Analysis of an Observational Studies - An Example Using Data from the Irish Cancer Registry.

Dooley, Cara
NUI Galway, School of Maths, Stats and Applied Maths
University Road
Galway, Ireland
E-mail: c.doooley6@nuigalway.ie

Hinde, John
NUI Galway, School of Maths, Stats and Applied Maths
University Road
Galway, Ireland
E-mail: john.hinde@nuigalway.ie

Newell, John
HRB, Clinical Research Facility
Newcastle Road
Galway, Ireland
E-mail: john.newell@nuigalway.ie

Egan, Larry
HRB, Clinical Research Facility
Newcastle Road
Galway, Ireland

Comber, Harry
National Cancer Registry
Cork, Ireland

Introduction

By their very nature, observational studies have an added layer of complexity when compared to Randomised Clinical Trials (RCT). There are often very many more controls than cases and the lack of design in the study may cause the two groups to be very different in terms of measured covariates. Some of the hurdles of analysing observational studies as well as methods to overcome them will be discussed.

Matching methods can be used to implement a 'design' by matching each case to a control or controls, minimising the distance (based on the observed covariates) between each control and case pair. While matching is a useful technique, it can often lead to omitting a large fraction of the dataset, particularly if we use one to one matching. Alternatively, we can use various weighting techniques in the analyses; these weights may be based on the number of controls matched to each case when matching is carried out using all patients, or the inverse propensity score. The propensity score is the conditional probability of being a case given the observed covariates.

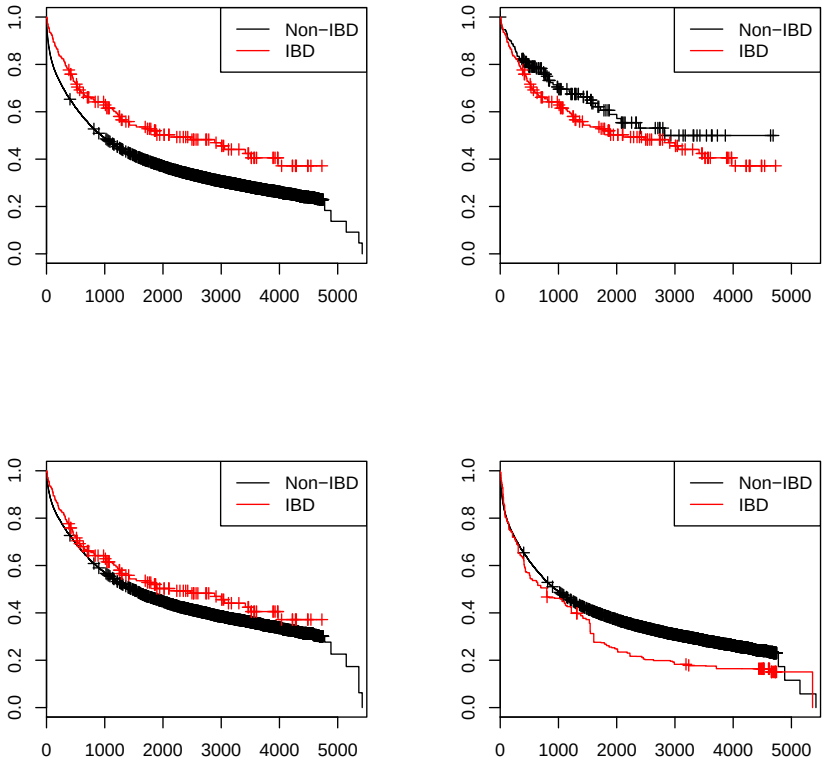
Below the above techniques are discussed further and are illustrated with data from an observational study carried out on colorectal cancer, where the cases of interest were patients with a secondary condition. The data set contained over 20,000 control patients and 170 patients with the secondary condition and is an extreme example of imbalance between the number of cases and controls.

The aim of the study was to compare survival of colorectal cancer patients in the whole popu-

lation against the survival of patients in a sub-population who also had inflammatory bowel disease (IBD). All individuals who suffered from colorectal cancer were drawn from the entire Irish population using data from January 1994 to December 2005 provided by the National Cancer Registry of Ireland (NCRI).

Analysis of the full dataset

Initially, the whole data set was analysed. Kaplan-Meier estimates were examined, as seen in Figure (a). A Cox proportional hazards model was fitted and all factors except for IBD were found to be significant ($p = 0.4121$). These factors included age, gender and various descriptors of the disease, including tumour type, location and stage of illness, i.e., the effect of IBD as seen in Figure (a), was eliminated by covariate adjustment.



Comparison of the four methods used to produced Survival Curves: (a) The whole data set using the conventional KM estimates; (b) The matched dataset using conventional KM estimates; (c) The whole dataset using the number matched as weights; (d) the whole dataset using Adjusted KM estimates.

Propensity Score

The conditional probability of being in the treated group ($Z = 1$) given the observed covariates x, is called the propensity score,

$$e(X) = P(Z = 1|x)$$

The propensity score $e(x)$ balances on observed bias, but not on the unobserved bias. In practice the estimated propensity score $\hat{e}(x)$ is used. To obtain the estimated propensity score, we fit a logistic regression model and use the estimated fit as the estimated propensity score, $\hat{e}(x)$, however other models may be used. The model can be over-fitted, including all variables available at time of diagnosis. The estimated propensity score, $\hat{e}(x)$ will not only balance on observed bias, but also on some of the unobserved bias (Rosenbaum and Rubin, 1983).

Matching

One approach to analysing Observational Studies is to implement a design into the observational study by using matching. In this example, we match the IBD patients to the nearest control by minimizing the Mahalanobis distance between them using the `optmatch` package in R (Hansen, 2007). First, the data was analysed to look at all differences between the two groups at time of diagnosis, those variables found to differ at time of diagnosis were used to create a distance matrix and a rank based Mahalanobis distance was calculated. The optmatch package uses an *optimal* matching algorithm to carry out the match, by minimizing the average distance between the matched pairs. An alternative to the *optimal* matching algorithm is to use a *nearest neighbour* or *greedy* algorithm, where each each IBD patient is matched to the nearest control, that has not already been matched to an IBD patient. When the number of controls is large the greedy and optimal algorithms will produce similar matched sets, Gu and Rosenbaum show that generally optimal matching results with smaller distances within each pair, even when the number of controls is large.

The Mahalanobis distance can have an added calliper (or penalty) calculated using propensity scores, as suggested by Rosenbaum (2010), this will penalise controls that have very different propensity scores from specific cases, making a match between them unlikely. For this analysis $1:1$ matching was carried out, that is, each case was matched to exactly one control. This is wasteful, as much of the data is discarded so we may allow an $n:1$ where for example each case is matched to n controls. Alternatively, all control data can be matched to the cases by allowing the number of controls matched to each case to vary, $n_i:1$ matching.

Analysis of the Matched Data

Following the $1:1$ matching, the resulting dataset was examined with regards to the quality of the match and the controls and cases were found to match almost exactly on the matching variables at of time diagnosis. Again, Kaplan-Meier estimates were again calculated, as shown in Figure (b). As there may still be heterogeneity between the members of a pair that is unexplained by the matching variables, a Cox proportional hazards model with a frailty term was also fitted to compare the risk of death for IBD and non-IBD patients while adjusting for the matching variables. Again, IBD was found to be non-significant ($p = 0.29$), the frailty term was also non-significant ($p = 0.92$), the two variables describing the severity of the illness were still found be significant, all other terms were non-significant.

Alternatives to Matching

While matching is a useful technique, in a simple 1:1 match much of the data remains unused. Some alternatives which are useful in this situation include the Weighted Kaplan-Meier (Winnett and Sasieni, 2002), the Adjusted Kaplan-Meier (Xie and Liu, 2005) and the adjusted Cox proportional hazard model (Sugihara, 2010).

Weighted Kaplan-Meier

Winnett and Sasieni (2002) suggest full matching, that is matching all available controls to cases, that is, allowing $n_i : 1$ matching, and then weighting the Kaplan-Meier estimates by the number of controls matched to each case (n_i).

$$\hat{S}^w(t) = \prod_{u:u\leq t}\left[1 - \frac{\sum_{j=1}^{k} w_j d_j(u)}{\sum_{j=1}^{k} w_j r_j(u)}\right]$$

where, $d_j(u)$ = number of events at time u in stratum j, $r_j(u)$= number at risk at u in stratum j and $w_j = 1/mj$ is the reciprocal of the stratum size. When the same number of controls are matched to each case this reduces to the usual KM estimates. The results of this can be seen in Figure (c).

Adjusted Kaplan-Meier Estimator - AKME

Xie and Liu (2005) suggest using the inverse of the propensity score to weight the Kaplan-Meier, assigning a weight $w_{ik} = 1/p_{ik}$ to each individual, where p_{ik} is the propensity score for individual i in group k.

So the AKME for the kth group is

$$\hat{S}^k(t) = 1 \quad \text{if } t < t_i$$

or

$$\hat{S}^k(t) = \prod_{t_j\leq t}\left[1 - \frac{d^w_{jk}}{Y^w_{jk}}\right] \quad \text{if } t_i \leq t$$

where, d^w_{jk} is the weighted number of events and Y^w_{jk} is the weighted number at risk.

The results of this are shown in Figure (d).

Adjusted Cox Proportional Hazards Model

In the same way that Kaplan-Meier estimates were adjusted using the inverse propensity score as weights, the Cox proportional hazards model may be modified as proposed by Sugihara (2010). After fitting the adjusted Cox proportional hazards model, except for gender, all factors, including IBD, were significant ($p < 0.0001$).

Results

As mentioned, matching is a useful technique, however, when using $1 : 1$ matching, much of the data remains unused. The three methods mentioned is Section 3, all use the whole dataset adjusting for the disparity in numbers between the two groups. The adjusted Cox proportional hazards model is the only model which finds a significant difference between the IBD group and the control. Further work is required to see if this an artifact of the weighting or a true difference.

The propensity score is known to be unstable when the data set is large or contains a great disparity between the number of cases and controls. There are stabilization techniques in the literature that attempt to address this issue, however one such method was applied to this data which showed little effect.

REFERENCES

Hansen, B. (2007). Optmatch: Flexible, Optimal Matching for Observational Studies. *R News*, **7**, 18-24.

Gu, X.S., and Rosenbaum, P.R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*,**2(4)**, 405-420.

Raja Ali, A., Dooley, C., Comber, H., Newell, J., and Egan, L. (2011). Clinical Features, Treatment, and Survival of Patients With Colorectal Cancer With or Without Inflammatory Bowel Disease *Clinical gastroenterology and hepatology* **Article in Press**.

Rosenbaum, P.R. (2010). it Design of observational studies. London: Springer Series in Statistics.

Sugihari, M. (2010). Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score , *Pharmaceutical Statistics*, **9**, 21-24.

Winett, A., and Sasieni, P. (2002). Adjusted Nelson-Aalen estimates with retrospective matching. *Journal of American Statistical Association*, **97**, 245-256.

Xie, J., and Lui, C. (2005). Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine* , **24**, 3089-3110.