, Donaldson MS, et al. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press, 2001.
12. **Berg M**. Implementing Information systems in health care organizations: myths and challenges. *Int J Med Inform* 2001;**64**:143–56.
13. **Hersh WR**. Medical informatics: improving health care through information. *JAMA* 2002;**288**:1955–8.
14. **Bates DW**, Gawande AA. Improving patient safety with information technology. *N Engl J Med* 2003;**348**:2526–34.
15. **Bradshaw K**. Computerized Alerting System Warns of Life-Threatening Events. *Proc Annu Symp Comput Appl Med Care* 1986;**10**:403.
16. **Kuperman GJ**, Teich JM, Tanasijevic MJ, et al. Improving response to critical laboratory results with automation: results of a randomized controlled trial. *J Am Med Inform Assoc* 1999;**6**:512–22.
17. **Bates DW**, Leape LL. Doing better with critical test results. *Jt Comm J Qual Patient Safety* 2005;**31**:66–7.
18. **BRIDG**. <http://bridgmodel.org/>
19. **UCUM**. <http://unitsofmeasure.org>
20. *caBIGR Compatibility Guidelines*. https://cabig.nci.nih.gov/guidelines_documentation/
21. **Shekelle PG**, Morton SC, Keeler EB. Costs and benefits of health information technology. *Evid Rep Technol Assess* 2006;**132**:1–71.
22. **Bates DW**, O'Neil AC, Petersen LA, et al. Evaluation of screening criteria for adverse events in medical patients. *Med Care* 1995;**33**:452–62.
23. **Bates DW**, O'Neil AC, Boyle D, et al. Potential identifiability and preventability of adverse events using information systems. *J Am Med Inform Assoc* 1994;**1**:404–11.
24. **Gandhi TK**, Seger DL, Bates DW. Identifying drug safety issues: from research to practice. *Int J Qual Health Care* 2000;**12**:69–76.
25. **Singh H**, Thomas EJ, Sittig DF, et al. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? *Am J Med* 2010;**123**:238–44.
26. **Lee EK**, Mejia AF, Senior T, et al. Improving patient safety through medical alert management: an automated decision tool to reduce alert fatigue. *AMIA Annu Symp Proc* 2010:417–21.