Statistical Modeling for Studying the Impact of ICD-10 on Health Fraud Detection

Yu Zhang, Trinity University, San Antonio, TX, USA
Tyler Olson, University of Iowa, Iowa City, IA, USA

ABSTRACT

When an individual is seen or treated by a healthcare professional, a series of alphanumeric codes are used to describe the medical diagnoses and services provided. This designated classification structure, the ninth iteration of ICD (International Classification of Diseases), implements the use of coding for healthcare management, public health and medical informatics, and insurance purposes. ICD-9 has been the coding standard in the healthcare industry for 30 years. On October 1st, 2015, the tenth revision ICD-10 was formally implemented in the United States. This paper explores the validity of predictions from domain professionals regarding fraud detection and the implementation of the ICD-10 code set. The notion that fraud detection systems using supervised learning algorithms will encounter an initial decline in performance due to ICD-10 is fairly unsupported at the moment. The authors claim that the results from their study will provide evidence that will support this notion of a preliminary negative transitional impact.

KEYWORDS
Clinical Classification, Healthcare Fraud Detection, ICD-10, Logistic Regression, Outlier Detection

INTRODUCTION

The implementation of the electronic health record (EHR) has allowed healthcare organizations to collect and externally report/provide a greater amount of data to the public sector. The EHR is a systematic collection of electronic health information about an individual patient or population, and can include a range of data, including demographics, medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics, and billing information. Fraud detection research in the field of healthcare management has primarily concentrated on medical billing data, which is derived from the EHR. After a patient’s medical record is updated by a physician or staff member, diagnosis and procedure codes are assigned by a medical coder. The appropriate medical codes and necessary data from the EHR are incorporated into an ANSI 837 file, which is submitted to the payer directly or via a clearinghouse. An insurance company is usually the recipient of this claim file, so the majority of appropriate fraud detection data comes from health insurance agencies.

Healthcare fraud literature originating from countries outside of the United States has used a variety of sources to acquire medical claims data. The National Health Insurance Administration...
(NHIA) in Taiwan has provided data to multiple research groups (Chan & Lan, 2001; Hwang, Wei & Yang, 2003; Liou, Tang & Chen, 2008; Wei, Hwang & Yang, 2000; Yang & Hwang, 2006), and an analogous NHIA program in South Korea, the National Health Insurance (NHI) system, has also contributed to studies (Shin, Park, Lee & Jhee, 2012). Two major Australian governmental health departments, the Health Insurance Commission (HIC) and Medicare Australia, have been reported as the data sources in numerous pertinent research projects (He, Graco & Yao, 1999; Hubick 1992; Shan, Jeacocke, Murray & Sutinen, 2008; Shan, Murray & Sutinen, 2009; Tang, Mendis, Murray et al., 2011; Williams, 1999). Healthcare claims data from private insurance companies located in Turkey and Chile has also been used by several researchers (Kirlidog & Asuk, 2012; Kumar, Ghani & Mei, 2010; Ortega, Figueroa & Ruz, 2006).

In the United States, various agencies within the Department of Health and Human Services (HHS) have been involved with research exploring the detection and prevention of healthcare fraud. The Centers for Medicare and Medicaid Services (CMS) and its predecessor, the Health Care Financing Administration (HCFA), both supplied members of academia with Medicare and Medicaid data (GAO 1996; Liu & Vasarhelyi, 2013; Musal, 2010; Shapiro, 2002). Researchers in the United States have also worked with private insurance companies and hospitals, using data collected by these organizations (Major & Riedinger, 2002; Ngufor & Wojtusiak, 2013).

Today, when an individual is seen or treated by a healthcare professional, a series of alphanumeric codes are still used to describe the medical diagnoses and services provided. This designated classification structure, the ninth iteration of ICD (International Classification of Diseases), implements the use of coding for healthcare management, public health and medical informatics, and insurance purposes. ICD-9 has been the coding standard in the healthcare industry since October 1st, 1984. The primary purpose of ICD-9 is to translate written information from a patient’s clinical statement regarding diagnoses and inpatient procedures into a series of universally understandable designations.

The ICD-9 code set contains approximately 13,000 distinct codes. Each ICD-9 code consists of a minimum of three digits and a maximum of five digits, with a decimal point after the third digit if more than three digits are used. Figure 1 illustrates the standard format of an ICD-9 code. The first three digits represent a single disease entity, or a group of similar or closely related conditions. The fourth digit identifies a subcategory, providing additional information regarding the etiology, site, or disease manifestation. Lastly, the fifth digit offers sub-classification of the subcategory, describing for example the mode of diagnosis or the anatomical site. ICD-9 is primarily numeric, with the exception of supplementary V-codes and E-codes. V-codes, characterized by a “V” as the first digit, are used when a patient seeks health care for reasons other than illness or injury. E-codes describe external causation of injury, poisoning, and adverse reactions, and where, why, and how an injury occurred. The structure of injuries described by ICD-9 codes are designated by the wound type, and the code omits laterality (left or right). An example of an ICD-9 code is 812.21, which describes a closed

Figure 1. Standard format of the ICD-9 code
Transformation of the Hospital Supply Chain: How to Measure the Maturity of Supplier Relationship Management Systems in Hospitals?
www.igi-global.com/article/transformation-hospital-supply-chain/53476?camid=4v1a

A Risk-Based Classification of Mobile Applications in Healthcare
www.igi-global.com/article/risk-based-classification-mobile-applications/67994?camid=4v1a