



Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

Title	Analysis of an Observational Studies - An Example Using Data from the Irish Cancer Registry
Author(s)	Dooley, Cara; Hinde, John
Publication Date	2011
Publication Information	Cara Dooley, John Hinde, John Newell, Larry Egan and Harry Comber (2011) Analysis of an Observational Studies - An Example Using Data from the Irish Cancer Registry ISI 2011
Item record	<a href="http://hdl.handle.net/10379/3123">http://hdl.handle.net/10379/3123</a>

Downloaded 2019-10-23T14:10:06Z

Some rights reserved. For more information, please see the item record link above.



# Analysis of an Observational Study

Cara Dooley<sup>1</sup>, John Hinde<sup>1</sup>, Harry Comber<sup>3</sup>, Larry Egan<sup>2</sup>,  
John Newell<sup>2</sup>

<sup>1</sup> School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, University Rd, Galway, Ireland. [c.dooley6@nuigalway.ie](mailto:c.dooley6@nuigalway.ie)

<sup>2</sup> HRB Clinical Research Facility, School of Medicine, National University of Ireland, Galway, Ireland.

<sup>3</sup> National Cancer Registry, Cork, Ireland.

**Abstract:** The study presented below aimed to compare survival of colorectal cancer patients against survival of a sub-population with a secondary disease, inflammatory bowel disease (IBD).

The data were taken from an observational study, that is there was no explicit design. The study had many complications, but the most significant aspect was that the number of controls was much greater than the number of cases of interest. Some techniques are used to overcome these obstacles, including: matching of the dataset, to make the controls and cases as similar as possible at time of diagnosis, effectively retrospectively fitting a design; weighting of the data, using both the propensity score and the number of similar patients found in matching.

**Keywords:** Observational Study; Propensity Score; Matching; Kaplan-Meier; Cox Model.

## 1 Introduction

The aim of the study was to compare survival of colorectal cancer patients in the whole population 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). The control group contained many more observations ( $n > 20000$ ) when compared to the IBD group ( $n = 170$ ). Given the number of control patients, there was large diversity in this group. In a conventional designed experiment or trial, patients entering the trial would be randomised across arms of the study, with similar numbers in each group. Usually patients would be similar in age, health, etc. As this was an observational study, there was no design prior to collecting the data and so no benefit, in terms of bias protection, from randomization in terms of the balance of the distribution of unobserved explanatory variables.

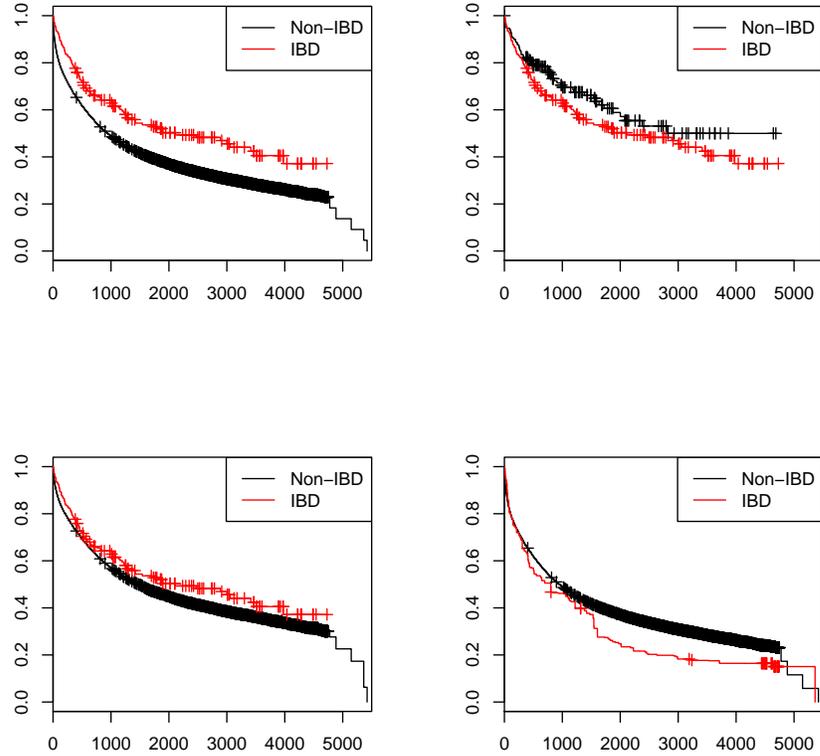


FIGURE 1. Comparison of the four methods used to produced Survival Curves: (a) The whole data set using the conventional KM estimates; (b) The matched dataset using conventional KM estimates; (c) The whole dataset using the number matched as weights; (d) the whole dataset using Adjusted KM estimates.

### 1.1 Analysis of the full dataset

Initially, the whole data set was analysed. Kaplan-Meier estimates were examined, as seen in Figure 1(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 1(a), was eliminated by covariate adjustment.

## 2 Matching

One approach to implement a design in an observational study is to use 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. The Mahalanobis distance has an added calliper (or penalty) calculated using propensity scores, as suggested by Rosenbaum (2010). Following the matching, Kaplan-Meier estimates were again calculated, as shown in Figure 1(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 to be significant, all other terms were non-significant.

### 2.1 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).

## 3 Analysis

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).

### 3.1 Weighted Kaplan-Meier

Winnett and Sasieni (2002) suggest full matching, that is matching all available controls to cases and then weighting the Kaplan-Meier estimates

by the number of controls matched to each case.

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

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  $w_j = 1/m_j$  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 1(c).

### 3.2 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  $k$ th group is

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

or

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

where,  $d_{jk}^w$  is the weighted number of events and  $Y_{jk}^w$  is the weighted number at risk.

The results of this are shown in Figure 1(d).

### 3.3 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 ( $p < 0.0001$ ).

## 4 Results

As mentioned, matching is a useful technique, however, when using 1 : 1 matching, much of the data remains unused. The three methods mentioned in 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.

**Acknowledgments:** This work was supported by Science Foundation of Ireland grant 07/MI/012

## References

- Hansen, B. (2007). Optmatch: Flexible, Optimal Matching for Observational Studies. *R News*, **7**, 18-24.
- Rosenbaum, P.R. (2010). *Design of observational studies*. London: Springer Series in Statistics.
- Rosenbaum, P. R. and Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**, 41-55.
- Sugihari, M. (2010). Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score. *Pharmaceutical Statistics*, **9**, 21-24.
- Winnett, 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.