Estimating Episodes of Care using Linked Medical Claims Data

Graham Williams[1], Rohan Baxter[1], Chris Kelman[2], Chris Rainsford[1], Hongxing He[1], Lifang Gu[1], Deanne Vickers[1], and Simon Hawkins[1]

[1] Enterprise Data Mining
CSIRO Mathematical and Information Sciences
GPO Box 664, Canberra, ACT 2601, Australia
Firstname.Lastname@csiro.au
http://datamining.csiro.au
[2] Commonwealth Department of Health and Ageing
Firstname.Lastname@health.gov.au

Abstract. Australia has extensive administrative health data collected by Commonwealth and state agencies. The value of the health data is how quickly and effectively the data can be converted into useful knowledge independent of its quantity and quality. Using a unique cleaned and linked administrative health dataset we address the problem of empirically defining episodes of care. An episode of care is a time interval containing medical services relating to a particular medical situation. In this paper the medical situation is a hospital admission. The medical services of interest are pathology tests, diagnostic imaging and non-invasive investigative procedures performed before or after the hospital admission, but 'associated' with the hospital admission. The task can be viewed as detecting a signal in a time series relating to a hospital admission, distinct from the background noise of on-going medical care. Our approach uses an ensemble (panel of experts) paradigm where we implement multiple agents (alternative predictive models) to independently estimate intervals and then combine good interval estimates using a voting scheme. The results have been used in a study for the Commonwealth Department of Health.
Keywords: applications, knowledge discovery and data mining, machine learning, record linkage, administrative data, health services.

1 Introduction

In collaboration with the Commonwealth Department of Health and Ageing and Queensland Health, CSIRO Data Mining has created a unique cleaned and linked administrative health dataset bringing together State hospital morbidity data and Commonwealth Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) data. The Queensland Linked Data Set (QLDS) links de-identified, administrative, unit-level data, allowing de-identified patients to be tracked through episodes of care as evidenced by their MBS, PBS and Hos-

pital records [1]. While the dataset is somewhat incomplete[3] it is likely to be invaluable in identifying service utilisation and cost trends and patterns in the overall delivery of Commonwealth and State funded health care.

An ongoing central issue in health services research is how to identify the groups of services and costs relating to a particular episode of care [2–5]. Episodes of care are defined as: *A block of one or more medical services, received by an individual during a period of relatively continuous contact with one or more providers of service, in relation to a particular medical problem or situation* [2]. One example of the application of episodes of care is their use in measuring the costs and services for a particular disease or condition of interest, such as diabetes, asthma or depression.

In our application, the particular medical problem is a hospital admission and the block of related medical services are pathology tests, diagnostic imaging and non-invasive investigative procedures. *A priori* these medical services are likely to be associated with a hospital admission rather than ongoing ambulatory health care. These medical services are not complete, but they contain the most expensive items in pre-admission preparation for a hospital admission and post-discharge care.

We anchor our episode of care around a particular hospital admission. The episode of care begins x days prior to the admission and ends at y days after discharge from hospital. All related medical services received in these two intervals are included in the episode of care. For our purposes services rendered whilst in hospital are ignored.

Previous work using episodes of care has concentrated on one or two narrow clinical areas. This meant that the model for estimating an appropriate episode of care interval could be hand-crafted and clinically assessed. In contrast, our application required episodes of care to be estimated for 666 Diagnostic Related Groups (DRGs) covering the full range of hospital admission types.

A previous study we performed assumed fixed intervals of a 90-day pre-admission interval and a 90-day post-discharge interval for each of the DRGs. These fixed interval assumptions are clearly not clinically valid. For example, the DRG for a broken femur should have a 0-day pre-admission interval (since broken legs are not planned for) and a 40 day post-discharge interval (the average recovery time). A delivery admission DRG for a birth will have a six-month pre-admission time (as women typically obtain pathology tests and diagnostic imaging six-months prior).

Time and resources do not allow for individual assessment of each of the 666 DRGs. Our solution is a robust and automatic means, based on machine learning techniques, to identify and extract the 666 intervals for estimating episodes of care.

[3] Only some 70% of the Queensland Health hospital records have MBS identifiers associated with them to allow matching with MBS and PBS. Further, PBS data is notoriously difficult to match, and incomplete because until recently only Safety Net and Concession Card holders were recorded and Safety Net records link to a family rather than an individual.

Others have explored the problem of identifying cut-points (change-points or segmentation). The problem arises in many applications in data mining, artificial intelligence and statistics, including segmenting time series [6], decision tree algorithms and image processing. A range of criteria have been proposed in the literature for determining if some time series data should be segmented into two or more regions [7]. [8] describe the segmentation of categorical time series data using a voting experts approach to combine evidence for segmentation boundaries. Interestingly, they offer the conjecture that their statistical characteristics for finding segmentation points are domain independent and provide examples to support this.

This paper presents our solution using data driven estimates for the DRG episodes of care resulting from an ensemble of alternative estimation models (or panel of experts) being combined through a voting scheme [9]. We use validation datasets to assess the confidence intervals for our episodes of care estimates and have obtained a preliminary clinical assessment of estimates for five of the 666 DRGs.

In Section 2 we review the data from which we determine the episodes of care. Section 3 presents our methodology for a data driven approach to estimating appropriate pre/post intervals for each DRG while Section 4 presents the results and reviews the usage of the identified intervals for a study performed for the Commonwealth Department of Health and Ageing.

2 Experimental Design

The data used for this study were extracted from the QLDS [1]. Hospital separations (data associated with an episode in hospital) have been used to identify admission and discharge information. For each separation all of that patient's relevant MBS services have been extracted and these form the basis of the service counts and aggregated cost used in this study.

The QLDS hospital data was filtered to remove hospital separations corresponding to changes in admission status that do not reflect end of episode. These records include statistical admissions and discharges as well as hospital transfers[4].

The data were also filtered to remove MBS items provided in hospital, corresponding to MBS item sub-groups with hospital flag set to 'h'.

Only MBS items in the following categories are considered:

- *diagnostic imaging*: MBS items from 55028 to 63946, inclusive;
- *pathology*: MBS items from 65060 to 73811, inclusive;
- *non-invasive investigative procedures*: MBS items from 11000 to 12533, inclusive.

[4] These records are identified by *admission source* codes 4, 6, and 11 and *separation mode*s 2, 6, 10, and 11.

QLDS includes 70% of the total Queensland hospital admissions. It is not expected that this will affect our estimation of episodes of care. However, comparing the results here with other published Queensland data, the number of services and total costs need to be adjusted appropriately to reflect total (100%) service counts and costs.

3 Ensemble Methodology

We have developed a data driven technique to time-frame-adjust DRG intervals of workup (pre-admission) and followup (post-discharge), motivated by [10]. The approach employs a *multiple experts* or *ensemble* paradigm [9] where several change-point estimators are employed and an averaged majority voting scheme is used to determine the final change-points, which then define the pre/post intervals.

A number of alternative schemes were investigated and four were finally chosen to form the *ensemble*: mean and variance optimisation; regression tree; multivariate adaptive regression splines; and multi point splines.

For each DRG a table containing a count of all pre-admission (and separately post-discharge) MBS services in diagnostic imaging, pathology, and non-invasive investigative procedures was constructed. Services were counted on a daily basis over all separations for the particular DRG. These were counted up to 180 days (6 months) pre-admission (and separately 180 days post-discharge). The daily counts were then normalised by dividing by the number of hospital admissions for that DRG.

The choice of 180 days pre/post was made to allow for a background pattern of servicing to be identified and then any intervals of increased servicing could be identified as being associated with the hospital separation.

We describe each of the four methods and then describe how the *ensemble* voting method is employed to determine the final set of pre/post intervals.

3.1 Mean and Variance Optimisation

The approach here is to search for a cut point t_c between T_1 and T_{180} ($T_1 < t_c < T_{180}$) that partitions the 180 day period (pre or post) into two parts $[T_1, t_c]$ and $[t_c, T_{180}]$. The search finds the value of t_c which maximises the difference between the mean of the two partitions and minimises the variance in each partition.

The original service count data are smoothed using a moving average of 3 days prior and 3 days post and then normalised to values between 0 and 1. Suppose then that μ_1 is the normalised mean of $[T_1, t_c]$ and σ_1 is the standard deviation of this interval. Similarly μ_2 and σ_2 for $[t_c, T_{180}]$. The t_c chosen is that which minimises:

$$\frac{1}{|\mu_1 - \mu_2|} + \beta(std_1 + std_2) \tag{1}$$

The parameter β allows fine tuning of the importance of the variance with respect to the mean. By experimentation β was set to 20.

3.2 Regression Tree

Regression trees [11] recursively partition data to build a regression model for separate parts of the data. *Rpart*, the regression tree routine provided by R [12], was used, fitting a constant model to the leaves of the tree. The splitting criterion is based on ANOVA (performing an analysis of the variance on the data) whereby the cut-point maximising the reduction in the squared error fit to the actual data is chosen.

Other parameters chosen for the modelling are:

- *Maximum number of splits*: 3. This was chosen to limit the choice of cut-points. Other settings could be investigated but have not been for this study.
- *Minimum number of data points in a leaf node*: 6. This is the default setting for *Rpart*. A consequence of this choice is that it is not possible for a cut-point to be less than 6 days from the boundaries $[181, 0]$.

3.3 Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) is a spline fitting regression approach where splines are fitted to distinct intervals of the data [13]. The cut points (called knots in MARS) are searched for through an exhaustive approach, optimising a so called loss of fit criterion. The R package [12] *mda* provides the implementation of the function *mars* used for this analysis.

The primary parameter chosen was:

- *Maximum number of terms*: $nk = 3$. By limiting the number of models to 3 (and thus the knots or cut point candidates to 2) the choice of cut-points is made easier.

3.4 Multi Point Splines

A series of natural splines were fitted to the data points using the R *spline* function with the *mean* function to approximate the background level of activity before the workup to admission from the first 100 (of 180) days of the data.

Once the splines have been fitted a search from the day of admission finds the first spline knot (cut point) that falls below the mean value (ignoring the knot at day 0).

The parameters chosen for the spline are:

- *Number of splines*: $n = 51$. Thus, the 360 days (pre and post period for each DRG) is effectively split into 51 segments of approximately one week each.

3.5 Panel of Experts

Each of the modelling approaches discussed in the previous four sections identifies a pre and post interval for each of the 666 DRGs. A successful approach in statistical modelling and machine learning has been the concept of multiple

predictive models being employed through a voting mechanism. This approach is adopted here to combine the proposed pre and post cut points from each of the four "experts" into single pre and post cut points for each DRG.

The method used is to select from the four proposed cut points the three having the least *variance*. We then calculate the *mean* value of these three as the final cut point.

4 Experimental Results

4.1 Example DRGs

We present here a small sample of DRGs to illustrate the resulting pre/post intervals. The first example is perhaps the 'most typical' of the patterns found. However there is a wide diversity of patterns that is not reflected in these example DRGs.

Each figure plots data for 180 days prior to a hospital admission for the specified DRG and 180 after discharge. The data plotted is the daily count of MBS items in diagnostic imaging, pathology, and non-invasive investigative procedures received by all patients, divided by the number of separations for the DRG.

In Figure 1 the pre-admission interval illustrates a marked increase in MBS activity 28 days prior to admission. The post-discharge interval of 56 days illustrates some activity after the episode in hospital. The post-discharge activity then stabilises to a new, but apparently increased, base line of activity.

Figure 2 illustrates what may be an emergency admission where there is very little or no pre-admission MBS workup but there is a period of post-discharge MBS followup activity. The *panel of experts* methodology has identified reasonable estimations in this case.

Figure 3 shows quite a different and unusual pattern, if not unexpected for this DRG (renal dialysis). In such cases the panel of experts approach has essentially identified no particular pre-admission or post-discharge interval and has instead used, approximately, a 90 day pre/post interval. Results for this DRG are then similar to those produced in the previous study where 90 days was used for all DRGs. This DRG is an example of where expert advice might estimate the best pre/post interval to be 0, in which case the methodology could be refined to automatically identify these situations.

Figure 4 shows another class of patterns with a long lead up time before hospitalisation (for birth). There is clearly a lot of activity about 5 months prior to birth, then again about 3 months prior to birth. The methodology has identified 100 days for the pre-admission interval. Post-discharge identifies increased activity about 1 month after discharge.

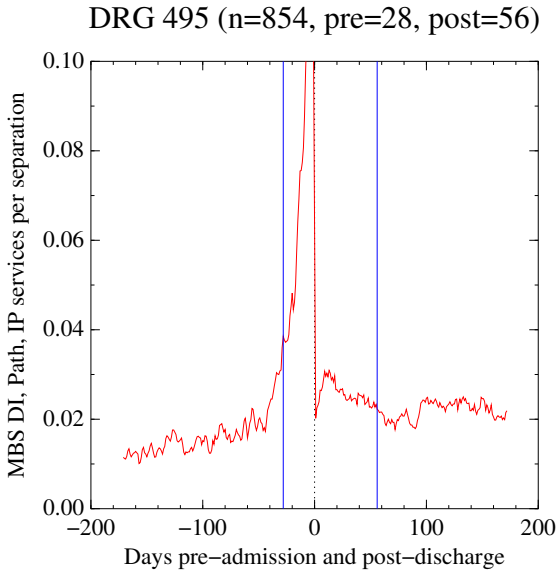

Fig. 1. DRG 495: Major procedures for malignant breast conditions. The estimated pre-admission and post-discharged intervals are shown as the solid vertical lines around the dotted central line at day 0 (day 0 essentially represents a period of time in hospital, collapsed to a single point in the plots). The plot title indicates the DRG population as 854 hospital separations, and the *panel of experts* methodology estimating a pre-admission interval of 28 days and a post-discharge interval of 56 days.

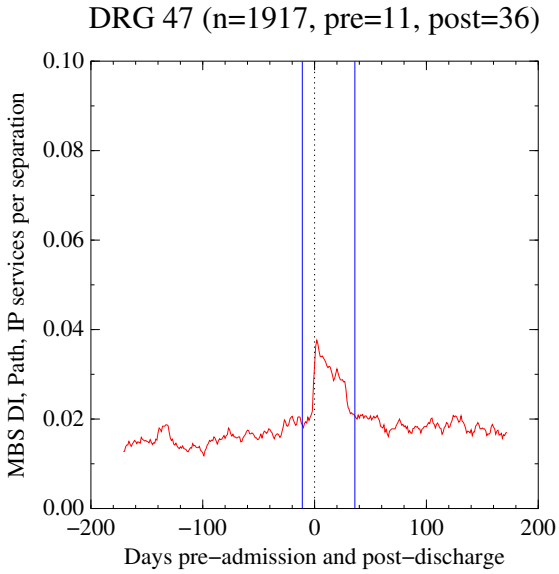

Fig. 2. DRG 47: Seizure, age < 65 without complication and/or comorbidity DRG.

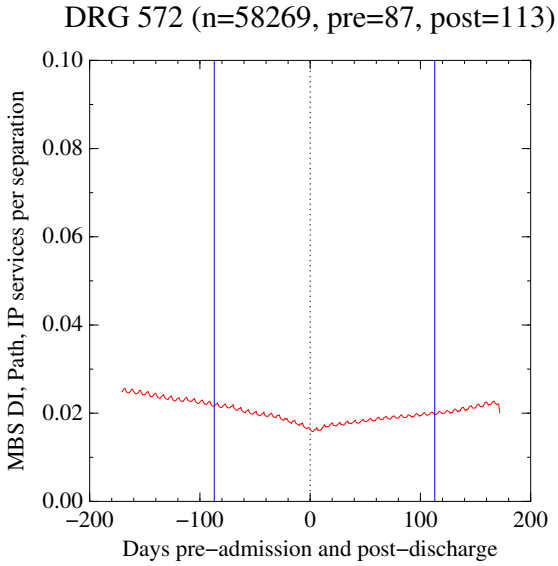

Fig. 3. DRG 572: Hospital admission for renal dialysis.

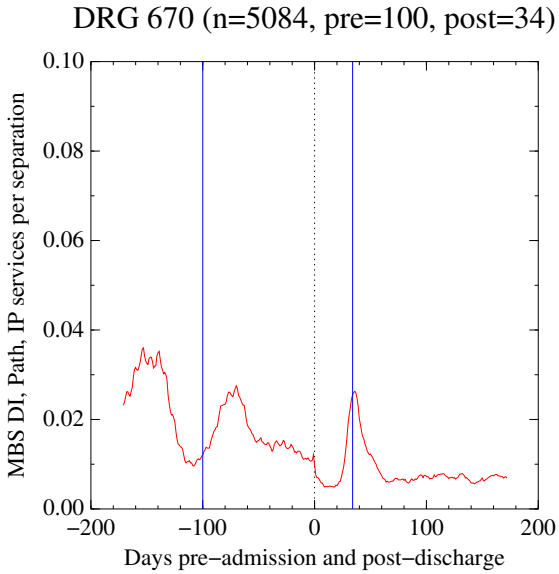

Fig. 4. DRG 670: Cesarean delivery without complicating diagnosis.

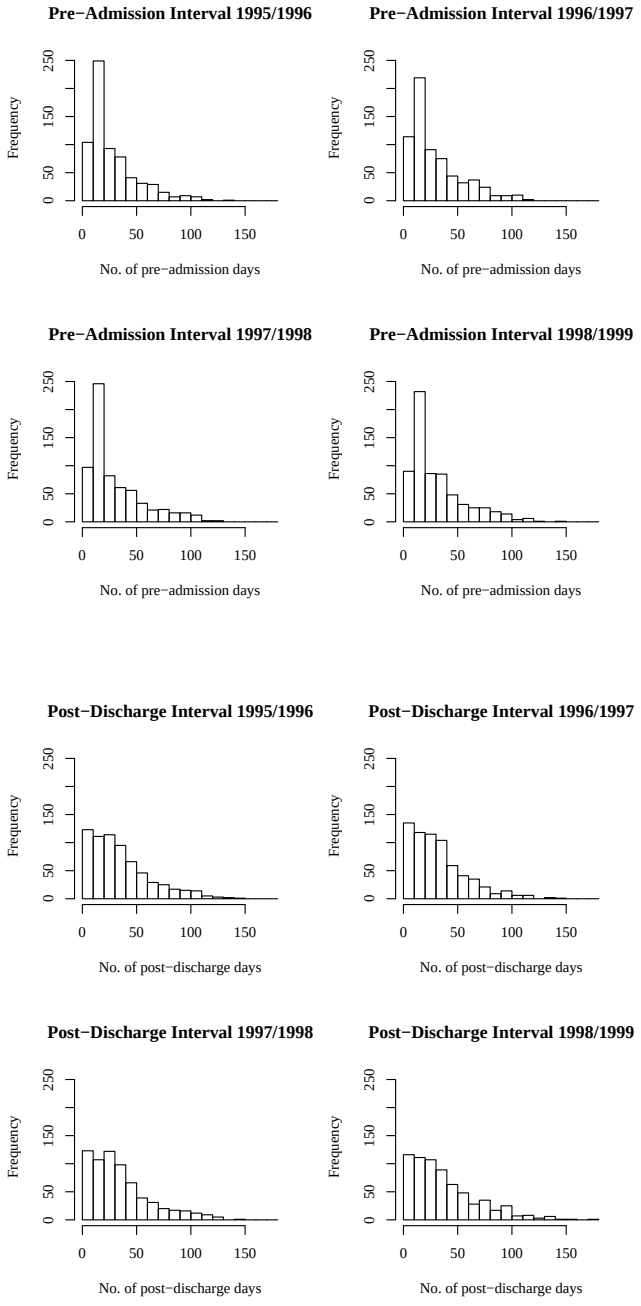

Fig. 5. Pre-admission and post-discharge histogram for all DRGs for all 4 years.

Method	Mean	StdDev
dev	20.5	21.50
tree	18.17	20.06
mars	15.14	18.73
spline	14.60	9.34
vote	13.00	12.21

Table 1. Mean and standard deviation of the mean variation of different methods' estimates across all DRGs and all years. A lower mean variation suggests that the method is more robust to noise (changes from year to year). The combined method, *vote* has the lowest mean variation. This is an advantage of an ensemble prediction method like *vote*.

4.2 Episode of Care Summary Statistics

4.3 Internal Validation of Results

We tested the sensitivity of our episode of care estimation method as follows. We applied the method to each of the four years of data. The variance of the resulting estimates were then calculated. The resulting error bars are shown in Figure 6.

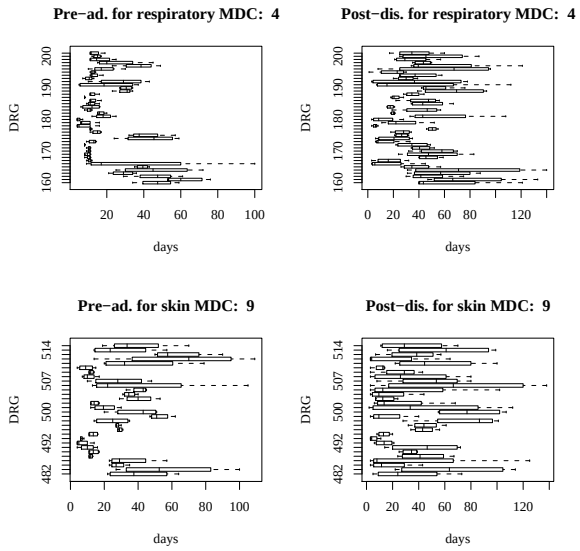

Fig. 6. Two example MDCs and ranges of estimates for pre-admission and post-discharge intervals over 4 years.

4.4 Clinical Assessment of Results

There were too many DRGs and not sufficient time or resources available to obtain a clinical assessment of all our results. However, some DRGs allow *a priori* judgements to be made about suitable intervals for a typical episode of care. There is ongoing validation work involving consultation with the clinical sub-committees responsible for various DRG areas.

5 Discussion and Conclusions

We have introduced an AI-based method for determining intervals of care relevant to hospitalisation. The approach begins with aggregating counts of all patient services prior to and after episodes in hospital. The resulting data is effectively a time series where we are searching for a significant change in the trends. An ensemble approach applies multiple "experts" to solve the problem in different ways and to then combine the results using an averaged voting method. This approach has been found to be quite effective in estimating relevant periods of pre-admission and post-discharge services.

A key assumption underlying this research is that a single pre-admission interval and a single post-admission interval is adequate for capturing individual DRG episodes of care. The implication is that all patients in a particular DRG have similar medical service utilisation patterns. This is more likely to be true for patients from similar age-gender mixes and with similar co-morbidities. For some DRGs patients do have similar characteristics. An obvious example is the DRG covering births where the patients are women of child-bearing age. For other DRGs though there is considerable diversity among the patients. Investigating different intervals of episodes of care within a single DRG was considered outside the scope of this application. It is an interesting area for future work.

Further validation of this approach, particularly by having the intervals automatically discovered reviewed by panels of relevant clinicians, is under way.

References

1. Williams, G., Vickers, D., Baxter, R., Hawkins, S., Kelman, C., Solon, R., He, H., Gu, L.: Queensland linked data set. Technical Report CMIS 02/21, CSIRO Mathematical and Information Sciences, Canberra (2002) Report on the development, structure and content of the Queensland Linked Data Set, in collaboration with the Commonwealth Department of Health and Ageing and Queensland Health.
2. Solon, J.A., Feeney, S.H., Jones, S.H., Rigg, R.D., Sheps, C.G.: Delineating episodes of medical care. American Journal of Public Health **57** (1967) 401–408
3. Wingert, T.D., Kralewski, J.E., Lindquist, T.J., Knutson, D.J.: Constructing episodes of care from encounter and claims data: some methodological issues. Inquiry **32** (1995) 162–170
4. Lestina, D., Miller, T., Smith, G.: Creating injury episodes using medical claims data. The Journal of Trauma **45** (1998) 565–569

5. Schulman, K.A., Yabroff, K.R., Kong, J., Gold, K.F., Rubenstein, L.E., Epstein, A.J., Glick, H.: A claims data approach to defining an episode of care. Pharmacoepidemilogy and Drug Safety **10** (2001) 417–427
6. Tong, H.: Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, New York (1990)
7. Oliver, J.J., Baxter, R.A., Wallace, C.S.: Minimum message length segmentation. In: Research and Development in Knowledge Discovery and Data Mining: Lecture Notes in Artificial Intelligence, Springer (1998) 223–233
8. Cohen, P., Adams, N.: An algorithm for segmenting categorical time series into meaningful episodes. In: Proceedings of the Fourth International Symposium on Intelligent Data Analysis, Lisbon Portugal (2001)
9. Dietterich, T.G.: Ensemble methods in machine learning. In Kittker, J., Roli, F., eds.: Proceedings of the First International Workshop on Multiple Classifier Systems (MCS00): Lecture Notes in Computer Science. Volume 1857., Cagliari, Italy, Spinger (2000) 1–15 `citeseer.nj.nec.com/dietterich00ensemble.html`.
10. Kelman, C.: Monitoring health care using national administrative data collections. PhD thesis, National Centre for Epidemiology and Population Health, Australian National University, Canberra (2000)
11. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Wadsworth, Belmont, CA (1984)
12. Venables, W.N., Smith, D.M., The R Development Team: An Introduction to R. 1.5.0 edn. (2002)
13. Friedman, J.: Multivariate adaptive regression splines. The Annals of Statistics **19** (1991) 1–141