Original Article   Open Access

A New Tool Preliminary Assessment on Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER)

Tim Chen[*]
School of Industrial Engineering, West Lafayette, Purdue University, USA
School of Computer Science and Informatics, Faculty of Technology, De Montfort University, LE1 9BH Leicester, UK

Hendri Daleanu
Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Institute of Space Sciences, Magurele¡VBucharest, Romania

J.C.-Y. Chen
NAAM Research Group, King Abdulaziz University, Jeddah 21589, Saudi Arabia
St Petersburg Univ, Dept Math & Mech, Univ Skii 28, St Petersburg 198504, Russia

Abstract

Patients' comorbidities, operations and complications can be associated with reduced long-term survival probability and increased healthcare utilisation. The aim of this research was to produce an adjusted case-mix model of comorbidity risk and develop a user-friendly toolkit to encourage public adaptation and incremental development. It has been shown in healthcare research that demographics, temporal dimensions, length-of-stay and time between admissions, can noticeably improve the statistical measures related to comorbidities. The proposed model incorporates temporal aspects, medical procedures, demographics, and admission details, as well as diagnoses. The research resulted in the development of Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) model using routinely collected hospital data.

Keywords: Multiple criteria decision making; Interval-valued belief distribution; Evolutionary algorithm; Accuracy; Efficiency.

1. Introduction

There is increasing evidence that the quantification of high-risk diagnoses, operations and procedures, and monitoring changes over time, can greatly improve the quality of readmission models with adequate adjustment. There have been two streams of work on risk scoring comorbidities to estimate future resource utilisation, emergency admission and mortality.

Firstly, one stream of research looks at the odds ratio of major diagnoses groups and therefore is highly reliant on the whole population statistics. These models stem from crudely summing up the derived weights for comorbidities, which are based on the most recent admission of patients with disregard of temporal patterns. A popular example is the Charlson Comorbidity Index (CCI) [1], which relies on twenty-two comorbidity groups. One of the recent translation of the CCI is the National Health Service (NHS) England version of the CCI (NHS-CCI), that is continuously being updated [2-6].

The second stream of models uses a diagnosis classification approach based on simi larities, type, likelihood or duration of care. However, they are usually very complex and specialised to highly particular settings and populations. Also, these models use a period of care records in past, but temporal patterns are greatly ignored. One prominent method is the Elixhauser Comorbidity Index (ECI) [7, 8], which relies on thirty comorbidity groups and 1-year lookback period. Unlike the CCI, the ECI is using Diagnosis-Related Groups (DRG), which was first developed by Fetter, et al. [9], Mistichelli [10]and is based on ICD (International Statistical Classification of Diseases) diagnoses, procedures, age, sex, discharge status, complications and comorbidities. A recent adaptation of the ECI is the AHRQ-ECI, which is actively being maintained by the US Public Health Service [7]. Another well-established method is the John Hopkin's [11], Adjusted Clinical Groups (ACGs), which is a commercial tool. The model uses a minimum of 6-month and maximum of 1-year prior care records, and it encapsulates 32 diagnoses groups, known as Aggregated Diagnosis Groups (ADGs), and their aggregations called Expanded Diagnosis Clusters (EDCs). Moreover, these indices are initially developed to adjust for particular risks, like mortality risk and care utilisation, but they are commonly used in a variety of risk adjustment problems in critical care health services research.

In the machine learning pipeline that was developed in the prior stage of our research [12], comorbidity index was an extremely significant factor and has a high potential for further improvement. Presently, comorbidity risk indices have four major weakness areas: robustness, temporal adjustment, population stratification, and the inclusion of associated factors to comorbidities and complications.

In this research, we are going to improve on these four major areas. Firstly, to make the risk score relevant to different environments, an approach must be used to model complex correlation between variables and states.

*Corresponding Author: chwong@sinica.edu.tw

Secondly, to better distinguish the short- and long-term conditions (i.e. prior admission, length-of-stay, and delta-time between admissions), the temporal dimension may be included in form of life-table or a polynomial weight function. Thirdly, population stratification is a major factor in the prevalence of medical conditions, and therefore must be adjusted. Fourthly, major correlated factors to diagnoses may be included directly or indirectly (latent) to improve the risk estimates, including secondary diagnoses, operations, procedures and complications.

2. Methods
2.1. Data

In this study, a bespoke extract of the HES inpatient data was used, which contains records from April 1995 to April 2010. Two main samples were randomly selected from this database, which includes 20% of total unique patients from 1999-2004 and 2004-2009 periods. Then, each main sample was divided into two equal half, to be used for training and testing. The specification of the data and selected samples are presented in Table 1.

Table 1
Selected samples from the HES inpatient.

Samples	Timeframe	Population size		Sample size		Selected Population		
		Episodes	Patients	Episodes	Patients	Patients	No prior spell	No post spell
Sample-1	1999/04 - 2004/03	18,885,777	7,206,133	6,347,067	1,441,227	1,157,873	492,458	148,950
Sample-2	2004/04 - 2009/03	31,731,488	8,104,748	11,394,152	1,615,347	1,410,923	395,522	110,961

Each time-frame was divided into one year of trigger-event, one year of prediction period, and three years of prior-history. The population includes all alive patients greater than one-year-old that have an admission within the trigger year. The prediction target variables are 30- and 365-day hospital emergency admission to the inpatient.

2.2. Features

After the data extraction step, several stages of data pre-processing and feature selection were carried out using the framework introduced by Mesgarpour, *et al.* [13]. Firstly, a set of data pre-processing steps are applied, then the feature selection steps are carried out. Also, before carrying out the feature selection steps, features are aggregated and split into temporal events, to capture the events through time.

2.2.1. Pre-Processing

The pre-processing stage implements data selection, removals of invalids and imputations of observations. Also, the feature re-categorisation was applied in this stage, to reduce sparsity and to better capture non-linear relationships.

Table 2
Definition of features.

Main Feature	Definition
gender	Gender of patient (Female, Male, Others)
ethnos	Ethnicity of patient (Bangladeshi, Black African, Black Caribbean, Black Other, Chinese, Indian, Pakistani, White, Others).
imd04rk	The Index of Multiple Deprivation (IMD) overall ranking of income (22.5%), employment (22.5%), health deprivation and disability (13.5%), education & skills (13.5%), barriers to housing & services (9.3%), crime (9.3%), & living environment (9.3%).
ageTrigger	Age of patient at the trigger event. Categorisation bins: {10-, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90+}.
gapDays (temporal)	Delta-times from discharge to the trigger admission.
epidur (temporal)	Spell durations.
preopdur (temporal)	Pre-operative durations.
posopdur (temporal)	Post-operative durations.
operOPCSL1 (temporal)	The level-1 categories (25 groups) of operating procedure codes (OPCS), the national standard [24] version 4.0 (¿4,000 codes).
diagCCS (temporal)	The level-1 categories (302 groups) of Clinical Classifications Software (CCS) for ICD-10 [8] diagnoses (69,800 codes).
admimeth (temporal)	The level-1 categories (3 groups) of admission method (20 codes): {Elective, Emergency, Others}.
mainspef (temporal)	The level-1 categories (33 groups) of the main specialities of the consultants (86 codes), based on the exploratory analysis.

In re-categorisation step, a clinical grouper, known as the Clinical Classifications Software (CCS), was used to categorise the diagnoses, to better capture comorbidities' patterns and cross-correlations. The CCS categorises the ICD-10 (10th revision of the ICD) diagnoses and operations into a number of categories that are clinically meaningful [14, 15]. Furthermore, operations and procedures were categorised using the major categories of the OPCS-4, but alternative coding categorisation may be used, like ICD-10-PCS. The OPCS-4 is an alphanumeric nomenclature (similar to ICD-10-PCS), and is used by the NHS England and has an implicit categorisation for operations based on clinical categories rather than cost or risks.

2.2.2. Life-Table and Aggregation

Healthcare administrative data are severely unbalanced regarding the amount of longitudinal (panel) data per patient and their distributions over the years. Statistical methods are not equipped to handle these type of unbalances directly. Therefore, the survival analysis's life-table approach was used to keep track of temporal events [16].

Based on previous studies and the initial statistical analyses, four levels of temporal features were generated: 0-30, 30-90, 90-365 and 365-730 days. These four levels capture part of the temporal aspect of comorbidities, in addition to the delta-time between admissions (gapDays) and the length-of-stay (epidur) features that include temporal metadata. Furthermore, in the modelling stage, we applied several techniques to capture the complex temporal patterns of patients' Comorbidities. The temporal features were summarised in each temporal level based on several aggregation functions, including prevalence, count and average. This stage increased the number of features by more than fifty folds.

2.2.3. Feature Selection

After feature generation, a feature pool was produced based on the developed features. Thereafter, the feature selection step has been carried out. Firstly, the features were filtered out based on their linear cross-correlation, as well as frequency and sparseness (percentage of distinct, and the ratio of the most common value to the second most common).

Thereafter, the continuous features have been transformed using two feature transformations methods: scale-to-mean and Yeo-Johnson [17]. Both methods can be used to transform the data, to improve normality. Although, feature transformations would not guarantee better convergence or very stable variance for any dataset, they have been applied to avoid inputting skewed features into models. Moreover, a downside of transformations is that they make model interpretation harder, and can negatively impact the relationship between correlated features in the model. Therefore, the highly correlated features were removed after transformations.

3. Discussion

We compared the performance of the T-CARER against commonly used comorbidity index models using different samples and population cohorts across a ten year period. Our analyses of the T-CARER and the NHS-CCI for different diagnoses categories demonstrated that our model performed best in the majority of comorbidity groups, and in overall T-CARER models show better results against previous surveys of CCIs and ECIs.

4. Conclusion Remarks

An advanced research will be sought to identify an approach to score commodities by the inclusion of diverse categories of diagnoses, operations and complexities. The T-CARER performs consistently across tests and validations, and it outperformed against Charl-son and Elixhauser indices which are widely used for prediction of comorbidity risks.

Funding Acknowledgement

Conflict of Interest Statement

The authors declare no conflict of interest in preparing this article.

References

[1] Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R., 1987. "A new method of classifying prognostic comorbidity in longitudinal studies, Development and val idation." *Journal of Chronic Diseases,* vol. 40, pp. 373–383.

[2] Aylin, P., Bottle, A., Jen, M. H., Middleton, S., and Intelligence, F., 2010. "Hsmr mortality indicators." Available: https://www1.imperial.ac.uk

[3] Bottle, A., Jarman, B., and Aylin, P., 2011. "Strengths and weaknesses of hospital standard ised mortality ratios." *B. M. J,* vol. 342, p. c7116. Available: https://doi.org/10.1136/bmj.c7116

[4] HSCIC, 2014. "National clinical coding standards OPCS-4 - accurate data for quality information." Available: http://systems.hscic.gov.uk

[5] HSCIC, 2015. "Summary hospital-level mortality indicator (shmi) - methodology de velopment log." Available: http://www.hscic.gov.uk

[6] HSCIC, 2016. "Summary hospital-level mortality indicator." Available: http://www.hscic.gov.uk

[7] AHRQ, 2016. "Ahrq - hcup - elixhauser comorbidity software for icd-10-cm." Available: https://www.hcup-us.ahrq.gov

[8] Elixhauser, A., Steiner, C., Harris, D. R., and Coff ey, R. M., 1998. "Comorbidity measures for use with administrative data." *Medical Care,* vol. 36, pp. 8–27.

[9] Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., and Thompson, J. D., 1980. "Case mix definition by diagnosis-related groups." *Medical Care,* vol. 18, pp. 1– 53.

[10]   Mistichelli, J. "Diagnosis related groups (drgs) and the prospective payment sys tem, Forecasting social implications." *Bioethics Research Library,* Available: https://www.researchgate.net/publication/268184574_Diagnosis_Related_Groups_DRGs_and_the_Prospective_Payment_System_Forecasting_Social_Implications

[11]   Weiner, J. and Abrams, C., 2011. "The johns hopkins ACG system - technical reference guide - version 10." Available: http://acg.jhsph.org

[12]   Mesgarpour, M., Chaussalet, T., and Chahed, S., 2017. "Ensemble of risk models for emer gency readmissions, Ermer." *International Journal of Medical Informatics,* vol. 103, pp. 65–77. Available: http://www.sciencedirect.com/science/article/pii/ https://doi.org/10.1016/j.ijmedinf.2017.04.010

[13]   Mesgarpour, M., Chaussalet, T., and Chahed, S., 2016. "Risk modelling framework for emer gency hospital readmission, using hospital episode statistics inpatient data." In *IEEE 29th International Symposium on Computer-Based Medical Systems, CBMS.* pp. 219–224.

[14]   AHRQ, 2016. "Ahrq - hcup - clinical classifications software (ccs) for icd-10-cm." Available: https://www.hcup-us.ahrq.gov

[15]   Elixhauser, A. and Steiner, C., 2006. "Readmissions to us hospitals by diagnosis, 2010, Sta tistical brief #153." Available: https://www.ncbi.nlm.nih.gov/pubmed/24006550

[16]   Singer, J. D. and Willett, J. B., 2003. "Applied longitudinal data analysis, Modeling change and event occurrence." Available: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195152968.001.0001/acprof-9780195152968

[17]   Yeo, I.-K. and Johnson, R. A., 2000. "A new family of power transformations to improve normality or symmetry." *Biometrika,* vol. 87, pp. 954-959.