

## Desiderata for a Computer-Assisted Audit Tool for Clinical Data Source Verification Audits

Stephany N Duda, Firas H Wehbe, Cynthia S Gadd

*Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA*

### Abstract

*Clinical data auditing often requires validating the contents of clinical research databases against source documents available in health care settings. Currently available data audit software, however, does not provide features necessary to compare the contents of such databases to source data in paper medical records. This work enumerates the primary weaknesses of using paper forms for clinical data audits and identifies the shortcomings of existing data audit software, as informed by the experiences of an audit team evaluating data quality for an international research consortium. The authors propose a set of attributes to guide the development of a computer-assisted clinical data audit tool to simplify and standardize the audit process.*

### Keywords:

Data auditing, Data quality, Software design

### Introduction

Analyzing low quality study data produces meaningless results, which is why interventional clinical trials focus so heavily on data quality control [1]. Many international, multi-center research networks that pool and analyze observational data, however, do not report a similar emphasis on data quality assurance. Without methods to assess and improve data quality, studies using the resulting observational databases may generate false research conclusions based on unreliable information.

Data auditing is a proven method of assessing the quality of routine clinical care data that have been reused for research[2]. Unfortunately, most verification audits of clinical data use paper audit forms, which have been shown in general to be less effective and efficient than electronic tools [3, 4]. This work aims to identify the core weaknesses of paper forms when used for clinical data auditing, as motivated by a series of data monitoring visits to seven clinics participating in an HIV observational research network. The authors propose a set of functional requirements for a computerized audit tool that may simplify the audit process and encourage research networks to measure and improve the quality of their data.

Auditing is an established technique for evaluating and improving the quality of products, services, or information, and

has been a staple of quality control activities for over a century [5, 6]. Audits take many different forms depending on the domain: in accounting, they identify fraud; in manufacturing, audits help assess both the quality of a product lot and the producer's compliance with Good Manufacturing Practice; and in information security, audits allow for the inspection of the security and reliability of computer systems and of the information they contain [7, 8]. In medicine, researchers have employed audit techniques to detect inconsistencies in terminologies, evaluate the quality of patient care and verify that medical services are properly documented, coded, and billed [9-11]. The U.S. Federal Drug Administration also requires auditing of many clinical trials to ensure that the operators of the trial are properly monitoring patient safety, accurately recording data generated by the study, and adhering to the study's protocol and Good Clinical Practice [12, 13].

Protocol-driven studies such as clinical trials often engage teams of clinicians and data managers to perform research data audits. These auditors compare research data to the source documentation, which often includes paper clinical charts, laboratory reports, or the contents of electronic medical record and laboratory systems at the study sites. Most tools to support clinical data audits are paper forms with lists and checkboxes, with various examples freely available online [10, 14, 15].

Although paper-based audits are still common in medicine, computer-assisted audit tools (CAATs) have improved the quality of audits in finance, manufacturing, and IT security by facilitating more thorough audits, generating more consistent documentation, and saving both time and money for auditors and auditees [16, 17]. CAATs can aid auditors during many stages of the audit process, from merging and analyzing data to generating audit reports. Auditors often use statistical or data extraction software as a CAAT in order to detect anomalous patterns in large data sets. Other software packages specifically designed for auditing (e.g., Audit Command Language (ACL) and Interactive Data Extraction and Analysis (IDEA)) even assist auditors in selecting the audit sample size and audit methodology [18, 19]. Each single-user ACL or IDEA license provides access to powerful data analysis tools, but also costs thousands of dollars. Less expensive audit-specific software includes TopCAATs, a Microsoft Excel audit plug-in, and Picalo, a Python-based, open-source data analysis and fraud

detection toolkit [20, 21]. Most of the advanced CAATs also require computer programming skills.

Unfortunately, these software packages focus on analyzing an existing, electronic dataset for errors and unusual patterns, rather than facilitating the comparison between the dataset and a physical source document. Indeed, in many accounting and security audits, the electronic database *is* the source document and no other records exist. As a result, these advanced software packages are not helpful for auditing paper source documents. Furthermore, the cost of tools such as ACL and IDEA makes purchasing them unfeasible in resource-limited settings.

## Motivation

### Data Audits at Seven HIV Clinics

The Caribbean, Central and South America Network for HIV research (CCASAnet) is one of seven collaborative research groups participating in the International Epidemiologic Databases to Evaluate AIDS (IeDEA) [22]. CCASAnet brings together researchers from HIV clinics in Argentina, Brazil, Chile, Haiti, Honduras, Mexico, and Peru to create an HIV observational database using routine patient care data. The project's data coordinating center (DCC) at Vanderbilt University conducts Good Clinical Practice-based audits on datasets submitted by CCASAnet member sites to identify sources of error in data collection, abstraction, and representation, and help the DCC determine the structure, quality, and reliability of the submitted data.

Between March 2007 and April 2008, a two- or three-person team including at least one physician and one informaticist visited each member clinic to compare the on-site medical documentation to the contents of the electronic database that the site previously submitted for analysis. The audit team used a multi-page paper audit form to record the results of the database-to-clinical record comparison. Data on the form were divided into categories such as demographics, clinical visit data, antiretroviral regimens, and laboratory results.

Two items per data element were preprinted on the form: the variable name (e.g., birth date, date of death, viral load result), and the database value for that variable. The team used the blank "audit value" field to record whether a data element was present in the source documents and whether the source value was correctly represented in the database. A small notes field – as well as the margin of the paper – was used to record additional information or possible causes of the error. Figure 1 shows a sample page of a completed audit form.

At the end of an audit visit, the auditors presented their preliminary findings during an exit interview with the site investigator and staff. After returning to the DCC, the audit team produced a report describing its findings and recommendations, which was sent to the site for review and comment.

The audit team inspected 184 records and 4223 unique data elements during seven audits. The average time between the end of an audit and the completion of the audit report was 101 days which meant the site data personnel rarely received im-

mediate, implementable recommendations on how to improve data quality.

Clinical Information			
Variable	Value	Audit Value	Notes
key	1049	✓	
Date of Birth	9/7/1970	✓	
Start of ARV Regimen	7/26/2003	✓	
Sex	M	✓	
Site	Hospital [redacted]	✓	
id	397	✓	
Stage	C	✓	
Date of Death		NA	
Cause of Death		NA	
Last Visit Date	1/26/2008	✓	Additional photo
Weight Date	---	---	
Baseline Weight	---	---	
Risk	Sexual	✓	

Figure 1- A neatly completed form from a CCASAnet site audit.

### Benefits of a Computer-Assisted Audit Tool

Feedback on data quality is most effective when it is communicated shortly after the audit takes place, but the use of paper audit forms makes generating reports a challenge [6]. In post-audit debriefings, the auditors identified several causes for delays in producing the audit report, including difficulties with

- handling multiple audits and reports simultaneously,
- reading other auditors' handwriting,
- interpreting underspecified notes without the presence of the source documents,
- deciding how to handle partially completed audit forms,
- classifying errors during post-visit audit form reviews,
- assessing whether an error was clinically meaningful,
- consulting with other auditors about error classifications or unclear information,
- sharing a single set of original audit forms among a team of auditors,
- tabulating errors,
- double-checking other auditors' error tables,
- calculating error rates,
- composing a thorough and detailed audit report, and
- formatting error tables for the final document.

A computer-assisted audit tool that replaces the paper forms could help standardize the audit process and increase the timeliness and reproducibility of audit results. A CAAT for clinical data auditing could guide users through the process of importing their study data, selecting records to audit, recording and categorizing data discrepancies, and generating audit results. The audit findings would be immediately available in an electronic format that could be used to generate tables or to feed corrected data back into the source database.

## Key Attributes of an Audit Tool

To improve the audit process, computer-aided software for source document verification should provide flexible functionality in five main areas: networking, audit data management, error categorization, audit decision support, and results reporting. The audit team's experiences that motivated these requirements are described in the text. The requirements are outlined as desiderata in Table 1.

### Requirement 1: Networking

Paper forms functioned as an excellent sharing tool during audits. Although each auditor worked independently on a set of records, difficult cases or records with cascading errors often required a group review in which the source document and audit form were passed around the table. A suitable CAAT should facilitate the same real-time communication between multiple auditors and allow collaborative editing of a shared database, in settings where Internet access is not guaranteed. CCASAnet data audits, for example, often take place in the record storage or meeting rooms of clinics in resource-limited settings, where auditors work collaboratively on complex records. An effective paperless audit tool needs to accommodate multiple users manipulating a single copy of the data. However, because of the unreliability of local network connections, a useful tool should take advantage of portable routers or alternate network structures, such as wireless ad hoc networks between auditor laptops.

### Requirement 2: Audit Data Management

Preparing paper audit forms in advance of each audit was a laborious, multi-day task for the audit team and the CCASAnet data manager. An audit tool that allowed users to import preformatted datasets could reduce the preparation time significantly. A standard XML data specification would permit auditors to load a copy of the audit data, as provided by the study data manager, before the audit begins. A standardized import/export format also allows the audit results to be routinely converted for use in statistical software packages.

Although the ongoing CCASAnet audit program evaluates the accuracy and completeness of HIV-related clinical and laboratory data, the proposed audit tool should be able to accommodate datasets with different medical content, such as data collected for studies of cancer or tuberculosis. The software's internal representation of the audit data, therefore, must be flexible enough to adapt to different content types. Such types include not only standard data representations, such as integers, character strings, or Boolean variables, but also the frequent irregular data forms that the audit team catalogued, such as malformed, partial, and approximate dates, miscoded values, and integer variables with character content (e.g. "<400").

### Requirement 3: Standardized Assessment of Errors

CCASAnet's paper-based audit process relied heavily on memory, interpretation, and opinion, and was difficult to replicate and standardize across sites. Indeed, when the DCC un-

dertook a reevaluation of the audit findings in mid-2008, the authors had a difficult time applying a standard error categorization system devised for the task, as the original paper forms had required auditors only to describe errors, not to classify them according to a formal error taxonomy.

A robust audit tool should assist auditors in assessing and classifying errors during the audit visit, rather than weeks afterward when the source documents are no longer accessible. The auditor should be able to import the most appropriate error categorization scheme for a given audit task, as errors in HIV data, for example, may be distinct from errors in cancer or tuberculosis data. The tool should also accommodate complex error classifications that prompt auditors to evaluate data errors on multiple axes, including the type of error, the severity of the error, the clinical relevance of the error, and the probable direct cause.

### Requirement 4: Audit Decision Support

Selecting the number and type of records to audit is a challenge for novice auditors. The audit team consulted a statistician in advance of each audit, but a useful CAAT could provide basic guidance on sample size calculations and selecting records for audit, using selection metrics that have been described in the literature.

The system should also allow the auditor to import an optional rule set that would guide the automatic classification of errors based on the data class and error type. A mismatched weight value of 37.5kg in the clinical record vs. 38kg in the database, for example, is likely to be a rounding error of limited clinical significance. This functionality could be useful in labeling complicated errors of drug prescription and discontinuation.

### Requirement 5: Results Reporting

The CCASAnet team found that preparing a post-audit report from paper audit sheets required time-consuming counting, description, tabulation, and confirmation of all the data discrepancies. Both auditors had to count the variables in each record, group any recurring errors, double check final numbers, and tabulate the results manually, which delayed preparing the final report. Software support for generating tables, graphing data, and displaying trends based the results of the audit could simplify the post-audit work and help auditors detect patterns of errors across records or data falsification that might be overlooked in manual review. Furthermore, such tools could help auditors adjust their sampling of variables during auditing to focus their efforts on evaluating variables that appear more prone to error.

By automatically generating tables of error rates and lists of specific errors, a computerized audit tool would assist auditors in preparing a summary for the exit interview and the final post-audit report. Audit support software should also provide basic quality improvement suggestions for the site based on patterns of error in the audit data and established user-imported rules.

Table 1- Obstacles encountered during clinical data audits and corresponding desiderata for computer-assisted audit tools.

<b>Obstacles Encountered during Audits</b>	<b>Solutions / Desiderata for a Computer-Assisted Audit Tool</b>
<b>Challenge: Collaboration</b>	<b>Solution: Networking</b>
Auditors need to work collaboratively on the same copy of a record.	Real-time Collaboration: networked laptops for auditors, shared databases, web-based systems
Audit sites may have no network infrastructure.	Portable Network Infrastructure: peer-to-peer networking, portable server and router
<b>Challenge: Audit Data</b>	<b>Solution: Audit Data Management</b>
Paper audit forms take a long time to prepare and validate.	Import Functionality: one-click import of data and data descriptions (metadata) from research database to CAAT, instant generation of basic electronic audit forms
Copying audit results from paper forms into a spreadsheet for analysis is time-consuming.	Export Functionality: export of audit results into structured data formats such as XML
Datasets may contain different medical content (e.g., HIV, Tuberculosis, or cancer data).	Metadata Management: customizable import interface, customizable display of data on screen, data dictionaries for special topic areas (HIV, TB, Cancer)
Data may violate syntactic rules; auditors may need to record corrected values.	Reasoning About Data Types: representing simple and complex data types, data syntax rules, codification of mismatch between research data and native records, handling malformed data
<b>Challenge: Types of Errors</b>	<b>Solution: Standardized Assessment of Errors</b>
Errors are not categorized and described clearly on paper forms, making it difficult to analyze and report error types and rates.	Representation of Error Types: hierarchical ontology of errors, clear operational descriptions of error types, specification of domain of error types (applies to specific variables within the audit record or applies to entire record), specification of error labels and default values and whether closed or open world assumptions apply
Auditors discover new and unexpected types of errors during the audit process.	Error Scheme Evolution: support for versioning and collaborative authoring of error schemes, interface to edit error schemes while in use
Some audits require different error classification schemes that are better suited to the data.	Error Scheme Management: storage, import, and export of audit-specific error terminologies
<b>Challenge: Audit Design</b>	<b>Solution: Audit Decision Support</b>
Auditors are unsure how many records should be audited to produce meaningful results.	Statistical Dashboard: guidance for sample size calculations, identification of grossly problematic records, pre-selection of records via statistical sampling
<b>Challenge: Analyzing and Presenting Results</b>	<b>Solution: Results Reporting Tools</b>
Tallying and tabulating errors by hand is a time-consuming and error-prone task for auditors.	Automatic Report Generation: software support for generating tables and graphs
Manual approaches may miss subtle patterns of data error.	Real-time Trend Detection: automatic checks for patterns of error suggesting data falsification, systematic errors

## Discussion

When an audit process lacks standardization, different auditors may produce different audit reports given the same source record. Without standardized quality measures, the changes in an organization's data quality cannot be compared from year to year, nor can audit results be compared from site to site.

Flexible audit support tools that have the attributes described will simplify and standardize auditors' work, thereby increasing an audit's potential benefit.

The CAAT recommendations described here stem from audit experiences at HIV clinics in Central America and the Caribbean. The perspective of a single international research network, however, may limit the diversity of experiences that

informed the five suggested requirements. Indeed, these recommendations may represent a necessary but not sufficient set of features needed for a successful paperless audit tool.

Future work will focus on developing a prototype CAAT for audits of HIV, tuberculosis, and cancer data. If the tool proves to be useful within CCASAnet, other similar networks could use it to evaluate the quality of their data, standardize their quality control activities, and identify areas for process improvement.

## Conclusion

The quality of routine clinical care data should be assessed before such data are included in observational databases. Current paper-based audit techniques can be both inefficient and inconsistent. Computer support tools may be able to simplify and standardize the preparation for and execution of a source document audit, if certain criteria for such a tool are met.

## Acknowledgements

This research was supported by NIH Cooperative agreement 5 U01 AI069923-04 (CCASAnet).

## References

- [1] Vantongelen K, Rotmensz N, Van Der Schueren E. Quality control of validity of data collected in clinical trials. *European Journal of Cancer and Clinical Oncology*. 1989;25(8):1241-7.
- [2] Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. *J Am Med Inform Assoc*. 1996;3(3):234-44.
- [3] Hanscom B, Lurie JD, Homa K, Weinstein JN. Computerized questionnaires and the quality of survey data. *Spine*. 2002 Aug 15;27(16):1797-801.
- [4] Ryan JM, Corry JR, Attewell R, Smithson MJ. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Qual Life Res*. 2002 Feb;11(1):19-26.
- [5] Dicksee LR, Montgomery RH. *Auditing; a practical manual for auditors*. Authorized American ed. NY 1905.
- [6] Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Medical care*. 2009 Mar;47(3):356-63.
- [7] Probitts J. Auditing in the manufacturing environment. *The Quality Assurance Journal*. 2000;4(4):193-6.
- [8] Dark M, Poftak A. How to Perform a Security Audit. *Technology and Learning*.24(7):18-27.
- [9] Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *J Am Med Inform Assoc*. 1998 Jan-Feb;5(1):41-51.
- [10] Grider D. *Medical Record Auditor: Documentation rules and rationales with exercises*. USA: American Medical Association 2008.
- [11] Johnston G, Crombie I, Alder E, Davies H, Millard A. Reviewing audit: barriers and facilitating factors for effective clinical audit. *Quality in Health Care : QHC*. 2000;9(1).
- [12] ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6 (R1); 1996.
- [13] Weiss RB. Systems of protocol review, quality assurance, and data audit. *Cancer Chemotherapy and Pharmacology*. 1998;42(0):S88-S92-S88-S92.
- [14] Hirayama K, Fukuda N, Satoh H, Itoh K, Chiba K, Nakae Y, Takezawa M, Gotoh K. Checklist for GCP compliance investigation (Medical institution). *The Quality Assurance Journal*. 2005;9(2):120-39.
- [15] DF/HCC Quality Assurance Office for Clinical Trials. *Clinical Trials Audit Manual*. 2008 [last accessed 2009 15 Oct]; <http://www.dfhcc.harvard.edu/clinical-research-support/quality-assurance-office-for-clinical-trials-qact/forms-policies-and-manuals/>
- [16] Abdolmohammadi M, Usoff C. A longitudinal study of applicable decision aids for detailed tasks in a financial audit. *International Journal of Intelligent Systems in Accounting, Finance & Management*. 2001;10(3):139-54.
- [17] Guidance Note on Computer Assisted Audit Techniques (CAATs). *The Chartered Accountant*. (Jan 2004):731-7.
- [18] Audimation Services I. IDEA Data Analysis Software. 2007 [last accessed 2009 15 Oct]; <http://www.audimation.com>
- [19] ACL Audit Analytics and Continuous Monitoring Software Solutions. 2009 [last accessed 2009 11 Mar]; <http://www.acl.com/>
- [20] Picalo: Data Analysis and Fraud Detection Toolkit. [last accessed; <http://www.picalo.org/>
- [21] TopCAATs - Audit Reinvented. [last accessed 2009 13 Mar]; <http://www.topcaats.com>
- [22] McGowan CC, Cahn P, Gotuzzo E, Padgett D, Pape JW, Wolff M, Schechter M, Masys DR. Cohort Profile: Caribbean, Central and South America Network for HIV research (CCASAnet) *International journal of epidemiology*. 2007 Oct;36(5):969-76.

## Address for correspondence

Stephany N. Duda, [stephany.duda@vanderbilt.edu](mailto:stephany.duda@vanderbilt.edu)