

# COMPUTER-ASSISTED AUDITING FOR HIGH VOLUME MEDICAL CODING

*Posted on July 20, 2009 by Administrator*

**Category:** [Conference Paper](#)

**Tags:** [coding](#), [Computer assisted coding](#)

by Daniel T. Heinze, PhD; Peter Feller, MS; Jerry McCorkle, BA; and Mark Morsch, MS

## **Abstract**

The volume of documents being processed by computer-assisted coding (CAC) has raised the bar regarding the need for audit methods suitable for production control and quality assurance. In this high-volume production environment, it becomes vitally important to adapt and implement techniques that have become a fundamental requirement for production operations management (POM). We present techniques and statistical methods that are developed and implemented for auditing the medical coding process and producing scores that accurately reflect the quality of the coding work, are comparable across time and between coders and auditors, and employ statistical methods for production control. The techniques and methods here described are patent pending and are implemented in the A-Life Medical, Inc. CoAudit™ system that is commercially available for auditing both computerized and human coding.

## **Introduction**

The advent of CAC in high-volume environments demands the use of modern statistical production control and QA methods. Traditionally, coding has been done “manually” by human coders. Because the volume of medical documents being manually coded at any one location has been relatively small, quality assurance (QA) has primarily depended on the individual skills, training, and continuing education of the coders. In the field of medical coding, QA methods historically consist of an ad hoc review of some fixed number or percentage of the coders' work product with ad hoc or subjective scoring and evaluation of audit results. Audit results across time and between coders are, therefore, not mathematically comparable. Additionally, these methods do not scale to high-volume processing.

Although some aspects of QA and production control can be handled automatically, there is still the need for human audit of the coding work product. However, coding is a complex matter, and for some significant percentage of medical documents there will be a measurable diversity of opinion as to how they ought correctly to be coded. Further, the process is sufficiently complex that even the auditors are expected to make errors, though presumably at a lower error level than the coders or processes that are being audited. Consideration for both matters of opinion (subjective judgment) and error must be taken into account when devising a medical coding audit methodology.

## **Background**

Faced with the problem of providing clients using the Actus® CAC application with a means to audit codes and perform production control, we embarked on a process of first soliciting client input and then developing consensus on a specification limit methodology for scoring around which we

developed and/or applied the necessary methodologies for the following research issues.

## Research Questions

Beginning from the description of how coding professionals audit one another, we address the following issues:

- *Sample Selection*: Calculating the sample size for audits
- *Specification and Control Limits*: Establishing and interpreting
  - *Specification limits* that measure the acceptability of individual coded documents with a method for audit scoring that produces results suitable for incorporation in statistical QA and production control, but which are also designed so that the composite sample scores track the subjective judgment of human auditors when evaluating a computerized or human coding process to be acceptable, marginally acceptable, or unacceptable
  - *Control limits* that measure the acceptability of a computerized or human coder
- *Calibrating for Auditor Variability*: Adjusting the statistical methods to calibrate for auditor subjectivity and error so audit results can be meaningfully compared across time and between auditors, coders, and CAC

## Methods

Corresponding to the research questions, the following methods are employed.

**Sample Selection.** Sample selection is governed first by identifying the population from which the sample will be drawn, second by applying some statistical method to determine the sample size, and third by selecting a random sample of the determined size from the population.

For purposes of code auditing, the population must be selected in accord with the objectives of the audit. Audit objectives should first be specified in terms of the target computerized or human coder to be audited with the population being then limited to codes produced by the target. The population may further be stratified according to subcharacteristics such as particular providers, procedures, or diagnoses. It may further be necessary to temporally limit the population to some period when the codes and guidelines for the audit objective were uniform.

Given a target population, the sample size must be calculated. We accept the sample size calculation employed by the Office of the Inspector General (OIG) and as described and implemented in the audit tool Rat-Stats as canonical for code audit purposes.<sup>1</sup> Considering the Rat-Stats function for Attribute Sample Size selection, we note that although the calculated sample size is guaranteed to be minimal, the confidence interval may be asymmetrical around the point estimate. With no harm to the accuracy or validity of the audit, we use a more basic calculation of sample size as given by many introductory statistics texts and also at the National Institute of

Standards and Technology, resulting in a symmetric confidence interval at the possible cost of a slightly larger sample size.<sup>2</sup>

**Specification and Control Limits.** Two sets of performance limits are defined, the specification limits and the control limits. The specification limits are with respect to individual components of the production items under test. These can be judged as either correct or incorrect (pass/fail), and if incorrect (fail) then, optionally, as either of consequence or not of consequence. The control limits are the statistically defined limits that indicate whether the overall coding process under audit is in control or not. For the medical coding application, only the upper control limit is of true interest in that there is no adverse consequence if the process, as measured in terms of proportion of errors, falls below the lower control limit (in fact that is a good thing and indicates that the process is performing better than required or expected).

[Formulas—Sample Size and Control Limits](#)

## Calibrating for Auditor Variability

An initial CV can be established by making an educated estimate of the auditor's accuracy, but auditors should be periodically tested to provide a benchmark CV value. Without this calibration, audit results across time and between auditors will not be meaningfully comparable. The objective of the testing is to track the CV value of each auditor across time using standardized benchmark tests. The benchmark test consists of a set of coded documents for the auditor to audit. The benchmark test must conform to three principles:

1. From one test session to the next, a significant portion of the test (at least 50 percent in the preferred implementation) must consist of the same documents with the same codes as were present on the previous test. The remaining documents will be new. In the preferred implementation, the order of the documents from test to test will be randomized.
2. Over time, the test documents must be selected so as to reflect the distribution of encounter and document types that coders would be expected to work with under actual production conditions.
3. Test sessions must be separated by sufficient time and test size must be sufficiently large in order that auditors would not reasonably be expected to remember a significant percentage of their edits from one test session to the next.

Auditor scores on the benchmark tests consist of two parts. First, determine CV as calculated on the recurring documents from one test session to the next. Second, the relative variances between auditors who take the same test are calculated and may be used as a cross-check on the intra-auditor CV variance.

## Results and Discussion

The CoAudit methodology is designed to correlate to the qualitative judgment of human auditors who may judge a coding process, as defined by the sample selection parameters, to be acceptable, marginally acceptable, or unacceptable. As such, the results of an audit are meaningful primarily when represented as a time series against the system control limits as in an X-bar chart. Because standards of acceptability may vary with time and between organizations, empirical tests must be performed periodically for calibration purposes, but  $P=0.1$ ,  $H=0.02$  and  $CV=0.03$  are recommended starting parameters. A process may have sample scores that are consistently in control (acceptable), occasionally out of control (marginally acceptable), or consistently out of control (unacceptable). As a starting point, monthly audits are recommended with more than two sample scores out of control in a year being considered unacceptable and requiring intervention to bring the system back in control. Statistical significance tests (e.g. -square) can be used to measure the effectiveness of interventions. At the time of writing, alpha tests on six coders indicate that the initial objectives have been met, but minor adjustments are expected as a result of beta testing.

Daniel Heinze, PhD, is the Chief Technology Officer of A-Life Medical, Inc. in San Diego, CA.

Peter Feller, MS, is a Senior Software Engineer at A-Life Medical, Inc. in San Diego, CA.

Jerry McCorkle, BA, is the Vice President of Client Services & Operations of A-Life Medical, Inc. in San Diego, CA.

Mark Morsch, MS, is the Vice President of NLP & Software Engineering at A-Life Medical, Inc. in San Diego, CA.

## Notes

1. Department of Health and Human Services – Office of Inspector General Office of Audit Services. *Rat-Stats Companion Manual*. September 2001. Available online at <http://oig.hhs.gov/organization/OAS/ratstats/ratstatsmanual.pdf>, last accessed July 18, 2006.
2. National Institute of Standards and Technology. *Engineering Statistics Handbook*. June 2003. Available online at <http://www.itl.nist.gov/div898/handbook/>, last accessed July 18, 2006.

Article citation: *Perspectives in Health Information Management*, CAC Proceedings; Fall 2006

[Printer friendly version of this article](#)

**There are no comments yet.**