Original Research Article

A novel algorithm for identifying risk factors for rare events: Predicting transient ischemic attack in young patients with low-risk atrial fibrillation

Chieh-Yu Liu[a,b,*], Hui-Chun Chen[c]

[a] National Taipei University of Nursing and Health Sciences, College of Health Technology, Department of Speech Language Pathology and Audiology, Biostatistical Consulting Lab,Taipei, Taiwan
[b] National Taipei University of Nursing and Health Sciences, College of Nursing. Department of Midwifery and Women Health Care, Taipei, Taiwan
[c] National Taipei University of Nursing and Health Sciences, College of Nursing, Department of Nursing, Taipei, Taiwan

ARTICLE INFO

ABSTRACT

Identification of risk factors for transient ischemic attack (TIA) is crucial for patients with atrial fibrillation (AF). However, identifying risk factors in young patients with low-risk AF is difficult, because the incidence of TIA in such patients is very low, which would result in traditional multiple logistic regression not being able to successfully identify the risk factors in such patients. Therefore, a novel algorithm for identifying risk factors for TIA is necessary. We thus propose a novel algorithm, which combines multiple correspondence analysis and hierarchical cluster analysis and uses the Taiwan National Health Insurance Research Database, a population-based database, to determine risk factors in these patients. The results of this study can help clinicians or patients with AF in preventing TIA or stroke events as early as possible.

Introduction

The prevalence and incidence of atrial fibrillation (AF) have increased worldwide in recent decades (Kim et al., 2017; Wilke et al., 2013; Yiin et al., 2017). Patients with AF have been reported to have a significantly higher risk of ischemic stroke (IS) (Dulli et al., 2003; Novello et al., 1993; Saxena et al., 2001). Currently, stroke is the second highest cause of mortality worldwide (van Rooy and Pretorius, 2015). Because of the improvement in IS treatment in recent years, the overall survival duration of patients with stroke has been prolonged. However, poststroke rehabilitation is expensive and requires more medical resources. Transient ischemic attack (TIA) is defined as a period of focal ischemia, the symptoms of which resemble a thromboembolic stroke, and TIA has been demonstrated as a precursor of IS in patients with AF (Appelros et al., 2017; Cruz-Flores, 2017). Compared with stroke, TIA symptoms typically occur for less than several hours with absence of necrosis, and stroke is often preceded by TIA. Therefore, TIA is a crucial predictor of stroke events. However, the incidence of TIA in young patients with low-risk AF is very low. Thus, identification of risk factors for TIA in young patients with low-risk AF is necessary to more effectively prevent stroke events, because they may have longer survival and require high costs and more medical resources (Abdelhafiz and Wheeldon, 2003).

In the current clinical practice, the risk of IS onset is mainly measured using two risk scoring systems: the congestive heart failure (HF), hypertension (HTN), age ($\geq$75 years), diabetes mellitus (DM), and prior stroke, TIA, or thromboembolism history (CHADS2) score (Rietbrock et al., 2008) and the CHA2DS2-VASc score (Pieri et al., 2011), which is a modification of the CHADS2 score aimed at improving stroke risk prediction in patients with AF by adding three risk factors: age (65–74 years), female sex, and history of vascular disease. Both scoring systems have been useful in predicting stroke events in patients with AF, particularly in older patients (aged 65–74 or $\geq$75 years). However, a risk scoring system for predicting TIA events in patients with AF is not available, probably because the incidence of TIA in young patients with low-risk AF is very low. These two scoring methods are probably not feasible for predicting the onset of TIA; therefore, a novel algorithm for predicting TIA events in young patients with low-risk AF is

* Author for correspondence: National Taipei University of Nursing and Health Sciences, College of Health Technology, Department of Speech Language Pathology and Audiology, Biostatistical Consulting Lab, 365 Min-der Rd., Beitou district, Taipei City, Taiwan.
E-mail addresses: chiehyu@ntunhs.edu.tw, chiehyuliu@gmail.com (C.-Y. Liu).

G Model
JAB 150 No. of Pages 6

ARTICLE IN PRESS

2   *C.-Y. Liu, H.-C. Chen / J. Appl. Biomed. xxx (2017) xxx–xxx*

required. Such an algorithm can be developed using a large-scale population-based database, namely the Taiwan National Health Insurance Research Database (NHIRD). In this study, a novel algorithm was developed for identifying risk factors for TIA events in patients with low-risk AF who are younger than 60 years.

Materials and methods

Study database

The claims database of the National Health Insurance (NHI) program in Taiwan, which was launched on March 1, 1995, was used in this study. The NHI program provided comprehensive healthcare services for approximately 99.5% of Taiwanese residents in 2010 (Yeh and Chang, 2015). The NHIRD contains nationwide population-based information including outpatient and inpatient clinical visits, dentistry services, prescription drugs, and traditional Chinese medicine services. The diagnostic and procedural codes of diseases are based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and Procedure Coding System (ICD-9-PCS).

Ethics statement

This study was approved by the Institutional Review Board of School of Nursing, National Taipei University of Nursing and Health Sciences (CN-IRB-2011-064). Because personal information that may potentially identify any individual patient was fully encrypted by the National Health Insurance Administration (NHIA), written consent from study patients was not obtained. The NHIA guarantees the confidentiality of the personal and health information of enrolled patients.

Study population

An incidence-based cohort of patients who had received a new diagnosis of AF (ICD-9-CM code 427.31), had at least two outpatient visits with AF being the primary disease, were aged between 20 and 59 years. The exclusion criteria were: (1) had any TIA or stroke event history (including hemorrhagic stroke [HS] and IS) in 2005; (2) AF patients who were aged <20 or >60 years old; (3) AF patients with severe baseline diseases: including cancers, coronary artery disease (CAD, including congenital heart defect (CHD), myocardial infarction, and heart failure), kidney failure (including chronic kidney failure (CKD)) and peripheral artery disease (PAD); (4) AF patients' CHA2DS2-VASc score was ≥ 2. AF patients who met the above criteria were identified and retrieved
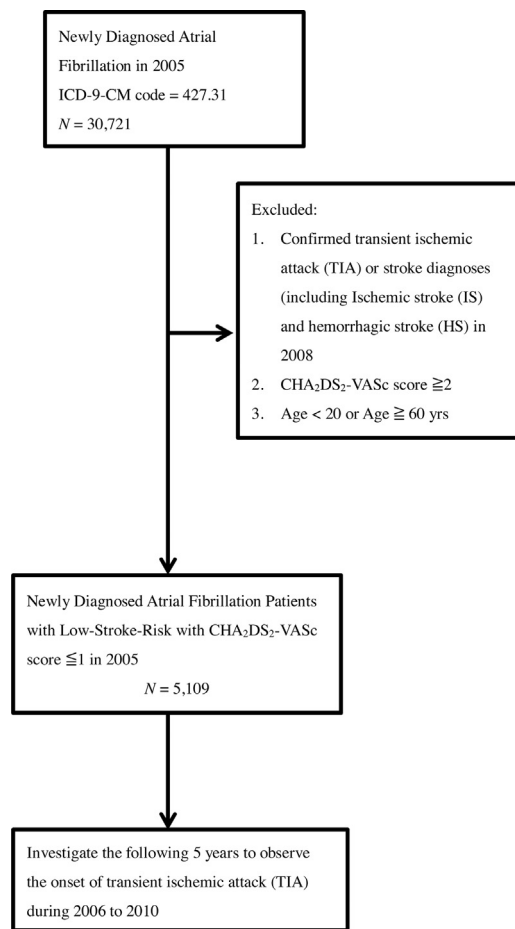

Fig. 1. Enrollment scheme of this study.

from the NHIRD. This present study retrieved only outpatient claims data from the NHIRD. We linked the data of study patients to their 2006–2010 medical claims database to observe if they had onset of TIA. The definition of low-risk AF patients was using AF patients' CHA2DS2-VASc score was <2 ($= 0$ or 1), which is a widely used definition worldwide in many published studies (Chao et al., 2014; Crivera et al., 2016; Manolis et al., 2016; Piyaskulkaew et al., 2014). The enrollment scheme of study patients is shown in Fig. 1.

Algorithm. (Havranek et al., 2016; Weller and Romney, 1990; SPAF III Writing Committee, 1998).