

Post-fielding Surveillance of a Guideline-based Decision Support System

Albert S. Chan, Susana B. Martins, Robert W. Coleman, Hayden B. Bosworth, Eugene Z. Oddone, Michael G. Shlipak, Samson W. Tu, Mark A. Musen, Brian B. Hoffman, Mary K. Goldstein

Abstract

Quality assurance (QA) processes for new technologies are used to ensure safety. Clinical decision support systems (DSS), identified by the Institute of Medicine (IOM) as an important tool in preventing patient errors, should undergo similar predeployment testing to prevent introduction of new errors. Post-fielding surveillance, akin to post-marketing surveillance for adverse events, may detect rarely occurring problems that appear only in widespread use. To assess the quality of a guideline-based DSS for hypertension, ATHENA DSS, researchers monitored real-time clinician feedback during point-of-care use of the system. Comments (n = 835) were submitted by 44 of the 91 (48.4 percent) study clinicians (median 8.5 comments/ clinician). Twenty-three (2.8 percent) comments identified important, rarely occurring problems. Timely analysis of such feedback revealed omissions of medications, diagnoses, and adverse drug reactions due to rare events in data extraction and conversion from the electronic health record. Analysis of clinician-user feedback facilitated rapid detection and correction of such errors. Based on this experience, new technologies for improving patient safety should include mechanisms for post-fielding QA testing.

Introduction

*ALL technology introduces new errors, even when its sole purpose is to prevent errors.*¹

Information technology has been cited as a key to improving the safety of health care delivery. In its 1999 report, *To Err Is Human: Building a Safer Health System*, the Institute of Medicine (IOM) emphasized the need for technologies specifically engineered to prevent medical errors. Such technologies include automated order entry systems, drug-drug interaction software, and decision support systems.¹ Leveraging clinical data from electronic health record systems (EHR), such technologies hold the promise of reducing errors in medical decision making that are due to inadequate information at the point of care.

The IOM and others have cautioned, however, that new technologies for health care providers can introduce unanticipated errors.¹⁻⁴ An implementation of computer interpretation of electrocardiograms (ECGs) at a U.S. academic medical center revealed that incorrect advice can significantly influence physicians: 67.7 percent of the physicians agreed with an inaccurate, computer-generated

interpretation of an ECG versus 33.1 percent when such advice was not presented.⁵ A study of drug interaction software used by community pharmacists revealed that these systems failed to detect clinically relevant drug-drug interactions in one-third of the cases.⁶ Goldstein et al. summarized a number of potential sources of error or harm in delivering drug recommendations via an automated decision support system (Table 1).³ The IOM concluded that prevention of errors introduced by the implementation of new technology requires careful attention to the vulnerabilities of any system.¹

Table 1: Potential sources of error in automated drug recommendation systems

Reasons why introducing a decision support system for drug recommendations may introduce new errors into the clinical workflow:

- Missing data leading to recommendation of a contraindicated drug.
- Potential interaction of the recommended drug with another drug prescribed for the patient.
- Inaccuracies in program logic which could lead to erroneous recommendations.
- Potential harm due to rearranging clinician priorities with required use of the decision support system (e.g., invoking a decision support system for hypertension when hypertension is not a clinical priority for that visit).
- The clinician-user has knowledge gaps that are directly relevant to the decision support system recommendations (e.g., promoting the use of a guideline-recommended drug without simultaneously providing information about dose limits).
- Generation of false expectations on the part of the clinician-user that the system will alert them to all problems.
- Potential for data overload.
- Potential for incorrect recommendations on cases that the system was not designed to handle.

An analogous challenge is faced with the approval of prescription drugs for the general public. Despite rigorous testing to establish safety and efficacy prior to drug approval, some problems are discovered only after widespread use of the drug. Phase III clinical trials are typically conducted over a short time period and may involve too few subjects to detect all adverse outcomes, especially when the events are rare. Also, the target population for use of the product expands after U.S. Food and Drug Administration approval.⁷ In recognition of these problems, the U.S. Food and Drug Administration Center for Drug Evaluation and Research (CDER) has a Post-Marketing Surveillance (PMS) system designed specifically to monitor and report adverse events data. CDER created MEDWatch, an Internet-based resource that includes an online submission form for health care professionals to report adverse events observed during use of medical products.⁸

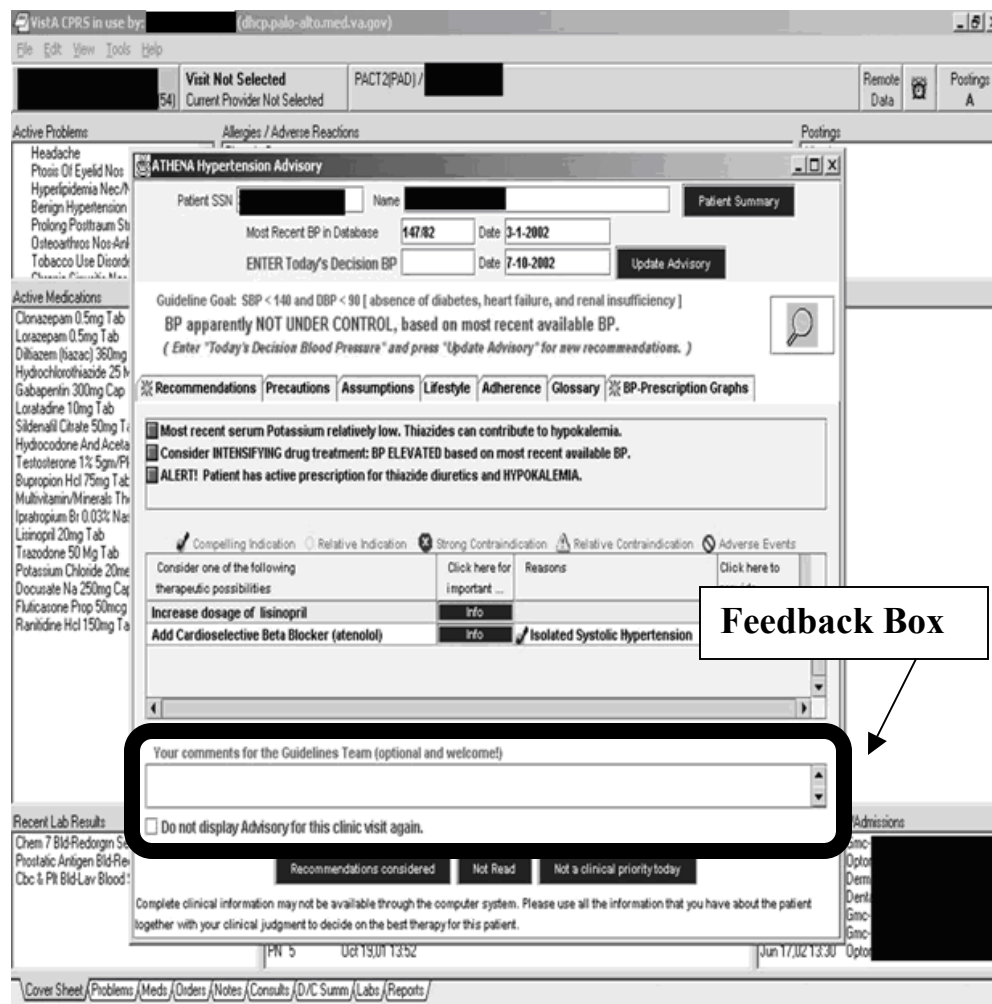
Clinical decision support systems may undergo predeployment testing prior to introduction into the clinical workflow. However, rarely occurring problems in

data sources, for example, may not be detected in predeployment testing. Miller and Gardner recognized the need for post-fielding surveillance of clinical software systems.⁹ Summarizing the findings of a large consortium of professional organizations, including the American Medical Informatics Association, Medical Library Association, and the Computer-based Patient Record Institute, they noted that software products may work well in isolation but face challenges when integrated into complex systems involving multiple clinical software systems and multiple vendors. Also, “raw complaints” from individual users were identified as potentially useful in monitoring and evaluating clinical software systems for potential sources of error.⁹ Gould and Lewis also noted the need for “design, test, measure, and redesign,” that is, the power of iterative user testing to discover and repair system problems.¹⁰

In this paper, researchers describe the method of maintaining quality assurance for a hypertension guideline system, Assessment and Treatment of Hypertension: Evidence-Based Automation Decision Support System (ATHENA DSS). Based on widely accepted national guidelines for hypertension (Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure [JNC] 6 and the Department of Veterans Affairs [VA]),^{11, 12} ATHENA DSS delivers treatment advisories to clinicians at the point of care. Built with EON technology for guideline-based decision support,¹³ ATHENA DSS consists of two main components: a hypertension knowledge base modeled in Protégé¹⁴ and a guideline interpreter that applies the information in the knowledge base to the clinical information retrieved from the computerized patient record system (CPRS) to create patient-specific recommendations for a patient encounter, on a visit-by-visit basis.¹⁵⁻¹⁷ ATHENA DSS displays advisories via an interface to the VA CPRS, a uniform EHR in patient care delivery settings. Recommendations generated by ATHENA DSS were displayed to primary care physicians during clinic visits with hypertensive patients (Figure 1). The system was accessible only to licensed clinicians individually enrolled in the system by study staff. The recommendations screen included a statement emphasizing the limitations of computer data and the importance of applying clinical judgment to decisionmaking.

ATHENA DSS included two features to maintain quality assurance after deployment. First, the user interface of ATHENA DSS, in addition to delivering drug recommendations, included a feedback box for clinician-users to enter free text comments during point-of-care system use. Impressions of clinician-users were captured in real-time, linking comments to the specific patient scenario in which the error was observed and reducing the potential for recall biases. Second, the actions of the ATHENA DSS program, including unanticipated error conditions known as “exceptions,” were logged during program execution. A scripting program scanned these log files for exceptions and automatically e-mailed messages to the ATHENA DSS programmers for investigation of potential sources of error in program function. To monitor for any unanticipated problems not detected in predeployment testing of ATHENA DSS, researchers monitored both the real-time feedback provided by clinicians and program exceptions during point-of-care use of the system.

Figure 1. Sample hypertension advisory pop-up window in computerized patient record system window



An example of ATHENA DSS displays an advisory pop-up window on top of the VA electronic medical record. In keeping with national guidelines for management of hypertension, the ATHENA DSS encourages use of thiazide diuretics; however, it also monitors for potential problems in their use and, for example, alerts clinicians to hypokalemia. Other parts of the system, not shown here, provide information about thiazide dosing to avoid hypokalemia. The feedback box at the bottom of the window allows clinician-users to enter free text feedback comments to the ATHENA DSS knowledge management team.

Methods

As part of a randomized trial to assess the overall effect of ATHENA DSS on choice of drug therapy and blood pressure control, recommendations were generated on a daily basis for 15 months at nine geographically dispersed clinical sites within the VA Durham, Palo Alto, and San Francisco Health Care Systems. Ninety-one primary care providers in the ATHENA arm of the study received the ATHENA Hypertension Advisory.

Clinicians were encouraged to comment about their interactions with the ATHENA DSS. Clinicians could phone or e-mail study staff with any concerns or questions about the program. Clinicians were also given the option to enter feedback during the viewing of the ATHENA Hypertension Advisory. Feedback comments submitted by clinician-users were initially logged into a Microsoft[®] SQL database. These comments were subsequently imported into a Microsoft Access[®] database to facilitate data analysis and monitoring by the ATHENA DSS team. Typically within 1 week of receipt of feedback, a member of the research staff reviewed the comments to identify any indication of possible error in the program.

Feedback data analysis

The feedback comments were later analyzed using a qualitative research approach. An ATHENA team member met with the project Principal Investigator to review all comments and classify the principal idea(s) expressed. New categories were created as needed. After all comments had been reviewed, the entire set of comments was reviewed again to classify them into the final set of categories. At least two members of the research team reviewed each comment. Any classification questions were resolved by consensus.

Text logs

Text log files of ATHENA DSS program actions were also monitored for potential errors. During the generation of ATHENA DSS recommendations, messages describing the actions of the program were recorded in text files, including any program exceptions that occurred. A Perl language text searching script extracted and compiled exception messages into a file that was sent to the ATHENA DSS team for daily analysis.

Results

ATHENA DSS displayed advisories for 19,859 clinic visits and 10,806 distinct patients during the study period. A total of 835 free text feedback comments were submitted by 44 of the 91 (48.4 percent) study clinicians via the feedback window. A median of 8.5 comments per clinician-respondent were received during the study period (range 1–140).

Free text comments were investigated as potential sources of error in the program or its data sources. In most cases, clinician reports were false positives for error. For example, a diagnosis questioned by the clinician as “inaccurate” could be found in the EHR, although apparently unknown to the clinician. Similarly, “missing” adverse drug reactions noted by clinicians were confirmed by chart review as not documented in the EHR.

However, in 23 (2.8 percent) feedback comments, investigation revealed 4 distinct problems in clinical data due to rare events in data conversion or data extraction routines from the VA EHR (Table 2). Nineteen comments were reports

of the same problem: an error in data conversion where active prescriptions were omitted. Two comments led to the identification of a programming stop code introduced when testing an update of the extraction routine that constrained the number of ICD9 codes extracted. One comment highlighted an instance where an adverse drug reaction (ADR) reaction to a substance was incorrectly labelled to a drug. Finally, one feedback comment led to the identification of an expired prescription being inadvertently labeled as active.

Table 2: Rarely occurring problems in data extraction or conversion detected by analysis of clinician feedback

Problem	Number of comments	Cause of problem
Active medications omitted from the prescription list	19	Method of updating a VISTA* medication status led to labeling active medications as inactive.
Incorrect adverse drug reaction (ADR) notification	1	Error in the data conversion, leading to the incorrect labeling of an ADR.
Omission of diagnoses (ICD-9 codes)	2	Programming stop code introduced into the extraction routine led to dropping of ICD-9 codes.
Expired medications appeared as active in the prescription list	1	VISTA* medication status flag led to inaccurate labeling of expired medications as active.

* Veterans Health Information Systems and Technology Architecture

Clinician feedback also identified areas where recommendations from ATHENA DSS benefited from additional clarification. For example, one provider commented that a display about specific drug contraindications was more beneficial than a general alert about the presence of a drug contraindication, so this change was implemented to improve the system.

Clinicians rarely contacted study staff by e-mail or phone about program problems. In the few cases in which they did, it was to report a technical malfunction (such as failure of any recommendations to appear, as would happen if the network connection was lost).

Tracking of ATHENA DSS program exceptions provided additional insight. In general, such tracking revealed very few program exceptions in ATHENA DSS function. Almost all of the exceptions reported a technical glitch not related to ATHENA DSS. Changes in drug formulary were forecasted and therefore a tracking system was implemented to alert the ATHENA team to new antihypertensives in the VA pharmacy formulary. As an example, the angiotensin-receptor blocker irbesartan was added to the formulary during the course of the trial. Thus, tracking of program exceptions facilitated maintenance of drug databases necessary for ATHENA DSS to make proper drug recommendations.

There were no additional programming, logic, or treatment errors identified by study providers. In all of the above examples, ongoing surveillance of system

performance ensured the continued delivery of accurate decision support recommendations.

Discussion

Failure to maintain a commitment to quality assurance has historically led to some devastating results. A presidential commission on the NASA space shuttle Challenger accident concluded that “a disproportionate reduction may have occurred in the safety, reliability, and quality assurance staff at NASA.... The decreases may have limited the ability of those offices to perform their review functions.”¹⁸ An authority on software development has argued that successful software development demands a similar commitment to quality assurance.¹⁹ It is essential to recognize that all human work is susceptible to error, and that even those working on systems aimed at lowering error rates must plan for anticipated errors in their own systems.

Clinical software systems such as decision support systems implemented to reduce medical errors pose similar challenges to maintaining quality assurance after deployment. Scaling up from a limited set of pilot users to a larger, more widespread audience and a larger number of patient record extracts may reveal unanticipated consequences, especially over a distributed network of clinicians. Clinicians may attempt to bypass time-consuming physician order entry systems by asking allied health professionals to write verbal or handwritten medication orders.² This may not be observed or anticipated in testing with clinician early adopters.

One must also consider the inevitable changes to a clinical decision support system or the integrated network of software to which it is connected. A clinical decision support system may evolve to incorporate new evidence regarding best treatment practices, as with upgrading ATHENA DSS for JNC 7. The software to which the clinical decision support system is connected may be upgraded or updated over time. An analysis of the Health Evaluation through Logic Processing (HELP) system at Latter Day Saints Hospital and its interconnected referral centers in Salt Lake City, UT, revealed 10^{24} possible software configurations.⁹ Implementing any changes after deployment, either to the decision support system itself or other software dependencies, can lead to new unanticipated errors.

Our approach to these challenges was to create several methods to monitor system accuracy. The free-text window on the recommendations screen promoted direct interaction between ATHENA DSS and its clinician-users. Despite extremely busy clinical workloads and limited time per patient visit, study participants interacted with the ATHENA DSS advisory in some fashion for 63 percent of the patients.²⁰ Nearly 50 percent of the study providers entered feedback comments during point-of-care use of ATHENA DSS. This level of participation is substantial, given other published reports of clinician interactions with decision support systems. One tertiary academic medical center reported that clinicians chose to interact with a guideline-based decision support system for hyperlipidemia in only 20 of 2,610 visits (0.8 percent).²¹ Clinician interaction

with ATHENA DSS proved to be crucial to our study. Surveillance of feedback data provided a sensitive method of detecting potentially important problems in data extraction and conversion.

Given the complex milieu of data sources on which a decision support system relies, it is not surprising that rarely occurring errors may not be detected despite rigorous predeployment testing. Clinician feedback submitted via the ATHENA DSS user interface and program execution tracking provided important facilities for monitoring quality assurance after deployment of the system. With such data, a small team of investigators monitored the interactions of 91 ATHENA DSS clinician-users distributed over three geographically dispersed sites for a 15-month study period. More important, this method provided an efficient means of detecting and subsequently correcting inaccurate recommendations based on rarely occurring problems. New technologies for improving patient safety should include mechanisms and funding for post-fielding surveillance such as point-of-care feedback and other monitoring of the system to prevent introduction of new errors into the clinical workflow.

Acknowledgments

The ATHENA project was supported by VA Health Services Research and Development Service grant CPI 99275. The EON project was supported by National Library of Medicine grant LM 05708. Dr Chan was supported by the National Library of Medicine grant LM 07033. Views expressed are those of the authors and not necessarily those of the Department of Veterans Affairs or other affiliated institutions.

Author affiliations

VA Palo Alto Health Care System, Palo Alto, CA (ASC, SBM, RWC, MKG). Stanford Medical Informatics, Stanford, CA (ASC, SWT, MAM, MKG). Durham VA Medical Center, Durham, NC (HBB, EZO). Boston VA Medical Center (BBH). Center for Primary Care Outcomes Research, Stanford, CA (MKG). Harvard Medical School, Boston (BBH), San Francisco VA Medical Center (MGS). University of California, San Francisco Department of Family and Community Medicine (ASC) and Department of Medicine (MGS).

Address correspondence to: Mary K. Goldstein, M.D., M.S.; VA Palo Alto Health Care System; GRECC 182B; 3801 Miranda Avenue; Palo Alto, CA 94304. Phone: 650-858-3933; e-mail: Mary.goldstein@med.va.gov.

References

1. Kohn LT, Corrigan JM, Donaldson MS, editors. *To err is human: building a safer health system*. A report of the Committee of Quality of Health Care in America, Institute of Medicine. Washington, DC: National Academy Press; 2000.
2. Bates DW, Cohen M, Leape LL. Reducing the frequency of errors in medicine using information technology. *J Am Med Assoc* 2001;8:299–308.
3. Goldstein MK, Hoffman BB, Coleman RW, et al. Patient safety in guideline-based decision support for hypertension management: ATHENA DSS. *J Am Med Assoc* 2002 Nov–Dec suppl;9:S11–S16.

4. Ash JS, Berg M, Coiera M. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 2004;11:104–12.
5. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc* 2003;10:478–83.
6. Haslet TK, Lee TA, Hansten PD, et al. Performance of community pharmacy drug interaction software. *J Am Pharm Assoc* 2001;41(2):200–4.
7. Rodriguez EM. Application of new methods to drug safety surveillance: beyond spontaneous reports. Rockville, MD: U.S. Food and Drug Administration Center for Drug Evaluation and Research; 1999. Available at: <http://www.fda.gov/cder/present/ispe-1999/evelyn/>.
8. U.S. Food and Drug Administration Center for Drug Evaluation and Reporting. MEDWatch Program. FDA/CDER. Available at: <http://www.fda.gov/cder/handbook/medwatch.htm>.
9. Miller RA, Gardner RM. Recommendations for responsible monitoring and regulation of clinical software systems. *J Am Med Inform Assoc* 1997;4:442–57.
10. Gould JD, Lewis CH. Designing for usability: key principles and what designers think. *Communications of the ACM* 1985;28(3):300–11.
11. National High Blood Pressure Education Program. The sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Washington, DC: National Institutes of Health; 1997.
12. Veterans Health Administration, Department of Defense. Diagnosis and management of patients with hypertension in the primary care setting. Washington, DC: VHA/DoD; 1999 May.
13. Musen MA, Tu SW, Das AK, Shahar Y. EON: a component-based approach to automation of protocol-directed therapy. *J Am Med Inform Assoc* 1996;3:367–88.
14. Musen MA, Gennari J, Eriksson H, et al. Protege II: computer support for development of intelligent systems from libraries of components. Medinfo 95, The Eighth World Congress on Medical Informatics, Vancouver, Canada; 1995.
15. Advani A, Tu SW, O'Connor M, et al. Integrating a modern knowledge-based system architecture with a legacy VA database: the ATHENA and EON projects at Stanford. *Proc AMIA Symp* 1999;653–7.
16. Goldstein MK, Hoffman BB, Coleman RW, et al. Implementing clinical practice guidelines while taking account of evidence: ATHENA, an easily modifiable decision-support system for management of hypertension in primary care. *AMIA Annual Symposium*, Los Angeles; 2000. pp. 303–4.
17. Goldstein MK, Hoffman BB. Hypertension recordkeeping and electronic management systems. In: Izzo JL and Black HR, editors. *Hypertension primer: the essentials of high blood pressure*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2003. pp. 393–6.
18. Report of the Presidential Commission on the space shuttle Challenger accident; 1986. Available at: <http://science.ksc.nasa.gov/shuttle/missions/51-1/docs/rogers-commission/table-of-contents.html>.
19. McConnell S. *Software project survival guide*. Redmond: Microsoft Press; 1998.
20. Goldstein MK, Coleman RW, Tu SW, et al. Translating research into practice: sociotechnical integration of automated decision support for hypertension in three medical centers. *J Am Med Inform Assoc* 2004 Sep–Oct;11(5):368–76.
21. Maviglia SM, Zielstorff RD, Paterno M, et al. Automating complex guidelines for chronic disease: lesson learned. *J Am Med Inform Assoc* 2003;10:154–65.

