MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Generating a Comprehensive Disease Pattern of Obstructive Sleep Apnea: A Data Mining Approach

Qi Rong Huang[a], Zhenxing Qin[b], Shichao Zhang[b] ,Chin Moi Chow[c]

Discipline of Health Information Management, Faculty of Health Sciences, The University of Sydney
Faculty of Information Technology, The University of Technology Sydney
Delta Sleep Research Unit, Discipline of Exercise and Sport Science, The University of Sydney

Abstract

Obstructive sleep apnea (OSA) is a sleep disorder with a wide range of comorbid conditions. Despite its known aetiology and manifestations, the strength of its association with comorbid conditions has not been fully explored. This research explored a large dataset (1999-2004) derived from the New South Wales Inpatient Data Collection, using a data mining approach, to provide a comprehensive disease pattern of OSA. All diagnoses were coded with the International Classification of Diseases (ICD)-Australian Modification codes. A total of 60,197 OSA cases (4% of total records) were identified. Their onset of occurrence was age-related with a peak occurrence at age 50-59y for both genders with a higher incidence in males. Comorbid conditions, in order of strength of association with OSA, were essential hypertension, obesity, hyper cholesterolemia, type 2 diabetes, and ischemic heart disease. When categorised according to disease systems, all cormorbid conditions clustered mainly into endocrine/ metabolic, cardiovascular and respiratory systems. The data mining approach unveiled the full OSA disease pattern and illustrated the progression of comorbid conditions.

Keywords:

obstructive sleep apnea, data repository, data mining, ICD-10I

Introduction

Obstructive sleep apnea (OSA) is characterised by repetitive, complete or partial occlusion of the posterior pharyngeal airways during sleep [1]. It affects approximately 2–15% of the general population [2]. The symptoms of OSA are snoring, transient cessation of breathing, and excessive daytime sleepiness. OSA patients suffer a poor quality of life due to fragmented sleep and a wide range of comorbid conditions of obesity, diabetes and cardiovascular diseases (CVD) [3]. Although much is now known about its prevalence, demographic characteristics, risk factors and comorbid conditions, a comprehensive disease pattern of OSA remains to be explored.

Data mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [4]. It assumes no prior knowledge but has the potential to uncover previously unknown relationships. The study aimed to apply a data mining technique to a clinical data repository to identify any unknown disease patterns and relationships with OSA.

Materials and methods

Data sources and definition of OSA

The data repository for this study is the New South Wales (NSW) Inpatient Data Collection. All public and some private hospitals report monthly all inpatient data to the Department of Health. A random dataset (1999-2004) was extracted. All diagnoses and procedures were coded with the International Classification of Diseases (ICD-10) Australian Modification (AM) codes [5]. G47.32 is the code for OSA. When comorbity conditions were subsequently explored, 4th and 5th digits for sub-classification were ignored. The number of codes generated for a patient can range from 5-20.

Data mining technique

The technique used was the association rule analysis that aims to discover the conjunctions of attribute-value pairs that are statistically related in huge data repository [6]. Association rule is an implication of the form X → Y, where X and Y are itemsets and X Y= Φ to predict that 'if X occurs in a transaction, then Y will likely also occur in the same transaction'. Each rule has two measures of its strength called support (the frequency of causality) and confidence (the strength of the causality):

$$\text{Support} \quad (X \rightarrow Y) = \frac{|\{t \in T \mid (X \cup Y) \subseteq t\}|}{|T|}$$

$$\text{Confidence} \quad (X \rightarrow Y) = \frac{|\{t \in T \mid (X \cup Y) \subseteq t\}|}{|\{t \in T \mid X \subseteq t\}|}$$

Identified dataset containing a diagnosis of OSA was a *focused dataset*. Frequency patterns discovered by association mining were noted as *focused patterns.*

Results

The dataset contained a total of 1,511,078 patient records (47.5% males and 52.5% females) from 1999-2004. A total of 60,197 (4%) had a diagnosis of OSA, listed as either principal or secondary diagnosis. Of these OSA cases, 72.2% were male and 27.8% were female (Figure 1). Comorbid conditions in order of the strength of association with OSA were essential hypertension, obesity, and type 2 diabetes mellitus (IDDM2), ischaemic heart diseases (IHD), hypercholesterolemia, congestive heart failure (CHF), cardiac arrhythmias, chronic obstructive lung disease, hypertrophy of tonsils/adenoids and disorders of nose and nasal sinuses, and joints problems such as gout, arthrosis, gonarthrosis and osteoporosis. Figure 2a,b show the age distribution for obesity, IDDM2, hypertension and IHD.

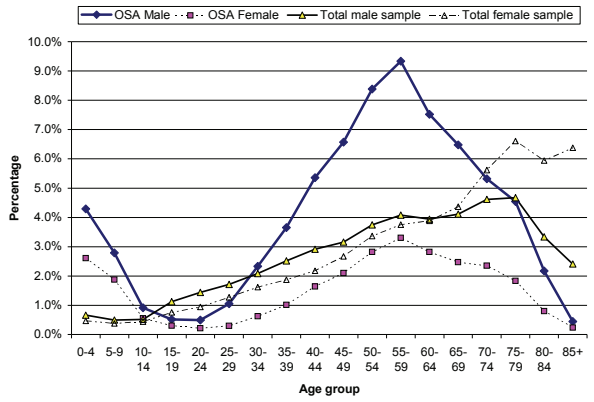
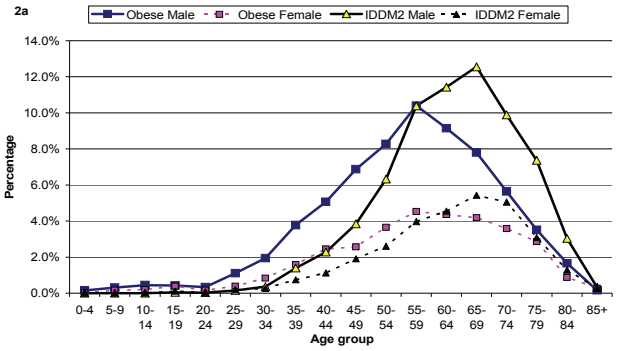


Figure 1 - Age and gender distributions of the sample and OSA cases



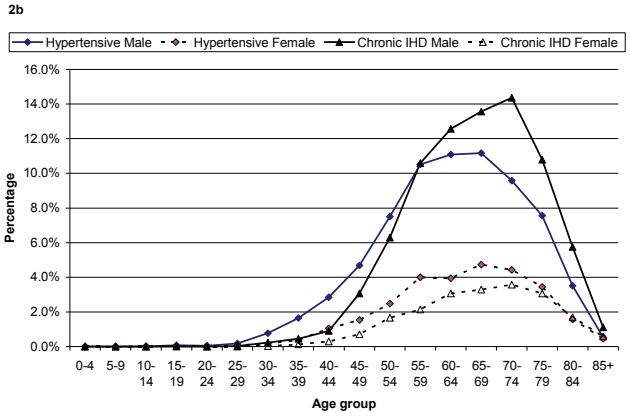


Figure 2 - Age and gender distributions of obestity and IDDM2 (2a), and hypertension and chronic IHD (2b)

Past or current use of tobacco was frequently listed in OSA patients. The comorbid conditions were clustered into the endocrine/ metabolic, cardiovascular and respiratory systems. There was high usage of health services amongst these patients (data not shown).

Discussion

This study has for the first time generated a comprehensive disease pattern of OSA. It confirms previous findings of a higher male prevalence, obesity as a risk factor, and associated comorbid conditions, namely IDDM2, and CVD (hypertension, IHD, CHF) [3].

The strong association of obesity, CVD, hypercholesterolemia and the metabolic syndrome with OSA suggests that a complex relationship of cause and effect exists between these conditions. Increase in obesity cases preceded that of hypertension, diabetes, and IHD. Such observation is in keeping with the knowledge of development and progression of disease patterns. Smoking may be a contributory factor to IHD and pulmonary diseases [7]. Joint diseases (arthrosis and gonarthrosis) are comorbid with OSA, although they are much less common compared to CVD. This observation has not been previously reported. Heavy use of health services related to treatments of OSA and comorbid conditions directly bear on the cost of health services and health burden on society.

In conclusion, data mining uncovers complex relationships between OSA and its associated comorbid conditions, and age and gender distribution of OSA. It offers an insight into the health cost of OSA and management of the disease.

Acknowledgements

We would like to thank the Centre for Epidemiology and Research, NSW Department of Health, for providing the dataset.

References

[1] Krieger J, McNicholas WT, Levy P, De Backer W, Douglas N, Marrone O, Montserrat J, Peter JH, Rodenstein D. Public health and medicolegal implications of sleep apnoea. Eur Respir J 2002; 20:1594-609.

[2] Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea. Am J Respir Crit Care Med. 2002;165:1217-1239.

[3] Young T, Skatrud J, Peppard PE. Risk factors for obstructive sleep apnea in adults. JAMA 2004;291: 2013–2016.

[4] Brossette SE, Sprague AP, Hardin JM, Waites KB, Jones WT, Moser SA. Association rules and data mining in hospital infection control and public health surveillance, J Am Med Inform Assoc. (1998);5:373–381.

[5] National Centre for Classification in Health. ICD-10-AM. National Centre for Classification in Health. Sydney 2000.

[6] Klemettinen M, Manilla H, Ronkainen P, Toivonen H, Verkamo A. Finding interesting rules from large sets of association rules. In: Proceedings of the 3rd International Conference of Information and Knowledge Management. New York: ACM Press, 1994:401–7.

[7] Shepard JW Jr. Cardiopulmonary consequences of obstructive sleep apnea. Mayo Clin Proc 1999;65: 1250–1259.

Address for correspondence

Qirong Huang, MD, PhD, Lecturer, Discipline of Health Information Management, Faculty of Health Sciences, The University of Sydney, Po Box 170, Lidcombe NSW 1825, Australia; Email: joe.huang@usyd.edu.au