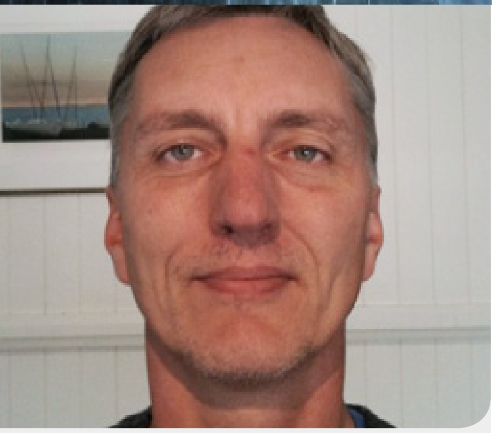

Norm Good

Senior Statistician
CSIRO

norm.good@csiro.au

Norm Good is a Statistician within the CSIRO's Computational Informatics Division. He has been based in the Australian E-Health Research Centre (Brisbane) for the past eight years conducting health related research. Mr Good has considerable expertise in developing and applying variable selection techniques and survival models to health data. Examples of this include novel approaches for estimating optimal colonoscopy screening intervals and developing a risk profiles for patients who are likely to be readmitted to hospital. He has also worked in the field of confidential health data, developing risk and utility measures for regression modelling and promoting the use of remote server statistical querying. Recently Mr Good has exploring the utility of visualisation methods for analysing high-dimensional geometry and multivariate data.

Health informatics visualisation engine: HIVE

Norm Good[a], Chris Bain[b], David Hansen[a], Simon Gibson[a]

[a]The Australian e-Health Research Centre, CSIRO, Queensland
[b]Health Informatics, Alfred Health, The Alfred, Victoria

SUMMARY

This project was designed to integrate, analyse, synthesise and present essential health and hospital information in a highly accessible, agile and visual form – because pictures are worth a thousand words.

We developed a prototype software tool that is;

- capable of drawing on standardised data files that replicate known industry standard, or are easily derivable from such standards
- provides the user (analyst, operational manager, financial manger, executive) with a customisable view of the relative outcomes of, and resources used in, care in a number of dimensions- clinical (LOS, number of adverse events, number of drug doses, attending doctor etc) and financial (surgical, pharmacy, nursing etc) - in one setting
- and identify outliers using advanced statistical modelling techniques.

This tool will generate immediate value for a hospitals' endeavour in continuous operational improvement and will be of particular interest to potential customers throughout Australia given the move to nationally provided Activity Based Funding for hospital services. The tool is a useful way to harness the power of "big data" through advanced analytics.

INTRODUCTION

In Australia, there has been a paradigm shift from input to output (activity) based funding for hospital service delivery. A common output based payment system for managing healthcare provisions involves casemix. Casemix systems are information tools used to classify episodes of patient care. Diagnosis-Related Groups (DRGs) are the most popular casemix system. The rationale from going from an input to output based funding are: i) improving cost efficiency in providing healthcare services; ii) promoting equity for all people to access health services; and iii) increasing incentives to hospitals for providing efficient and quality services.

This shift in funding type has highlighted a number of issues. The one with a significant impact on the costs of care is focused on outliers[1,2]. An outlier in this context is an observation for a patient which is outside the "normal" range expected. Current practices of outlier identification involve a case-by-case analysis, which can be very time consuming. Another issue is to assess the relative clinical and financial value of components of care delivery. These components consist of equipment, drugs and specialist staff which are assessed against mortality and measures of morbidity. Once again, current practice is on a case-by-case basis. Significant cost savings and identification of efficient care models can be achieved if an appropriate software tool can be developed and utilised by front line care.

DESCRIPTION

ANALYTICS ENGINE - Behind the software interface are two main analytics modules. In the first module we can develop models of predicted costs for treating individual patients. In addition to the standard cost attached to a specific DRG, extra costs based on clinical and demographic parameters can be estimated. The main purpose of this modeling was to develop a profile for the "average" patient. The second module focused on identifying outliers in a data set. Given the predicted cost of care calculated above, the actual cost of care is calculated. An outlier in this context is the difference between the actual and predicted cost for a patient which is outside the "normal" range expected. The current method for identifying outliers in Victoria is the L3H3 method[3]. It uses three times the average length of stay for a particular DRG as the high cut point for outliers. We will be using a more robust method based on statistical discordancy to identify outliers[4]. What is potentially more informative is the detection of "inliers". That is, people whose costs are much lower than predicted. Examination of these patients may reveal insights into optimal care models. Given the multivariate nature of the cost and clinical data, dimension reductions tools such as principal components analysis have been employed to project patients and their associated cost of care and clinical status onto a viewable two or three dimensional space.

The baseline data comes from clinical and financial databases from The Alfred Hospital, statistical analyses are undertaken with the R statistical package and the GUI is developed using web-based Java script.

RESULTS

Figure 1 is a four dimensional representation of some of the hospital data. A simple point and click screen in the software produced this plot. The x-axis represents the costs associated with ICU, Allied and Pathology. These costs tend to be related to each other. However, the greatest cost is ICU. The y-axis represents costs associated with Nursing and Medical non-surgery. These axes are derived from a method known as "principal components". The size and colour of the bubbles are representative of total costs. The "Yes" and "No" overlaid onto the bubbles indicate whether the patient was alive at discharge. As you can see some of the biggest costs are associated with patients that died, "No". Figure 2 is a screenshot from the software tool showing a parallel coordinates plot. This plot easily shows outliers or inliers and associated clinical/financial measures. A "brush" can be applied to any vertical axis to select subsets of the data. This is only one example of showing the data. Another option is to undertake statistical discordancy analysis on a subset of disease related groups. This effectively "standardises" costs so that we can compare them in a relative way across patients.

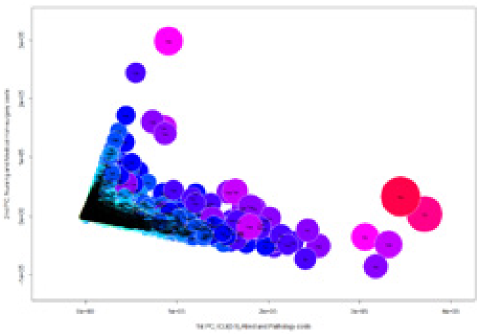

Figure 1. Principal component bubble plot

Figure 2. Parallel coordinates plot.
Different coloured lines represent DRG categories

CONCLUSION

The development of advanced visual analytics capabilities especially those in the bioinformatics sphere[5] can give greater insight into an organisation data than standard reports alone such as those given by platforms such as Tableau ® and Qlikview®. Such capabilities could serve a very useful purpose when it comes to quickly gaining insights from "big data" data sets. Adding advanced multivariable reduction techniques such as principal components analysis and statistical discordancy can add additional insights into identifying outliers and inliers in hospital administrative and clinical data. It is hoped that this tool will aid in the timely identification of outliers and inliers and provide insight into reducing costs in the future.

REFERENCES

1. M Harley, M.A. Mohammed, S Hussain, et al. Was Rodney Ledward a statistical outlier? Retrospective analysis using routine hospital data to identify gynaecologists' performance, BMJ (2005), 330:929.
2. E Polverejan, J.C. Gardiner, C.J. Bradely, et al. Estimating mean hospital cost as a function of length of stay and patient characteristics, Health Economics (2003), 935-42.
3. T.E. McGuire, J.A. Bender, C Maskel. Casemix episodic payment for private health insurance. Canberra: AGPS (1995).
4. A.H. Lee, J Xiao, S.R. Vemuri, Y Zhao. A discordancy test approach to identify outliers of length of hospital stay, Statistics in Medicine 17 (1998), 2199–206
5. Visualising Biological Data: VISBI, http://visbi.org [accessed 10/10/2014].