Observational Studies, Matching and Propensity Scores: Applied to Colorectal Cancer Data

John Hinde & Cara Dooley

School of Mathematics, Statistics and Applied Mathematics,
National University of Ireland, Galway

john.hinde@nuigalway.ie

Research Supported by SFI Award 07/MI/012

ROeS 2013, Dornbirn

10 September 2013

NUI Galway
OÉ Gaillimh

Outline

Observational Studies

- Observational studies are studies with no randomised allocation of treatment.

Observational Studies

- Observational studies are studies with no randomised allocation of treatment.
- There may be no treatment allocation process and subjects may, in some way, be self-selecting to the treatment they receive.

John Hinde (NUI Galway) 10 September 2013 3 / 34

Observational Studies

- Observational studies are studies with no randomised allocation of treatment.
- There may be no treatment allocation process and subjects may, in some way, be self-selecting to the treatment they receive.
- They are often used when a Randomised Control Trial may be unethical or impossible.

Observational Studies

- Observational studies are studies with no randomised allocation of treatment.
- There may be no treatment allocation process and subjects may, in some way, be self-selecting to the treatment they receive.
- They are often used when a Randomised Control Trial may be unethical or impossible.
- Much routinely collected data (registries) can be considered as observational studies.

Observational Studies

- Observational studies are studies with no randomised allocation of treatment.
- There may be no treatment allocation process and subjects may, in some way, be self-selecting to the treatment they receive.
- They are often used when a Randomised Control Trial may be unethical or impossible.
- Much routinely collected data (registries) can be considered as observational studies.
- Since they lack design or control of potentially important confounding factors, the estimate of treatment effect may be biased.

Randomised Control Trials

- Randomised Control Trials (RCT) are the gold standard in design of clinical experiments.
- Individuals are recruited to the study subject to some explicit criteria and then randomly allocated to the treatment and control groups.
- Using the design of the RCT, any factors which are likely to introduce bias in the estimate of the treatment effect can be controlled for.
- The hope is that randomisation accounts for all other factors, on average.

Colorectal Cancer and IBD

- An observational study using data from the National Cancer Registry of Ireland (NCRI)

  (Thanks to Raja Affendi, Harry Comber and Laurence Egan)
- All individuals with colorectal cancer from January 1994 to December 2005 (22323 patients)
- Subgroup of interest those with a secondary disease **inflammatory bowel disease (IBD)** (170 patients)

Colorectal Cancer and IBD

- An observational study using data from the National Cancer Registry of Ireland (NCRI)

  (Thanks to Raja Affendi, Harry Comber and Laurence Egan)
- All individuals with colorectal cancer from January 1994 to December 2005 (22323 patients)
- Subgroup of interest those with a secondary disease **inflammatory bowel disease (IBD)** (170 patients)

Question of interest:

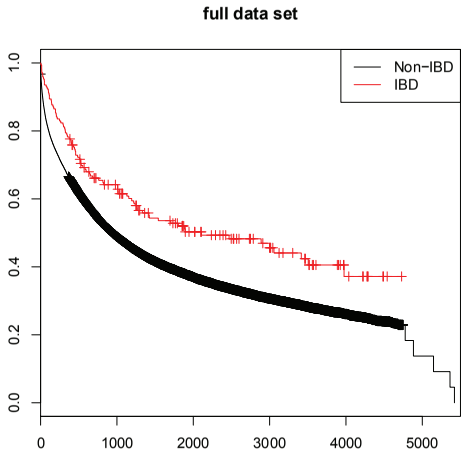
Do individuals with the secondary condition have better or worse survival prospects?

The Data

Patient ID	Unique anonymised identifier for each patient.
Age	Age of patient at time of diagnosis.
ICD10	Classification of disease location, from the International Classification of Diseases for Oncology.
Grade	Grade of the tumour at time of diagnosis.
Stage	Stage of the tumour at time of diagnosis.
Smoking	Smoking status of the individual.
Gender	Patient's gender.
Survival	Survival time for the individual after diagnosis. The study runs from January 1994 to December 2005, individuals who survived beyond this time are censored.
Censor	Indicates if the data is censored or not, that is if the individual survived past end of study.
IBD	Indicates if patient had IBD or not.

Complete Data Set : Kaplan-Meier

Strong evidence of a marginal effect of IBD



Whole Data Set

- Median survival time for the 170 IBD patients is 2107 days
- Median survival time for the 22153 control group is 942 days
- But . . .

Whole Data Set

- Median survival time for the 170 IBD patients is 2107 days
- Median survival time for the 22153 control group is 942 days
- But . . .
- Strong associations between IBD and other covariates

Whole Data Set

- Median survival time for the 170 IBD patients is 2107 days
- Median survival time for the 22153 control group is 942 days
- But ...
- Strong associations between IBD and other covariates
- Cox PH model was fitted with main effects for the observed covariates
  - IBD was non-significant with a p-value of 0.41
  - All other main effects were significant

Propensity Score (Rosenbaum and Rubin, *Biometrika*, 1983)

A commonly used tool in Observational Studies, where there is little or no design to balance the data on the observed covariates

Propensity Score (Rosenbaum and Rubin, *Biometrika*, 1983)

A commonly used tool in Observational Studies, where there is little or no design to balance the data on the observed covariates

The **propensity score**, for subject i, is the conditional probability of being in the group of interest ($Z = 1$) given the observed covariates x,

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

Propensity Score (Rosenbaum and Rubin, *Biometrika*, 1983)

A commonly used tool in Observational Studies, where there is little or no design to balance the data on the observed covariates

The **propensity score**, for subject i, is the conditional probability of being in the group of interest ($Z = 1$) given the observed covariates x,

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

- To obtain the propensity score:
  - fit a model to the groups of interest (IBD or not)
  - use the estimated fit as the estimated propensity score, $\hat{e}(x)$
  - here used a logistic regression model, but any suitable family of models may be used.
  - the model can be over-fitted, including all variables available.

Uses of the Propensity Score

- The propensity score is used to remove the bias due to the difference in the observed covariates (at baseline).
- This difference is sometimes referred to as the *overt bias*.
- It cannot control for the *hidden bias*, that is the bias due to the unobserved covariates.

Uses of the Propensity Score

- The propensity score is used to remove the bias due to the difference in the observed covariates (at baseline).
- This difference is sometimes referred to as the *overt bias*.
- It cannot control for the *hidden bias*, that is the bias due to the unobserved covariates.

The propensity score can be used

- as a covariate in the model
- to stratify the data
- to **match** the data
- as a **weight** in the model — Inverse Probability Treatment Weights (IPTW)

Matching

- Ideally, we would start with a designed experiment, but this is not always possible, e.g. Observational Studies
- Attempt to add design to the observational study, by matching controls and cases based on recorded covariates
- Aim to obtain two groups (cases/controls) of patients similar at time of diagnosis

Matching

- Form a similarity matrix using the variables that were found to differ between the two groups.
- Commonly used choices for the measure of similarity are:
  - propensity scores
  - Mahalanobis distances
  - a combination of the two.
- Here, we used a rank based Mahalanobis distance with a calliper (or penalty) based on the propensity score.
- The control and case that are closest are matched
- Use `optmatch` package in R (Hansen et al)
  - Uses an Optimal matching algorithm
  - Matches by minimizing the average distance between the matched pairs

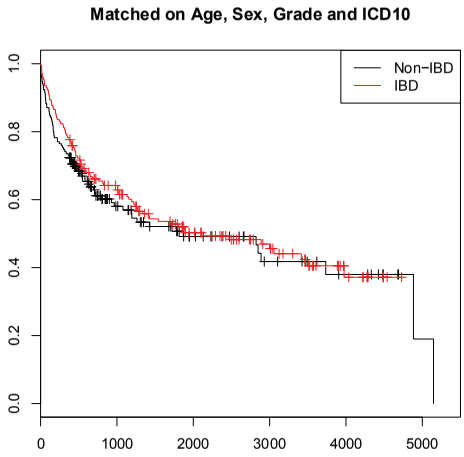
Matching

	1	2	3	4	5	6	7
A	3.6	5.9	6.4	7.2	5.1	4.1	**2.8**
B	4.2	11.1	16.3	17.9	9.1	4.1	7.1
C	4.8	4.2	∞	5.1	6.2	3.9	3.0
D	7.3	13.7	9.4	10.9	9.3	13.4	8.6
E	4.4	4.9	6.2	6.8	11.6	2.3	3.5
F	1.5	0.3	5.6	5.9	3.1	**1.3**	5.6
G	4.5	8.6	6.1	7.3	5.4	∞	5.8
H	4.0	8.2	8.4	9.7	8.6	7.5	7.8
I	3.1	**0.0**	4.6	4.6	5.1	2.1	6.2
J	5.1	2.1	9.5	9.5	3.3	6.0	12.2
K	**1.4**	4.0	8.8	9.7	**1.0**	5.5	9.3
L	5.5	9.0	9.5	10.7	7.6	11.1	11.9
M	5.1	3.3	3.6	3.9	9.6	5.8	6.9
N	6.1	3.1	**3.4**	**3.5**	10.6	6.2	7.5

Matching

	1	2	3	4	5	6	7
A	3.6	5.9	6.4	7.2	5.1	4.1	**2.8**
B	4.2	11.1	16.3	17.9	9.1	4.1	7.1
C	4.8	4.2	∞	5.1	6.2	3.9	3.0
D	7.3	13.7	9.4	10.9	9.3	13.4	8.6
E	4.4	4.9	6.2	6.8	11.6	2.3	3.5
F	1.5	0.3	5.6	5.9	3.1	**1.3**	5.6
G	4.5	8.6	6.1	7.3	5.4	∞	5.8
H	4.0	8.2	8.4	9.7	8.6	7.5	7.8
I	3.1	**0.0**	4.6	4.6	5.1	2.1	6.2
J	5.1	2.1	9.5	9.5	3.3	6.0	12.2
K	**1.4**	4.0	8.8	9.7	**1.0**	5.5	9.3
L	5.5	9.0	9.5	10.7	7.6	11.1	11.9
M	5.1	3.3	3.6	3.9	9.6	5.8	6.9
N	6.1	3.1	**3.4**	**3.5**	10.6	6.2	7.5

After Matching:Kaplan-Meier



After Matching

- The median survival time for the 170 IBD patients is 2107 days
- The median survival time for the 170 matched control patients is 1869 days
  - compared to a median survival time of 942 for the whole set of controls

After Matching

- The median survival time for the 170 IBD patients is 2107 days
- The median survival time for the 170 matched control patients is 1869 days
  - compared to a median survival time of 942 for the whole set of controls
- Cox PH model was fitted to the matched data with the observed covariates
  - Only Grade and Age are now significant
  - IBD is non-significant

After Matching

- The median survival time for the 170 IBD patients is 2107 days
- The median survival time for the 170 matched control patients is 1869 days
  - compared to a median survival time of 942 for the whole set of controls
- Cox PH model was fitted to the matched data with the observed covariates
  - Only Grade and Age are now significant
  - IBD is non-significant
  - However, Gender*IBD and Smoking Status*IBD interactions significant

John Hinde (NUI Galway) 10 September 2013 15 / 34

After Matching

- The median survival time for the 170 IBD patients is 2107 days
- The median survival time for the 170 matched control patients is 1869 days
  - compared to a median survival time of 942 for the whole set of controls
- Cox PH model was fitted to the matched data with the observed covariates
  - Only Grade and Age are now significant
  - IBD is non-significant
  - However, Gender*IBD and Smoking Status*IBD interactions significant

But, have thrown away most of the control data!

Multiple Matching

- Matching above is 1:1
- Matching can be m:1 — greater use of controls
  - m can be allowed to vary so that each case has a different number of matches, m_i:1 matching (match controls to closest cases)
  - This allows full use of all the control data

Weighted Kaplan-Meier

- Winnett and Sasieni (JASA 2002) suggest full matching with stratification and 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 u in stratum j
$r_j(u)$= number at risk at u in stratum j
$w_j = \frac{1}{m_j}$ reciprocal of the stratum size

- If the same number of controls are matched to each case this reduces to the usual KM estimates.

Adjusted Kaplan-Meier Estimator - AKME

- Xie and Liu (Stats in Med, 2005) suggest using the inverse of the propensity score to weight the Kaplan-Meier. (AKME)
- Assign a weight to each individual, using the inverse of the propensity score $w_{ik} = 1/p_{ik}$
- Weighted number of events

$$d_{jk}^w = \sum_{i:T_i=t_j} w_{ik}\delta_i I(Z_i = k) = \sum_{i:T_i=t_j} \frac{\delta_i I(Z_i = k)}{p_{ik}}$$

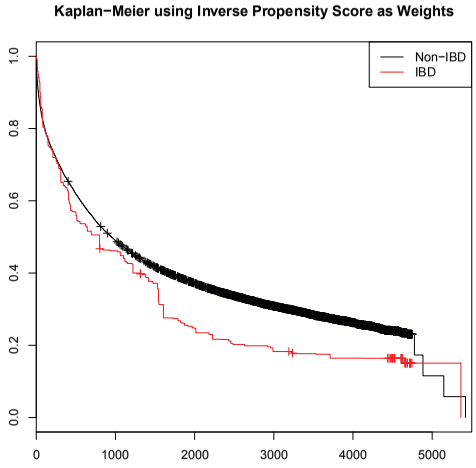
- Weighted number at risk

$$Y_{jk}^w = \sum_{i:T_i\geq t_j} w_{ik} I(Z_i = k) = \sum_{i:T_i\geq t_j} \frac{I(Z_i = k)}{p_{ik}}$$

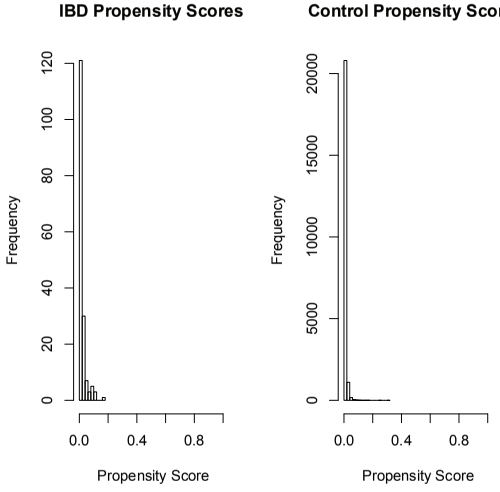
- So the AKME for the *kth* group is

$$\hat{S}^k(t) = \begin{cases} 1 & \text{if } t < t_i \\ \prod_{t_j\leq t}[1 - d_{jk}^w / Y_{jk}^w] & \text{if } t_i \leq t \end{cases}$$

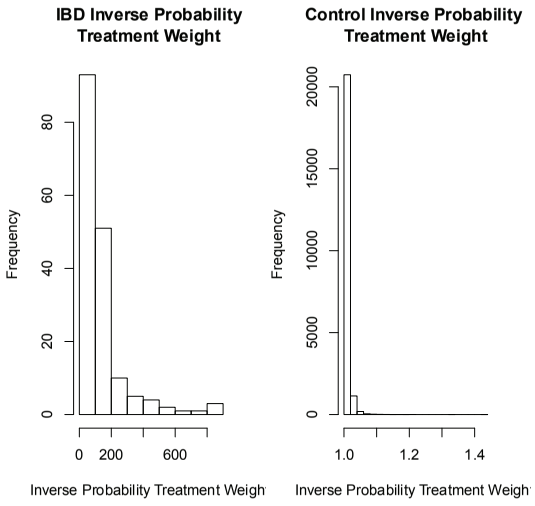
AKME



Propensity Scores



Propensity Scores



Stabilising Techniques for the Propensity Score

One approach to this problem is to adjust the propensity score weights and take account of the group size using **stabilised weights**

$$w_i = \frac{P(Z = k)}{P(Z = k|X = x_i)}$$

Where, k indicates the group ($Z = 0$ is the control group and $Z = 1$ is the treatment group).

Stabilising Techniques for the Propensity Score

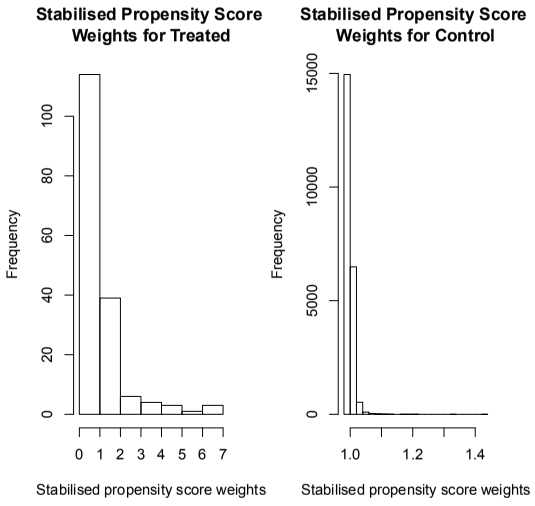
One approach to this problem is to adjust the propensity score weights and take account of the group size using **stabilised weights**

$$w_i = \frac{P(Z = k)}{P(Z = k|X = x_i)}$$

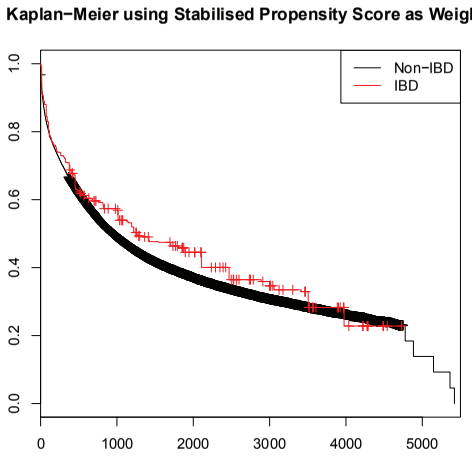
Where, k indicates the group ($Z = 0$ is the control group and $Z = 1$ is the treatment group).

Applying this here gives a slight improvement.

Stabilising Techniques for the Propensity Score



Stabilising Techniques for the Propensity Score



Repeated sampling estimate of Propensity Score

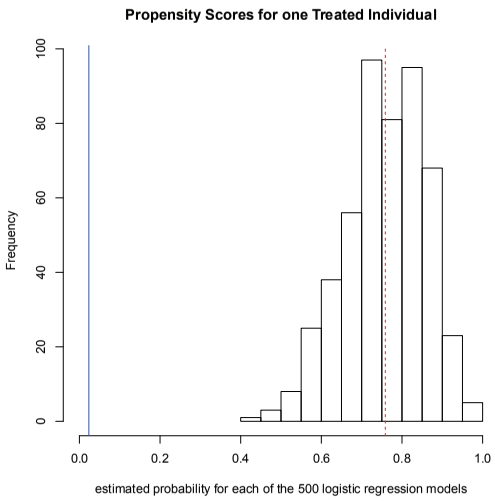
What about sampling the large group of controls?

Repeated sampling estimate of Propensity Score

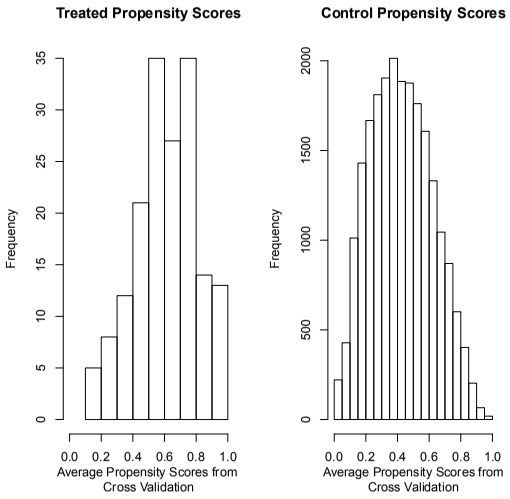
What about sampling the large group of controls?

- Take samples of *size 170* from the controls, *without replacement* ($n_c = n_t$)
- Fit logistic regression model to estimate propensity scores
- Use fitted model to calculate the estimated propensity score for each individual.
- Repeat 500 times and calculate the average propensity score for each individual.

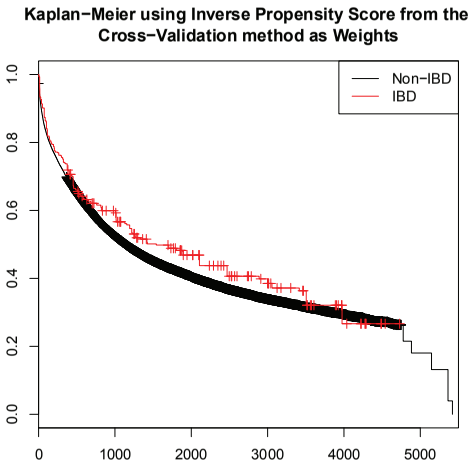
Repeated sampling estimate of Propensity Score



Average Propensity Score



Average Propensity Score



Bias reduction model

Can we improve the estimation of the propensity scores?

Bias reduction model

Can we improve the estimation of the propensity scores?

Firth (1993) proposed a bias reduction method for glms. It involves modifying the likelihood function

$$L(\beta)^* = L(\beta)|\mathcal{I}|^{0.5}$$

where $|\mathcal{I}|^{0.5}$ is Jeffrey's Invariant Prior based on the information matrix.

Bias reduction model

Can we improve the estimation of the propensity scores?

Firth (1993) proposed a bias reduction method for glms. It involves modifying the likelihood function

$$L(\beta)^* = L(\beta)|\mathcal{I}|^{0.5}$$

where $|\mathcal{I}|^{0.5}$ is Jeffrey's Invariant Prior based on the information matrix.

For binary logistic regression this amounts to modifying the response (0/1) and the binomial denominator (1).

Bias reduction model

Can we improve the estimation of the propensity scores?

Firth (1993) proposed a bias reduction method for glms. It involves modifying the likelihood function

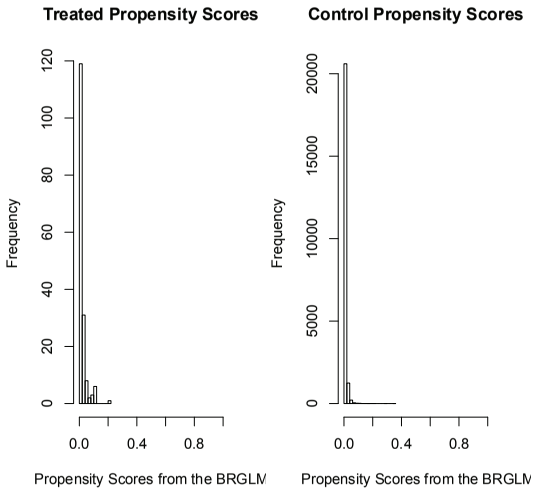
$$L(\beta)^* = L(\beta)|\mathcal{I}|^{0.5}$$

where $|\mathcal{I}|^{0.5}$ is Jeffrey's Invariant Prior based on the information matrix.

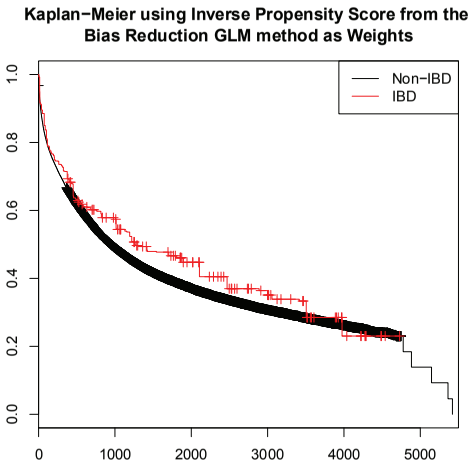
For binary logistic regression this amounts to modifying the response (0/1) and the binomial denominator (1).

Available in the R-package `brglm`.

Bias reduction model



Bias reduction model



Conclusions

- The analysis of observational studies is likely to become more commonplace as vast amounts of data are recorded routinely
- As in randomised control trials thought should be given the 'design' of the study before analysis
- More than one method for allowing for the observational facet of the data should be used
- While matching is the most intuitive, it often means that large amounts of the data are not included in the analysis
- When the numbers in the treatment and control groups are largely imbalanced estimation of the propensity score can be difficult

Further Work

- Which of these methods of propensity score adjustment is **better**?

  Simulation study under way.

Further Work

- Which of these methods of propensity score adjustment is **better**?

  Simulation study under way.
- What is to be gained by making greater use of available controls?

Further Work

- Which of these methods of propensity score adjustment is **better**?

  Simulation study under way.
- What is to be gained by making greater use of available controls?
  - improved precision
  - but are we answering the same question?
- The IPTW methods can be applied in more general modelling approaches . . .

John Hinde (NUI Galway) 10 September 2013 33 / 34

References

- David Firth.
  Bias reduction of maximum likelihood estimates.
  Biometrika, 80:27–38, 1993.
- Ben B. Hansen and Stephanie Olsen Klopfer.
  Optimal full matching and related designs via network flows.
  Journal of Computational and Graphical Statistics, 15(3):609–627, 2006.
- Paul R. Rosenbaum.
  Design of Observational Studies.
  Springer-Verlag Inc, 2010.
- Angela Winnett and Peter Sasieni.
  Adjusted Nelson-Aalen estimates with retrospective matching.
  Journal of the American Statistical Association, 97(457):245–256, 2002.
- Jun Xie and Chaofeng Liu.
  Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data.
  Statistics in Medicine, 24(20):3089–3110, 2005.