

# Development and Use of Automated Coding Software to Enhance Antifraud Activities

*by Jennifer Hornung Garvin, PhD, RHIA, CCS, CTR, CPHQ, FAHIMA; Valerie J.M. Watzlaf, PhD, RHIA, FAHIMA; and Sohrab Moeini, BSIS*

Completed under contract number HHSP23320054100EC. This report was prepared by the Foundation for Research and Education (FORE) of the American Health Information Association (AHIMA) and the University of Pittsburgh under contract with the Department of Health and Human Services, Office of the National Coordinator for Health Information Technology, and was submitted on July 11, 2005.

## Abstract

The deliberate submittal of false claims to private health insurance plans and tax-funded public health insurance programs is a serious and increasingly nationwide crime occurrence. This descriptive, exploratory study examined automated coding software and how it could enhance antifraud activities, detect errors, increase the accuracy of coded data, and detect false claims. The methods used to undertake this research included a review of the literature, interviews with federal agencies, use of a product information form, and interviews with vendors and users of automated coding and antifraud software. The research helped develop product information, flowcharts, an automated coding impact table, a table describing weak links in fraud and abuse software, and an antifraud model. The study concluded that fraud can be mitigated with appropriate technology, fraud prevention and detection processes, and ongoing educational efforts.

## Introduction and Background

The deliberate submittal of false claims to private health insurance plans and tax-funded public health insurance programs, such as Medicare and Medicaid, is a serious and increasingly nationwide crime occurrence. In 2003 alone, the National Health Care Anti-Fraud Association (NHCAA) estimates that at least 3 percent of the nation's healthcare expenditures, or \$51 billion, was lost to outright fraud. Other estimates by government and law enforcement agencies place the loss as high as 10 percent of our annual expenditures, or \$170 billion each year.<sup>1</sup> According to the Centers for Medicare & Medicaid Services (CMS),<sup>2</sup> fraud may take different forms, such as incorrect reporting of diagnoses or procedures to maximize payments, fraudulent diagnosis, and billing for services not rendered. This study examined automated coding software as an evolving technology and described it across healthcare settings and patient types as well as gauging its ability to reduce fraudulent activities. This study also looked at how automated coding software can help a healthcare organization enhance antifraud activities, detect errors, increase the accuracy of coded data, and detect false claims.

---

## Research Question and Objectives

The primary research question asked whether automated coding software can reduce fraudulent activities. This research project aimed to identify the characteristics of automated coding systems that have the potential to detect improper coding, identify the components of the automated coding process that could minimize improper or fraudulent coding practices and to relate them to the role of the electronic health record (EHR), and develop recommendations for software developers and users of coding products to maximize antifraud practices.

## Methodology

This descriptive study consisted of several segments beginning with a review of literature on automated coding software, antifraud software within automated coding systems, and the extent of fraud and abuse related to automated coding. This was followed by interviews with federal agencies to gather information about instances of improper reimbursement or potential fraud involving automated coding software. Next, a description of products was developed based on a product information form completed by vendors. Researchers then interviewed vendors and users of automated coding and antifraud software to ascertain how they used these products.

## Results

The study resulted in the development of product information matrices on the use of antifraud software and automated coding software across healthcare settings, the cost of these systems, and the use of coding optimization software and other coding tools across healthcare settings. Flowcharts were also created to demonstrate how automated coding and antifraud software tools are used with and without automated coding. The research team designed a table that summarizes the impact of automated coding tools on coding and billing accuracy. The team evaluated weak links in fraud and abuse during the coding process and presented these issues in tables. Based on the findings, researchers developed an ideal antifraud model that summarizes features, processes, and staffing. Please see the full report at the Office of the National Coordinator Web site at <http://www.hhs.gov/healthit/documents/AutomaticCodingReport.pdf>.

The study specifically found that code sets typically include the codes and code descriptions and the rules, conventions, and guidelines for proper use of the codes within them. However, payers do not always abide by such standards for proper application of the medical code sets. To read more on the complexity of the coding process, please see "Internet Resources for Accurate Coding and Reimbursement Practices," AHIMA Practice Brief, 2004.<sup>3</sup> As discussed in the AHIMA 2002 Position Statement on Consistency of Healthcare Diagnostic and Procedure Coding,<sup>4</sup> coded clinical data is used by many entities (healthcare providers, payers, researchers, government agencies, and others) for measuring many things (the quality, safety, and efficacy of care; managing care and disease processes; tracking public health and risks; providing data to consumers regarding costs and outcomes of treatment options; designing payment systems and processing claims for reimbursement; conducting research; epidemiological studies and clinical trials; designing healthcare delivery systems and monitoring resource utilization; identifying fraudulent practices; and setting health policy).

Researchers found that existing software can generate codes from electronic text. As in manual coding, errors can occur and fraud can be perpetrated. Automated coding can speed turnaround time for billing. Products vary and can have different names, but two common names are computer-assisted coding (CAC) and automated coding.<sup>5</sup> CAC is best defined as the use of software that automatically generates a set of medical codes for review, validation, and use based upon clinical documentation

---

provided by healthcare practitioners. Automated coding assignment differs from the manual process of coding in a significant way—it evaluates electronic text and determines the initial codes, rather than having a human user (coder or practitioner) assign codes from the start. However, in both methods the final determination of codes reported or stored should be done by a coding professional.

When combined with the electronic health record (EHR) or electronic documents, automated coding can streamline the way that healthcare organizations gather data and submit claims for services. It can help organize work and make documents easier to find. It can also provide a way to analyze health data and coding patterns to perform continuous auditing prior to billing and claims submission. Automated coding is commonly used in settings where there is limited variability of documentation such as when performing endoscopies; in the emergency, outpatient surgery, and radiology departments of a hospital; and in specialty physician offices.

There is a minimal number of coding software programs that code inpatient documents and these programs are not yet widely used. In primary care settings, the creation of text can be mapped to associated codes for physician validation. Software companies are rapidly responding to the marketplace and are planning to expand into new areas as natural language processing (NLP) coding engines become more familiar with the more complex clinical and surgical scenarios. In the short time since the development of the AHIMA Practice Brief “Delving into Computer-assisted Coding” (AHIMA, 2004),<sup>6-7</sup> there have been additional developments. These include the evolution of the use of statistics-based NLP or rules-based NLP used alone in an automated coding product, and the use of a combination of both methods in many products. Further, the majority of automated coding software companies interviewed reported that coded data is reviewed by qualified coding staff prior to use in the billing process. Of note, all parties interviewed discussed the continuing need for training coding professionals to evaluate and validate coded data. For the most part, users reported that the processes utilized by automated coding can enhance workflows so that coding staff can be better utilized.

Automated coding products can incorporate patient data generated from a variety of sources and analyze it. It can also evaluate record-specific information. Both of these aspects can help prevent fraud in reimbursement claims. It should be noted that some basic text-to-code-mapping products may not provide antifraud features and may contribute to it if not properly designed. The sophistication of the antifraud tools and software varies across products and can include basic tools, such as post-payment audits, or more complex data-mining techniques and machine learning. An example of the latter is artificial neural networks (ANNs).

ANNs can predict the potential for fraud in a specific claim based on the data in the claim and in the EHR. ANNs do not need constant updating, but rather continuously learn by analyzing certain pieces of information. Much like the text analytics in NLP, the medical data in ANNs is analyzed for any given claim and provides a statistical estimate that the data will either match or not match desired output. Training the system to detect fraud is improved by using examples of fraudulent cases. Once this is completed, the system uses its prior knowledge to determine whether a medical claim or data is falsified. These systems can be used for both prepayment and post-payment fraud detection.<sup>8-9</sup> Three mechanisms that help the ANN system deal with fraud detection include data profiling, advanced analytic models, and rank scoring.

Advanced analytical models that perform pattern recognition are also used in ANN systems. The data is compared to multiple sources of information to eventually try to find patterns that may suggest possible fraud or abuse. Since the antifraud software uses a combination of the systems described above, it continues to learn about the characteristics and patterns of legitimate and illegitimate claim behavior, becoming more intelligent and increasingly accurate in its detections over time.<sup>8-12</sup>

Just as there is a range of automated coding products, there is also a range of EHR products, from basic to sophisticated. In the primary care setting, there are software programs that suggest potential codes as patient records are generated. The practitioner often must select or validate the appropriate

code(s). In this model, there may also be edits or prompts to help the practitioner select the correct code. The code is not automatically assigned without validation and there are limited, if any, antifraud algorithms. The capabilities to combat healthcare fraud are possible when several types of technology are used together. Automated coding with NLP (rules-based and statistics-based combination) combined with ANNs and predictive modeling to detect fraud within an EHR is ideal. However, audit trails are also vital in order to continue to assess the patterns of use within the EHR as well as the patterns of coding and billing.

## **Conclusions**

This research provides recommendations for software developers, users, payers, consumers, and government agencies. The following is a summary of major recommendations. Please see “Detailed Recommendations” on page 31 of the final report published at the Office of the National Coordinator for detailed recommendations by stakeholders. Computer-assisted coding software should utilize a combination of statistics-based and rules-based automated coding and a standardized national database (as opposed to a facility-specific database) to train the statistics-based engine. Audit trails are essential in all coding and billing software and EHR application to ensure that codes are based on documentation by clinicians. Machine learning such as ANNs should be available for predictive modeling to reveal trends and scores to detect fraud and abuse before it happens. Users of automated coding should have an appropriate compliance program that includes continuous data analysis to detect potential patterns of abuse prior to claims submission and payment, appropriately trained coding professionals, and use of current coding references and appropriate coding practice standards. Product certification for computer-assisted coding products should be instituted. Certification should be based on criteria assessing the accuracy with which health record documentation is converted to codes based on standard coding principles and guidelines.

Payers and providers must work more closely to prevent fraud. Adherence to standard coding conventions and rules is essential, as is aggregate data analysis and continuous monitoring enabled by computerization. When making any software purchase, providers and payers should evaluate the potential impact on the accuracy of coding, billing, and claims processing so there are no unintended consequences. Coding experts should participate in the selection and implementation processes. Consumer education can help in detecting fraud. Information regarding claims accuracy could be included in a quality measures reporting and consumers might be alerted to potential billing problems. This effort could be assisted by widespread use of patient-friendly billing formats. There should be greater cross-industry collaboration to prevent fraud. This would involve multistakeholder collaboration including payers, billing organizations, and providers with the aim of fewer inaccurate claims, and reduced cost associated with the currently complex and often antagonistic processes. Joint education is needed on methods of prevention of fraud.

## **Discussion**

This research was based on data gathered from selected vendors of automated coding products and a limited number of users. It consisted of Web-based product demonstrations and telephone interviews with vendors, users, and government personnel. Many of the technologies described are newly applied to clinical code assignment, and are not yet in widespread use for this purpose. Generally, more thorough evaluation is needed regarding how these tools perform in a variety of settings with different types of health records. Particular attention should be directed to the coding features of primary care EHRs that prompt for evaluation and management (E & M) code assignment. It is necessary to develop some agreed upon measures so that these technologies can be evaluated over time.

Based on this research we concluded that there are short-term research and action plans. These include the following:

1. Instituting programs to improve national adherence to standard coding guidelines and rules by all stakeholders with education about the consequences of local policy and practice and incentives to drive compliance.
2. Standardizing code to improving data quality. It will also make it less costly to develop automated coding solutions, will permit more reliable trending for fraud detection, and will facilitate adoption of updated code sets.
3. Evaluating the use of computer-assisted coding technologies in production EHR settings, comparing and contrasting the benefits in terms of data integrity, productivity, and compliance monitoring for EHRs that feature structured versus unstructured text and those that are based on a reference terminology.
4. Creating use cases and test databases on which to evaluate the capability of computer software to generate codes according to standard coding guidelines, conventions, and rules. This will permit assessing how best to certify these technologies in the future.
5. Evaluating the potential of using automated code generation and antifraud software in conjunction with the EHR to relieve coding work force shortages.

Jennifer Hornung Garvin is a medical informatics postdoctoral fellow at the Center for Health Equity and Research at the Philadelphia VA Medical Center.

Valerie Watzlaf is an associate professor in Health Information Management at the University of Pittsburgh.

Sohrab Moeini is a graduate student at the University of Pittsburgh.

## Notes

1. National Health Care Anti-Fraud Association. "Health Care Fraud, a Serious and Costly Reality for All Americans." Available at [http://www.nhcaa.org/about\\_health\\_care\\_fraud/](http://www.nhcaa.org/about_health_care_fraud/). Accessed June 14, 2005.
2. Centers for Medicare and Medicaid Services. "HIPAA Part 4: Protecting Medicare from Fraud and Abuse." Available at <http://www.medicare.gov/Publications/Pubs/pdf/10111.pdf>. Accessed June 14, 2005.
3. AHIMA Coding Practice Team. "Internet Resources for Accurate Coding and Reimbursement Practices." (AHIMA Practice Brief) *Journal of the American Health Information Management Association* 75, no. 7 (July/August 2004): 48A–G.
4. AHIMA Position Statement on Consistency of Healthcare Diagnostic and Procedure Coding, 2002. Available at [http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1\\_013732.hcsp?dDocName=bok1\\_013732](http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_013732.hcsp?dDocName=bok1_013732).
5. AHIMA e-HIM Work Group on Computer-Assisted Coding. "Delving into Computer-assisted Coding. Appendix G: Glossary of Terms." *Journal of AHIMA* 75, no. 10 (November-December 2004): web extra.
6. AHIMA e-HIM Work Group on Computer-Assisted Coding. "Delving into Computer-assisted Coding. Appendix C: Advantages and Disadvantages of CAC Technology." *Journal of AHIMA* 75, no. 10 (November-December 2004): web extra.
7. AHIMA e-HIM Work Group on Computer-Assisted Coding. "Delving into Computer-assisted Coding. Appendix E: Summary of Use Cases" *Journal of AHIMA* 75, no.10 (November-December 2004): web extra.

8. Fair Isaac. "Automated Exception Management for Efficient Claims Processing." A Fair Isaac White Paper. January 2003. Available at <http://www.fairisaac.com/NR/rdonlyres/43B4D6E0-6FC8-4C5E-81BA-0279564B5B92/0/AutomatedFraudDetectionExceptionMgmtWP.pdf#search=%22Automated%20Exception%20Management%20for%20Efficient%20Claims%20Processing%22>.
9. Fair Isaac. "Prepayment Fraud and Abuse Detection." A Fair Isaac White Paper. January 2003. Available at <http://www.fairisaac.com/NR/rdonlyres/3186F67E-ACA7-44BB-AAB0-0B51569F8927/0/PrepaymentFraudWP.pdf#search=%22Prepayment%20Fraud%20and%20Abuse%20Detection%22>.
10. Popowich, F. "Using Text Mining and Natural Language Processing for Health Care Claims Processing." *SIGKDD Explorations* 7, no. 1 (June 2005). Available at <http://www.acm.org/sigs/sigkdd/explorations/issues/7-1-2005-06/9-Popowich.pdf>.
11. Popowich, F. "Use of Text Analytics and Taxonomies for Fraud and Abuse Detection in Medical Insurance Claims." Burnaby, BC, Canada: Axonwave Software Inc., 2004.
12. Sordo, Margarita. "Introduction to Neural Networks in Healthcare." October 2002. Available at [www.openclinical.org/docs/int/neuralnetworks011.pdf](http://www.openclinical.org/docs/int/neuralnetworks011.pdf).