Improving patient safety via automated laboratory-based adverse event grading

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

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. 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. Chart review to assess the presence and severity of AEs is expensive, inefficient, and imperfect. Problems include under-reporting of low grade/recurrent AEs, and inconsistent or incomplete characterization and reporting of high grade AEs. Without accurate AE reporting, treatments may appear less toxic than they are, potentially endangering patients.

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), a graduated scale for evaluating the severity of ~550 qualitative and quantitative AEs, from grade ‘1’ (least severe) to ‘4’ (most severe), with grade ‘5’ signifying AE-related death. Approximately 15% 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. 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.’ 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, 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 experimentally 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
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®).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

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.

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.
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

| Study number | Study phase | Number of COH patients graded | Number of laboratory results evaluated |
|--------------|-------------|-------------------------------|---------------------------------------|
| 1            | Pilot       | 8                             | 9775                                  |
| 2            | Pilot       | 5                             | 5117                                  |
| 3            | I/II        | 1                             | 614                                   |
| Subtotal     |             | 14                            | 15,506                                |

| Solid tumor protocols | Study phase | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|-------------|---|---|---|---|---|---|---|---|---|---|
|                       | Number of | 13 | 3 | 1 | 2 | 1 | 4 | 1 | 2 | 4 | 1 |
|                       | laboratory | 1382 | 379 | 313 | 150 | 65 | 542 | 266 | 3097 |
| Subtotal              |             | 26 | 177 |
| Total                 |             | 40 | 603 |

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

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 | AE correctly detected manually | AE missed by manual method | AE 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 |        |                                 |                             |                       |
| Aciddosis/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/hyperglycemia      | 40       | 39                              | 7                           | 3                     |
| Hyper/hypokalemia        | 29       | 24                              | 3                           | 2                     |
| Hyper/hypomagnesemia     | 34       | 26                              | 6                           | 2                     |
| Hyper/hyponatremia       | 26       | 24                              | 2                           | 0                     |
| Hypertiglyceridemia      | 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 undergrading; 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’).
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. Yet to date the nation’s healthcare system has fallen short in applying new technology safely and appropriately to enhance the translation of new biomedical discoveries into practice. Strategies for AE detection that incorporate electronically screened data can cost significantly less per AE detected, an attractive improvement over pure manual review.

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

A very high overall accuracy level was seen in our evaluation (18.502 correct assessments, 99.5%). The fact remains that 17% of all severe grade 3 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, for example, to uncover pharmacoge-
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. Bates DW, Evans RS, Murff H, et al. Detecting adverse events using information technology. JAMA 2003;290:115—28.
5. Murff HJ, Patel VL, Hripcsak G, et al. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36:131—43.
6. Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. J Clin Oncol 2006;24:3933—8.
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, Seltzer 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/ctcaev3.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—6.
14. Bates DW, Gowanwe AA. Improving patient safety with information technology. N Engl J Med 2003;348:2528—34.
15. Bradshaw K. Computerized Alerting System Warns of Life-Threatening Events. Proc Annu Symp Comput Appl Med Care 1998;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 Saf 2005;31:66—7.
18. BRIDG. http://bridgmodel.org/
19. UCUM. http://unitsofmeasure.org
cdbIGR 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 1998;36:452—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, Segal 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, Meija 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.