

# A data mining system for chronic diseases surveillance for primary care unit in Taiwan

HM Hua<sup>1</sup>, T. Lee<sup>2</sup>.

Graduate school of Gerontic Technology and Service Management, Nan Kai Institute of Technology, Caotun, Nantou, Taiwan. <sup>1</sup>

Doctor Lee Family Medicine Clinic, Tamshui, Taipei Hsien, Taiwan. <sup>2</sup>

**Abstract**—Chronic diseases affect a large proportion among elderly and cause a major public health burden in Taiwan. Complications of hyperlipidemia, hypertension, triglyceride and diabetes highly increase the risks of medical treatment and long-term care for the elderly. In this study, we provide a data mining system for chronic diseases surveillance to assist clinics doctors in medical treatment concerns for their patients. It is not only able to reduce the cost of national health insurance but also to afford the better quality of life for elderly people.

## I. INTRODUCTION

CHRONIC diseases affect a large proportion among elderly and cause a major public health burden in Taiwan [1] (Figure 1). Elderly with chronic disease is not only to increase death but also to raise health insurance cost. Complications of hyperlipidemia, hypertension, triglyceride and diabetes highly increase the risks of medical treatment and long-term care for the elderly. In this study, we provide a data mining system for chronic diseases surveillance to assist clinics doctors in medical treatment concerns for their patients. It is not only able to reduce the cost of national health insurance but also to afford the better quality of life for elderly people.

uses database and statistical methods including structured patient information with diagnosis, drugs and biochemical tests or physical examinations.

We infer the values of several medical outcomes, described by a set of variables of interest. Data collection of such variables include whether a patient has a particular disease, whether a patient has received a certain type of medication, lab recordings for blood glucose, whether a patient has specific contraindications for a class of medication (Figure 2).

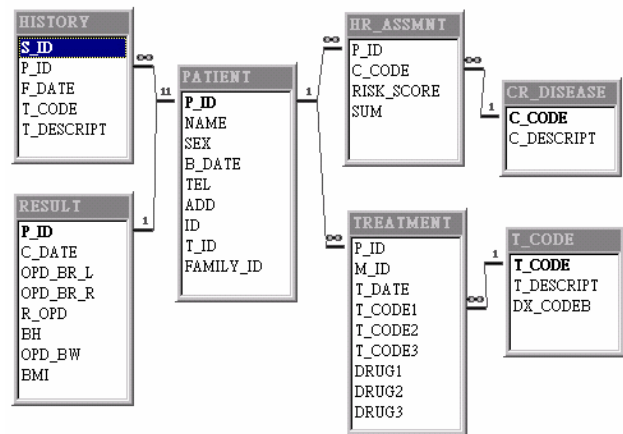


Figure 2 Electronic Structured Patient Records with Clinical Data

## Classification of Diseases

To structured information about patient demographics, the domain knowledge of chronic diseases is classified according to the internationally accepted standards, ICD-9 (International Classification of Diseases, 9th Revision [2]). Many of the criteria used to determine if a patient is eligible for or treated according to a particular guideline, are based upon diagnostic information.

## Machine Learning Approach

Many studies have shown that the clinical accuracy of ICD codes is only 60%-80% [3]. Thus, we we provide a machine learning approach to data mining and knowledge discovery based on the induction of clinical classification, association and sequence rules [4].

Chronic diseases require multiple decisions concerning medications. A decision tree is regarded as a hierarchically organized set of clinical rules. Association rule is mainly for discovering chronic diseases possible synchronize. Both genetic algorithm and case base reasoning in hybrid data mining technology are used to acquire of knowledge from database [8]-[12]. These techniques also map personal, familial, and professional values to meet clinical diagnosis

表1 1999至2007年全民健康保險之門診門診醫療服務統計數據

疾病類別	CCS序號	CCS中文名稱	99年				04年					
			門診人數	門診次數	門診費用	門診藥品	門診人數	門診次數	門診費用	門診藥品		
1	120	高血壓性心臟病	1	20,520	20,299	121	20,020	12,171	20,694	10,519	122	12,567
2	150	慢性腎臟病	3	23,055	23,164	928	20,020	18,721	20,629	9,398	539	20,128
3	125	急性上呼吸系統感染	2	25,162	24,608	554	24,066	4,335	23,636	3,356	539	2,688
4	131	成人呼吸器病	4	12,618	42	12,596	12,039	2,541	97	55,776	12,843	2,281
5	122	肺炎	8	8,786	523	8,722	10,966	14,448	720	37,966	9,352	13,058
6	98	高脂血症(非家族性)	5	9,016	9,855	61	9,612	3,078	9,371	0,866	51	16,726
7	205	糖尿病(包括併發症)	6	9,368	7,235	1,842	9,009	3,833	7,214	4,176	1,880	9,076
8	70	糖尿病(無併發症)	9	8,448	2,893	2,956	8,862	4,881	2,642	6,106	6,339	6,399
9	101	糖尿病(家族性)	7	9,216	3,896	5,320	8,810	4,411	3,827	1,799	5,098	4,181
10	49	糖尿病(非併發症)	11	6,309	6,302	7	7,344	16,408	7,338	16,476	6	11,099
11	50	糖尿病(併發症)	10	7,055	4,058	2,897	5,528	17,424	4,263	12,298	1,091	5,528
12	203	尿石症(腎臟)	13	5,420	3,303	2,118	5,377	6,858	3,235	2,666	2,489	3,438
13	99	高脂血症(家族性)	12	5,036	4,988	668	5,332	5,408	4,898	1,436	742	11,114
14	110	關節炎	14	5,242	2,505	2,273	5,196	6,088	2,592	3,898	2,738	4,076
15	211	肌力衰弱(肌肉、神經系統)	15	5,201	4,624	577	4,925	5,313	4,404	4,746	656	13,688
16	69	癱瘓性神經病	17	4,445	3,000	1,844	4,698	5,709	3,308	10,276	1,890	3,788
17	125	急性心肌梗塞	18	4,358	3,723	405	4,603	5,778	3,990	7,188	681	2,906
18	127	慢性阻塞性肺病	16	4,511	2,091	2,559	4,509	1,113	3,025	1,395	2,568	2,329
19	45	肺炎(非家族性)	20	3,376	17	3,359	4,170	23,511	49	14,626	4,151	23,555
20	134	上呼吸系統感染(非家族性)	20	3,919	3,426	513	4,087	1,741	3,598	5,029	504	1,999
合計			194,222	105,727	105,498	146,706	179,719	113,133	268,111	368,111	48,613	54,474
前十大疾病佔比			48,613	48,613	48,613	48,613	48,613	48,613	48,613	48,613	48,613	48,613
前二十大疾病佔比			105,508	141,302	92,206	195,455	1,016	143,526	54,474			
前30大疾病佔比			48,613	48,613	48,613	48,613	48,613	48,613				

Figure 1 The National Health Insurance Statistics--by diseases, 2004 data from Bureau of National Health Insurance, Taiwan

## II. METHODS

### Samples and Data Collection

Medical patient data in electronic form in database in this study has been obtained from a local clinic in Taipei, Taiwan from 1999 to 2007. Data also consists of all the information required to document the physician's diagnoses and the procedures performed. Family tree and history also correspond with patients. The central role of data mining

requirement and long-term care decisions (Figure 3).

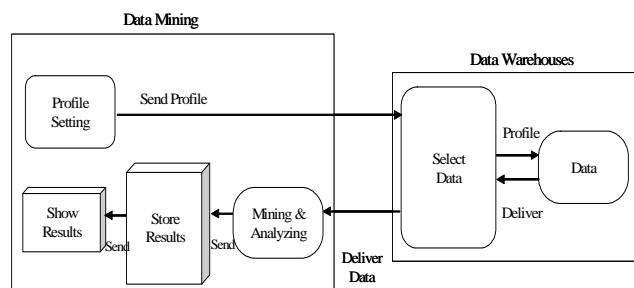


Figure 3 The system structure of chronic diseases surveillance

### Auto-remind System

Traditional medical treatment on patients relies upon manual conversion of data by medical experts. However, it is expensive, time consuming, and is only possible for a small subset of patients or at institutions with a strong research focus. Based on the data statistic and analysis, we count risk factors related with chronic diseases and patients. An auto-remind mechanism on multi-source data based on prior medical knowledge is processed with observations to assist clinics doctors in medical treatment concerns for their patients.

### III. RESULTS AND DISCUSSION

Chronic diseases treatment requires long term caring. Chronic medical data is also highly complex and difficult to analyze. Lack of cost and staff, the primary care currently focuses on non-serious diseases but requiring urgent treatment. However, our chronic diseases surveillance is built on structured information in the clinical repository, typically as unstructured free text in patient history and physicals, discharge summaries, progress notes, radiology reports, etc. To treat and monitor these patients with chronic diseases complications, such as diabetic patients with hyperlipidemia or hyperlipidemia patients with triglyceride, highly improve the quality of clinical care.

Chronic disease caring relies on patients' taking the initiative for return treatment. Our system is designed to provide the patients list according to their return treatment dates. Medical staffs take the initiative to remind patients for their return treatment. It highly improves chronic diseases treatment.

Clinical records and medical treatment analysis give a practical controlled picture to both doctors and chronic diseases patients. These analyses give great assistance to the medical staff in monitoring the patients' symptoms.

Hypertension is an important factor among many other cardiovascular diseases. However, other risk factors such as diabetes, high cholesterol, smoking, sex, and family history etc. should also be assessed and analyzed in order to monitor the level of danger of other heart and blood related disorders. In principle, the high risk patients (i.e. the high risk groups) should have priority to be the monitored subjects.

Our system has also built up family medical history records. It provides interdisciplinary medical and health service in prevention, emergency and chronic disease treatment. Through medical treatment analysis, clinical doctors at primary care unit are able to build up a good long term relationship with chronic diseases patients. Chronic diseases patients obtain better quality of clinical care. Clinics, doctors and patients all benefit from our chronic diseases surveillance system.

In the more distant future, drugs analysis will be added in to measure the clinical effects associated with the provision of medication therapy management services to cut down national health insurance cost.

### REFERENCES

- [1] The National Health Insurance Statistics--by diseases, 2004 data from Bureau of National Health Insurance, Taiwan. Available: [http://www.nhi.gov.tw/webdata/webdata.asp?menu=1&menu\\_id=4&webdata\\_id=805&WD\\_ID=19](http://www.nhi.gov.tw/webdata/webdata.asp?menu=1&menu_id=4&webdata_id=805&WD_ID=19).
- [2] World Health Organization. Manual of the international statistical classification of diseases, injuries, and causes of death. World Health Organization, Geneva, 1977.
- [3] C Benesch, DM Witter Jr, AL Wilder, PW Duncan, GP Samsa, DB Matchar. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology*, Vol. 49, pp. 660-664, 1997.
- [4] T Imamura, S Matsumoto, Y Kanagawa, B Tajima, S Matsuya, M Furue, H Oyama. A technique for identifying three diagnostic findings using association analysis. *Medical and Biological Engineering and Computation*. 45, pp 51-59, 2007.
- [5] DH Arnold, T Gebretsadik, PA Minton, S Higgins, TV Hartert. Assessment of severity measures for acute asthma outcomes: a first step in developing an asthma clinical prediction rule. *Am J Emerg Med*. 26(4):473-9, May, 2008.
- [6] SS Young, N Ge. Recursive partitioning analysis of complex disease pharmacogenetic studies I. Motivation and overview. *Pharmacogenetics*. vol. 6, pp. 65 - 75, 2005.
- [7] YC Wang, MH Chung, KL Lai, CC Chou, S Kao. Preferences of the elderly and their primary family caregivers in the arrangement of long-term care. *Journal of Formosa Medicine Association*, 103(7), 533-539, 2004.
- [8] GO Barnett, JJ Cimino, AJ Hupp, EP Hoffer. DXplain: an Evolving Diagnostic Decision-Support System. *JAMA*. Vol. 258(1), pp. 67-74, 1987.
- [9] B Kappen, W Wiegerinck, E Akay, J Neijt, A Van Beek. Promedas: A Clinical Diagnostic Decision Support System in Bayesian Modeling Applications. Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence (BNAIC'), Nijmegen, The Netherlands, October 23-24, 2003. pp. 455-456.
- [10] RB Rao, S Krishnan, RS Niculescu. *SIGKDD Explorations*. 8, pp 3-10, 2006.
- [11] E Armengol, A Palaudaries, E Plaza. Individual Prognosis of Diabetes Long-term Risks: A CBR Approach. *Method Inform Med*. Vol.40, pp.46-51, 2001.
- [12] C Artur, S Zdzisława. Diabetes Mellitus and Cancer. *European Journal of Internal Medicine*, Vol.11, pp.245-252, 2000.