HEDG Working Paper 08/20

Using propensity score methods to analyse individual patient – level cost – effectiveness data from observational studies

Andrea Manca

Peter C Austin

July 2008

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

Using propensity score methods to analyse individual patient-level cost-effectiveness data from observational studies

Andrea Manca [1,*,†] and Peter C Austin [2,3]

[1] *Centre for Health Economics, The University of York, York, United Kingdom*
[2] *Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada*
[3] *Department of Health Policy, Management and Evaluation, University of Toronto, Ontario, Canada*

SUMMARY

The methodology relating to the statistical analysis of individual patient-level cost-effectiveness data collected alongside randomised controlled trials (RCTs) has evolved dramatically in the last ten years. This body of techniques has been developed and applied mainly in the context of the randomised clinical trial design. There are, however, many situations in which a trial is neither the most suitable nor the most efficient vehicle for the evaluation. This paper provides a tutorial-like discussion of the ways in which propensity score methods could be used to assist in the analysis of observational individual patient-level cost-effectiveness data. As a motivating example, we assessed the cost-effectiveness of CABG *versus* PTCA – one year post procedure - in a cohort of individuals who received the intervention within 365 days of their index admission for AMI. The data used for this paper were obtained from the Ontario Myocardial Infarction Database (OMID), linking these with data from the Canadian Institute for Health Information (CIHI), the Ontario Health Insurance Plan (OHIP), the Ontario Drug Benefit (ODB) program, and Ontario Registered Persons Database (RPDB). We discuss three ways in which propensity score can be used to control for confounding in the estimation of average cost-effectiveness, and provide syntax codes for both propensity score matching and cost-effectiveness modelling.

KEY WORDS: Cost, cost-effectiveness, propensity score, revascularisation, statistical methods

* Correspondence to: Dr A. Manca, Centre for Health Economics, The University of York, Alcuin A, Heslington, York, YO10 5DD, United Kingdom. Tel. +44 (0)1904 321430; Fax: +44 (0)1904 321402;
† E-mail: am126@york.ac.uk

1 Introduction

The methodology relating to the statistical analysis of individual patient-level cost-effectiveness data collected alongside randomised controlled trials (RCTs) has evolved dramatically in the last ten years. Major contributions have focused on the appropriate way for estimating sampling uncertainty around the incremental cost-effectiveness ratio (ICER) [1-10] and the consequent paradigm shift towards the net benefits formulation [11, 12], the development of the cost-effectiveness acceptability curve (CEAC) as a vehicle to represent decision uncertainty [5, 13-15], the use of a regression-based framework for cost-effectiveness analysis (CEA) [11, 16-18], and the application of Bayesian concepts to model cost-effectiveness data [19-30].

This body of techniques has been developed and applied in healthcare economic evaluation mainly in the context of experimental study design (i.e. the RCT). In this setting, randomisation of the units of interest (e.g. patients) ensures the balance (on average) of measured (and unmeasured) characteristics between treatment arms, hence protecting against bias in the estimation of the *average treatment effect* (e.g. log-odds ratio, differential mean cost).

There are, however, many situations in which a RCT is neither the most suitable nor the most efficient vehicle for the evaluation. Ethical considerations may prevent randomisation, providing a strong argument for exploring existing non-randomised data before setting up a new clinical trial. Similarly, financial reasons may suggest that funding trials in certain disease areas, or for which an amount of (randomised and non-randomised) evidence already exists, may not be an efficient use of Research and Development (R&D) resources. Furthermore, RCTs are often characterised by a follow-up duration which is too short to allow an accurate evaluation of the long-term costs and effects of a given health technology. Again, funding trials with very long follow up periods may require too large a sample size to accommodate inevitable attrition rates. Even if such long-term follow up studies were funded, there is always the risk that by the time the study results become available they may no longer be relevant (e.g. change in clinical practice, different relevant comparators, etc). Another instance in which a RCT may not be the most efficient way to produce clinical and economic evidence is when, although trial evidence exists, this relates to a different study population, hence limiting the external validity of the available evidence to the context or jurisdictions in which decisions are to be made.

In all these circumstances the analysis of observational data (e.g. surveys, registries, administrative records, and census data) [31] offers a potential solution to reconcile the need for an efficient use of limited healthcare R&D resources [32, 33] and timely generation of relevant evidence for decision-making.

Evidence derived from observational studies has been traditionally considered prone to bias. Here, patients' allocation (i.e. selection) to a given treatment is not under the control of the investigators, with the consequent potential risk that the average treatment effect could be confounded with subject characteristics (i.e. treated subjects may differ systematically from untreated subjects). Statistical methods developed in both medical statistics and economics during the last twenty years provide a powerful set of tools for the analysis of non-experimental effectiveness and cost-effectiveness data.

The use of propensity score methodology in healthcare research [34-51] is rapidly gaining popularity [52], although examples of its application in cost-effectiveness analysis are still limited [43, 53-57]. The objective of this paper is to illustrate different ways in which propensity score methods can be used to analyse observational individual patient-level cost-effectiveness data with an aim to estimate average measures of cost-effectiveness. The methods are illustrated using data from the Canadian province of Ontario, comparing the cost-effectiveness of Percutaneous Transluminal Coronary Angioplasty (PTCA) versus Coronary Artery Bypass Grafting surgery (CABG) in post-Acute Myocardial Infarction (AMI) patients.

The manuscript is structured as follows. In the next section we review the general principles of propensity score methodology and cost-effectiveness analysis, showing how to integrate the two in a coherent framework. In section 3 we introduce the motivating example. The results are presented next, followed by the discussion and conclusion section.

2 Methods

2.1 Propensity score methods

While the analysis of RCT data relies on the fact that randomisation ensures (on average) estimated treatment effects are unbiased, the same estimates derived from observational data may be prone to bias in that patients' characteristics could be confounded with treatment allocation.

Traditionally, researchers have tried to address the issue of confounding using multivariate matching methods, regression adjustment or stratification. These approaches though have limitations. Multivariate matching is impractical and often impossible when there is a large number of covariates. Regression adjustment requires the joint distribution of the covariates to be approximately the same between treatment groups. Stratification has limitations in that as the number of covariates increases the number of subclasses increases exponentially, making it difficult to create strata that contain both treated and untreated subjects.

An alternative approach is to use *propensity score* methodology. The propensity score for an individual is the probability of being assigned to either treatment or control, given the value of a set of observed covariates [37, 58]. It has been shown under the assumption that there are no unobserved factors which might give rise to selection bias (*ignorable treatment assignment assumption*), conditioning on the propensity score ensures allocation to either the control or the intervention therapy for a given individual is independent of the treatment outcome [37]. In other words, conditioning on the propensity score allows unbiased estimation of average treatment effect [58].

More formally, the propensity score for subject *i* is defined as

$$\Pr(T_i = 1 \mid \mathbf{X}_i = x_i) \tag{1}$$

where $T_i = 1$ if the subject receives the intervention and 0 if he or she receives the control. It is assumed that, conditional on a set of explanatory variables $\mathbf{X}_i$, the T_i are independent. This probability can be easily estimated using either a *logit* or a *probit* regression. In this paper we use the former.

Once estimated, it is necessary to predict the individual-level *propensity* to be assigned to either the control or the intervention therapy. This facilitates the creation of a *quasi-randomised* experiment, in that two individuals - one receiving the treatment, the other the control – having very similar propensity scores, can be thought of as having been randomised.

Using the relationship between odds and probabilities, the predicted propensity score ($\hat{Z}_i$) for individual *i* is obtained as

$$\hat{Z}_i = \left[\frac{\exp(\hat{\alpha}_o + \sum_{k=1}^{K} \hat{\alpha}_k X_{ik})}{1 + \exp(\hat{\alpha}_o + \sum_{k=1}^{K} \hat{\alpha}_k X_{ik})} \right] = \frac{1}{1 + \exp[-(\hat{\alpha}_o + \sum_{k=1}^{K} \hat{\alpha}_k X_{ik})]} \tag{2}$$

where the $\hat{\alpha}'s$ coefficients are those obtained from the *logit* model in (1).

Let us assume, for the sake of argument, that the analysis is concerned with a single continuous outcome measure (e.g. total cost). The predicted propensity score at individual level can now be used in three different ways [37].

First, it can be included as a covariate in a regression model developed to estimate the average treatment effect (*regression adjustment*), as illustrated in (3),

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 \hat{Z}_i \qquad (3)$$

where Y_i is the i^{th} individual outcome, and the coefficients β_0, β_1, and β_2 represent, respectively, the adjusted mean outcome in the control group, the adjusted mean differential outcome, and the change in mean outcome associated with a marginal change in the propensity score.

Alternatively, the propensity score can be used to stratify patients into subgroups (usually defined by the quintiles of the propensity score distribution) within which to estimate average treatment effect (*stratification* or *subclassification on the propensity score*),

$$Y_{iq} = \beta_0 + \beta_1 T_i + \sum_{q=2}^{5} \beta_{2q} Q_{iq} + \sum_{q=2}^{5} \beta_{3q} T_i \cdot Q_{iq} \qquad (4)$$

where Q_{iq} is a dummy variable taking value 1 if the i^{th} subject belongs to quintile q, and 0 otherwise. The coefficients β_0 and β_1 in (4) can be interpreted as the mean outcome in the control and the differential outcome in the reference group (usually the 1st quintile) respectively. Similarly, β_{2q} and β_{3q} are the mean outcome in the control group and the treatment-by-quintile interaction term in the quintiles 2 to 5, essentially indicating the outcomes over and above those reported by the reference group. The differential outcome in the q^{th} quintile is then $(\beta_1 + \beta_{3q})$.

Finally, $\hat{Z}_i$ can be used to create propensity score matched pairs (in case of 1-to-1 matching) of treated and untreated subjects with similar propensity scores (*propensity score matching*). The latter approach is similar to traditional matching, and it is usually employed when the number of controls is larger than the numbers of treated patients. The objective here is to select, for each treated individual, a

'match' from the control group in order to create a quasi-experimental comparison. There are various algorithms that can be employed to carry out matching on the propensity score, and reviews of these have been published elsewhere.[59] In this paper we apply nearest neighbour 1-to-1 matching within a caliper of 0.25 standard deviations of the propensity score [60].

2.2 Cost-effectiveness analysis of individual patient-level RCT data

Let us denote C_{ij} and E_{ij} the cost and effect of individual i $(=1,....,n_j)$ in treatment arm j (where *j*=1 control and *j*=2 intervention). Using the notation of Nixon and Thompson [27], a general formulation of the cost-effectiveness model for individual patient-level data can be written as follows

$$\begin{aligned} C_{ij} &\sim Dist(\phi_{Cij}, \sigma_{Cj}) \\ E_{ij} &\sim Dist(\phi_{Eij}, \sigma_{Ej}) \end{aligned} \qquad (5)$$

which assume that costs and effects follow a certain probability distribution (*Dist*) with parameters ϕ and σ representing a measure of location and spread, respectively. Since costs and effects at an individual level are expected to be jointly distributed, model (5) needs to be parameterised in such a way as to reflect the correlation between these two outcomes.

In the simplest scenario, costs and effects can be assumed to follow a bivariate normal distribution, in which case ϕ's and σ's represent, respectively, the marginal means and standard deviations. The correlation between costs and effects in each treatment group can be preserved by modelling their mean as follows [27],

$$\begin{aligned} \phi_{C_{ij}} &= \mu_{C_j} + \beta_j \cdot (E_{ij} - \phi_{E_{ij}}) \\ \phi_{E_{ij}} &= \mu_{E_j} \end{aligned} \qquad (6)$$

which essentially assumes that, in each trial arm, the mean cost is linearly related to the departure of the individual level health outcome (E_{ij}) from its mean.

Since costs and effects at individual patient-level are often characterised by non-Normal distributions, the factorisation approach used in (5) and (6) is particularly helpful in that it allows selection of a wider range of possible distributions outside bivariate Normality. An alternative formulation of (5) to accommodate the typical

right skewed nature of cost data [21, 26] could employ a Gamma distribution for instance. Similarly, in the case of binary health outcomes (e.g. survival or cure status during the period of interest) such an event could be modelled using a Bernoulli process. This situation is illustrated in (7)

$$\begin{aligned} C_{ij} &\sim \Gamma(\eta_{Cj}, \rho_{Cij}) \\ E_{ij} &\sim Bernoulli(p_{ij}) \end{aligned} \tag{7}$$

where the Gamma is specified in terms of its shape (η_j) and rate (ρ_{ij}), a formulation that allows expression of the rate parameter as the ratio between the shape and the mean of the Gamma.

As in (6) the correlation between costs and effects at patient level is captured using a factorisation approach, conditioning the *logit* of the binary outcome on costs as illustrated in (8)

$$\begin{aligned} \phi_{Cij} &= \mu_{Cj} \\ \rho_{Cij} &= \eta_{Cj} / \phi_{Cij} \\ logit(p_{ij}) &= \mu_{Ej} + \beta_j \cdot (C_{ij} - \phi_{Cij}) \\ p_j &= (1 + e^{-\mu_{Ej}})^{-1} \end{aligned} \tag{8}$$

where p_j is estimated using the anti-logit transformation.

Regardless of the model used, the parameters of interest in CEA are the differential mean costs ($\Delta C = \mu_{C_2} - \mu_{C_1}$) and effects ($\Delta E = \mu_{E_2} - \mu_{E_1}$). Once estimated these quantities need to be related to each other giving rise to one of the following scenarios:

a) the intervention is both more effective (that is, generates larger health benefits), or at least as effective as, the control and is less costly;

b) The control is less effective than the intervention and more (or at least as) costly.

c) The control is both less (more) effective and less (more) costly compared with the intervention.

In CEA if the results indicate either scenarios (a) or (b) above, then one approach is clearly most cost effective, that is, it *dominates* the other. However, if the results indicate scenario (c), then a 'decision rule' is required to assess which is the most cost effective treatment. The decision rule typically requires the calculation of the incremental cost-effectiveness ratio (ICER), defined as $(\Delta C / \Delta E)$. The ICER represents the additional cost that the decision maker is (on average) expected to pay to achieve an additional unit of health benefit in this population. A treatment strategy is considered to be cost-effective if the decision maker's willingness to pay for an additional unit of health outcome (i.e. λ) is at least equal (or greater) to the ICER. The mean estimates of ΔC and ΔE together with their joint distribution can also be reported onto a cost-effectiveness plane,[61] or combined to obtain the incremental net benefit (INB) statistics [11] (a reformulation of the ICER), as follows

$$INB(\lambda) = \lambda \cdot \Delta E - \Delta C \tag{9}$$

which states that, at a given level of λ, an intervention can be considered cost-effective if $INB(\lambda) > 0$.

2.3 Integrating the propensity score in the cost-effectiveness analysis framework

The first way in which we can use the propensity score in cost-effectiveness analysis of observational data is through simple regression adjustment by including the individual-level predicted propensity score in the equations defining the mean costs and effects. In the case of the bivariate normal model these equations can be re-written as follows,

$$\begin{aligned} \phi_{C_{ij}} &= \mu_{C_j} + \beta_j \cdot (E_{ij} - \phi_{E_{ij}}) + \gamma_{Cj} \cdot (\hat{Z}_{ij} - \phi_{\hat{Z}_{ij}}) \\ \phi_{E_{ij}} &= \mu_{E_j} + \gamma_{Ej} \cdot (\hat{Z}_{ij} - \phi_{\hat{Z}_{ij}}) \end{aligned} \tag{10}$$

where $\hat{Z}_{ij}$ is the propensity score for individual *i* receiving treatment strategy *j*, and $\phi_{\hat{Z}_{ij}}$ is the mean of the distribution of the propensity score in treatment strategy *j*. By extension, equation (8) can be re-expressed as follows

$$
\begin{aligned}
&\phi_{Cij} = \mu_{Cj} + \gamma_{Cj}(\hat{Z}_{ij} - \phi_{Zij}) \\
&\rho_{Cij} = \eta_{Cj} / \phi_{Cij} \\
&logit(p_{ij}) = \phi_{Ej} + \beta_j \cdot (C_{ij} - \phi_{Cij}) + \gamma_{Ej}(\hat{Z}_{ij} - \phi_{Zij}) \\
&p_j = (1 + e^{-\mu_{Ej}})^{-1}
\end{aligned}
\tag{11}
$$

An alternative analytical strategy is to carry out the CEA within each propensity score quintile, either through the regression framework illustrated in section 2.2 or equivalently by running five separate analyses (one for each quintile), and subsequently combine the outputs to obtain an overall measure of cost-effectiveness. In the latter case, (6) and (8) can be re-expressed as,

$$
\begin{aligned}
&\phi_{C_{ij}} = \mu_{C_j} + \beta_j \cdot (E_{ij} - \phi_{E_{ij}}) + \gamma_{Cq} \cdot Q_{ijq} \\
&\phi_{E_{ij}} = \mu_{E_j} + \gamma_{Eq} \cdot Q_{ijq}
\end{aligned}
\tag{12}
$$

and

$$
\begin{aligned}
&\phi_{Cij} = \mu_{Cj} + \gamma_{Cjq} Q_{ijq} \\
&\rho_{Cij} = \eta_{Cj} / \phi_{Cij} \\
&logit(p_{ij}) = \phi_{Ej} + \beta_j \cdot (C_{ij} - \phi_{Cij}) + \gamma_{Ejq} Q_{ijq} \\
&\mu_{Ej} = (1 + e^{-\phi_{Ej}})^{-1}
\end{aligned}
\tag{13}
$$

The final, and perhaps most straightforward way to use propensity score methods in cost-effectiveness modelling, is by applying the models described in section 2.2 to the propensity score matched cohort.

3 Motivating example

The methods presented in section 2 were applied to the analysis of administrative data from the Ontario Myocardial Infarction Database (OMID). The OMID was created by researchers at the Institute for Clinical Evaluative Sciences in Ontario (Canada), with support from the Medical Research Council of Canada, to study population-based quality and patterns of care, readmissions, drug use and short- and long-term mortality for Ontario citizens who had an acute myocardial infarction between fiscal 1992/1993 and 2006/2007. The OMID links all of Ontario's major healthcare administrative data bases.[62] More specifically, patients' demographics

and clinical information, hospital-based services (procedures and diagnoses), in-hospital outcomes, length of hospital stay and healthcare resources intensity weights, were obtained from the Canadian Institute for Health Information (CIHI) data base. Physicians' fees relating to consults or assessments in private offices, acute care, and long-term care facilities; technical and professional components of diagnostic and therapeutic procedures; surgical procedures; and laboratory services were derived from the Ontario Health Insurance Plan (OHIP) data base. Drug costs for all adults aged 65+ in Ontario were extracted from the data base of the Ontario Drug Benefit (ODB) program. Finally, demographics and survival status were extracted from the Ontario Registered Persons Database (RPDB), developed, and maintained by the Ontario Ministry of Health and Long Term Care. Costs data were converted into comparable figures using the resource intensity weights provided by the CIHI and, where relevant, up-rated to 2005 figures using the Canadian price index for healthcare services and technologies [63].

The cohort used in this paper consists of patients who had either PTCA or CABG within 365 days of an index admission for AMI. The dates of the hospitalizations for AMI were between 1st April 1994 and 31st March 2004. We excluded patients who had a first PTCA or CABG prior to the index hospitalization or following the observation period, cases who had both PTCA and CABG, and observations with multiple of repeated procedures.

Propensity score matching was carried out in `STATA 9.0` [64] using the user written ado file `psmatch2.ado` [65], whereas the cost-effectiveness models described in section 2.3 were developed and estimated in the freely available software `WinBUGS 1.4.2` [66]. Finally, the user written collection of ado files `wb` were used to call WinBUGS from within `STATA` and to examine the resulting Markov chain Monte Carlo (McMC) simulations. The latter were obtained running 3 parallel chains for 10,000 iterations following a burn-in of 5,000 iterations. Convergence of each individual chain was assessed using the Gelman-Rubin convergence criteria [67], as implemented in `WinBUGS`. The `STATA` syntax codes used for the propensity score matching procedure and for creating the propensity score matched cohort, as well as the WinBUGS implementation of the cost-effectiveness analysis of the propensity score matched cohort data are reported in the appendix.

4 Results

4.1 Baseline characteristics of the cohort

Administrative data used in this motivating example included the following variables that were potential confounders of the treatment effect: age, gender, cardiogenic shock, acute and chronic renal failure, diabetes with complications, congestive heart failure, cerebrovascular disease, malignancies, pulmonary oedema,

cardiac dysrhythmia, the Charlson co-morbidity index, and household median income. With the exception of household median income, and the Charlson co-morbidity index, the remaining variables comprise the Ontario AMI mortality prediction model, which uses administrative data to predict mortality within 30 days from admission for an AMI [68]. Table 1 reports the baseline characteristics of the study cohort by treatment group. Continuous and dichotomous variables were compared between treatment groups using t-tests (or Wilcoxon rank sum tests) and chi-squared tests, respectively.

<<Table 1 here>>

Most of the baseline covariates of the two groups display a statistically significant imbalance. Compared to the PTCA group, individuals undergoing CABG were approximately 3 years older ($p < 0.001$), tended to be male ($p < 0.001$), and would be less likely to present with cardiogenic shock ($p = 0.001$), but more likely to have some form of diabetes related complications ($p < 0.001$), congestive heart failure ($p<0.001$), and cerebrovascular disease ($p<0.001$). Furthermore, individuals in the CABG group also presented with a higher frequency of pulmonary oedema ($p<0.001$), cardiac dysrhythmia ($p=0.01$), and co-morbidities ($p<0.001$).

Figure 1 shows the distribution of the predicted propensity score between CABG and PTCA group in the study cohort. While there is good overlap between the distributions of the propensity score in the two treatment groups, it can be seen that for values of the propensity score higher than 0.5 the number of individuals who underwent CABG in the cohort is larger than those who underwent PTCA. An opposite trend can be seen for values of the propensity score lower than 0.4.

<<Figure 1 here>>

4.2 Application of the propensity score methodology to the study cohort

Table 2 reports the standardised differences between treatment groups for the unmatched cohort, the propensity score matched cohort, and within each quintile of the propensity score. Standardized differences that exceed 10% are frequently taken to denote meaningful imbalance in a baseline variable between treatment arms [35, 69, 70]

<<Table 2 here>>

While the unmatched initial cohort shows some serious imbalance in at least four of the covariates of interest (i.e. age, gender, congestive heart failure, Charlson score),

the matched cohort displays good balancing of the same covariates between the two treatment strategies (see Figure 2 for visual inspection).

<<Figure 2 here>>

Table 3 reports the summary statistics of the propensity score matched cohort.

<<Table 3 here>>

Returning to the information reported in Table 2, one can observe that the within propensity score quintile distribution of these baseline covariates is also well balanced, although some borderline values close to ±10% difference are observed for one or two of the variables in the first (e.g. diabetes with complications) and fifth (e.g. age) quintile. This is also reflected in Figure 3, which shows the box-plots by treatment group and quintile of the propensity score.

<<Figure 3 here>>

4.3 Integrating propensity score and cost-effectiveness methodology for the analysis of the study cohort data

Costs and survival status at one year post-procedure in the two groups were analysed using the methods presented in section 2.3. Table 4 reports the mean differential costs and odds ratio, together with their 95% credibility intervals (CrI). For comparative purposes the results of the unadjusted analysis are also reported.

All four analytical strategies lead to the same conclusion (i.e. PTCA *dominates* CABG, in other words PTCA costs less and produces a higher probability of survival at one year post procedure compared to CABG). The estimated differential costs and odds ratios obtained using the propensity score through either regression adjustment, matching or sub-classification are considerably different than those obtained from the unadjusted analysis. More specifically, the differential costs obtained from the analyses using the propensity score are approximately 26% lower than those estimated in the unadjusted analysis, while their estimated odds ratios are closer to one than that obtained from the unadjusted analysis. Furthermore, two of the analytical strategies using propensity score – regression adjustment and matching – suggest a non-statistically significant survival advantage of PTCA versus CABG in our cohort.

The differential costs estimates obtained using propensity score through either regression adjustment or matching are very similar in terms of point estimates, with the propensity score matching analysis giving wider credibility intervals. Interestingly, the same two methods give slightly different odds ratio estimates.

<<Table 4 here>>

5 Discussion

Healthcare economic evaluation of individual patient-level data has been typically carried out using information collected as part of randomised clinical trials. There are, however, a number of cases in which it is unfeasible to adopt a trial design and recourse to the analysis of non-randomised experimental data becomes the only way forward to address a particular research question. Issues of bias in this type of data have been traditionally dealt with using multivariate regression adjustment, multivariate matching, or stratification, each of which suffers from some form of limitation (e.g. unfeasibility or impracticability). The use of propensity score offers a potential solution for the analysis of observational data, by conditioning the probability of treatment allocation on a set of baseline covariates. In this sense, propensity score methodology conveys all the information contained by the set of covariates included in the prediction model into a single variable, which can then be used (i) as a covariate in a regression model, (ii) to created matched pairs of individuals with the same propensity score (and hence the same distribution of covariates), and (iii) to create strata of equal size (usually defined by the quintiles of the distribution of the propensity score) within which the estimation of the average treatment effect can be carried out.

This paper offered a tutorial-like discussion of ways in which cost-effectiveness analysis and propensity score methodology to analyse observational healthcare data can be integrated. Using administrative data from the Ontario Myocardial Infarction Database (OMID), and linking these with data from the Canadian Institute for Health Information (CIHI), the Ontario Health Insurance Plan (OHIP), the Ontario Drug Benefit (ODB) program, and Ontario Registered Persons Database (RPDB), we assessed the cost-effectiveness of CABG *versus* PTCA – one year post procedure - in a cohort of individuals who received the intervention within 365 days of their index admission for AMI.

It was found that, regardless of which propensity score methodology was used to adjust for the risk of confounding, at one year from procedure PTCA was both cheaper and associated with better survival rate than CABG. However, a note of caution is in order here. A very short time horizon was selected for the analysis (i.e.

one year) and it could be argued that for reimbursement decisions one should be evaluating these interventions over the entire lifetime of patients [71]. For instance, PTCA is typically associated with a higher re-intervention rate during the first year post procedure; however, for simplicity we omitted the sub-group of patients who had a repeat procedure from our analysis. Similarly, if one were to assess the long-term cost-effectiveness of CABG *versus* PTCA it would be paramount to control for the changing nature of PTCA. The introduction of new devices (e.g. stents) and the use of drugs (e.g. glycoprotein IIB/IIIA inhibitors) alongside PTCA are all variables that should be accounted for.

On a related note, in this paper we have used routinely collected administrative data to illustrate the use of propensity score methods in cost-effectiveness analysis. It is important to acknowledge that administrative data are not typically collected for research purposes. Austin *et al* [34] compared accuracy of treatment effect estimates obtained using administrative *versus* clinical data and found that measures of treatment effect obtained using administrative data were larger than those obtained from clinical data. Furthermore, propensity scores developed using administrative data did not necessarily balance patients characteristics contained in the clinical data, since the latter are usually characterised by richer information.

The reader should be reminded at this point that propensity score methodology has no pretence to be able to control for unobserved confounders. As argued by Rubin [72], the propensity score is not in the same class as any of the 'selection models' [73], which instead attempt to model the probability of treatment assignment either directly through the observed outcome, or indirectly through instrumental variables [74] methods. Head to head comparisons of the various methods have been extensively carried out and are outside the scope of this paper, but what emerged from these studies – unsurprisingly - is that the appropriate method in any given circumstance depends on a combination of the data available and the parameter of interest [75].

Acknowledgments

AM is recipients of a Wellcome Trust post-doctoral Training Fellowship in Health Services Research (GR071304MA). The authors are grateful to Dr Jack Tu for the financial support which funded the programming time used to create the study cohort. The funding was from the Canadian Institutes of Health Research Team Grant in Cardiovascular Outcomes Research. We would like to thank Julie Wang and Jun Guan for their programming work relating to data extraction and cohort creation, Anirban Basu and Simon Walker for comments on a previous draft of this paper.

Errors or omissions, and the opinions expressed in this paper are those of the authors and do not represent the point of view of the funding institutions. This paper was conceived and the applications carried out while AM was visiting researcher at the Institute for Clinical Evaluative Sciences, Toronto, Ontario (Canada) between April and October 2007.

Appendices

A.1. STATA code for the propensity score matching and creation of the study propensity score matched cohort. This code requires STATA8.2 or higher. You will need to install the STATA ado file psmatch2.ado

```
************************************
* ESTIMATION OF THE PROPENSITY SCORE *
************************************
logit <treatment group> <varlist>, robust
predict double varname

**************************************************
* PLOTTING THE PROPENSITY SCORE BY TREATMENT GROUP *
**************************************************
psgraph, treated(varname) pscore(varname) bin(50)

************************************************************
* NEIREST NEIGHBOUR CALIPER MATCHING WITHOUT REPLACEMENT *
************************************************************
*random ordering the observations
set seed 123456
gen u=uniform()
sort u

*****************************
* PROPENSITY SCORE MATCHING *
*****************************
*carry out propensity score matching
psmatch2 varname, pscore(varname) outcome(varname) caliper(<caliper>) logit
noreplacement descending
*testing that balancing has been achieved
pstest <varlist>
*plotting the propensity score by treatment group
psgraph, treated(varname) pscore(varname) bin(50)

*********************************
* CREATING THE PS MATCHED SAMPLE *
*********************************
sort _id
g match=_id[_n1]
g costmatch=cost[_n1]
g survmatch=surv[_n1]
g treatnew=_id if _nn==1
g costtreat=cost if _nn==1
g survtreat=surv if _nn==1
drop if treat==.

keep costmatch survmatch survtreat costtreat

stack costmatch survmatch costtreat survtreat , into(cost surv)  clear

rename _stack arm

g group=arm
label define arm 1 PTCA 2 CABG
label values arm arm
```

A.2. WinBUGS code for the analysis of propensity score matched cost-effectiveness data

```
# Model costs as a Gamma and survival status at one year as a Bernoulli.
# Adapted from Nixon and Thompson (Health Economics 2005, 14(12):1217-29

model  {

for(i in 1:N)  {

cost[i]~dgamma(shape.c[arm[i]],rate.c[arm[i],i])
     rate.c[arm[i],i]<-shape.c[arm[i]]/phi.c[arm[i],i]
     phi.c[arm[i],i]<-mu.c[arm[i]]

surv[i]~dbern(pi.alive[arm[i],i])
    logit(pi.alive[arm[i],i])<-mu.e[arm[i]]+beta.e[arm[i]]*(cost[arm[i],i]-mu.c[arm[i]])
                  }
# node transformations
for (j in 1:2) {p.e[j]<-exp(mu.e[j])          }

# prior distributions
for (j in 1:2) {

shape.c[j]~dunif(shape.c.low[j],shape.c.up[j])
mu.c[j]~dunif(mu.c.low[j],mu.c.up[j])
mu.e[j]~dunif(mu.e.low[j],mu.e.up[j])
beta.e[j]~dunif(beta.e.low[j],beta.e.up[j])

             }

# ce[1]=c1, ce[2]=c2, ce[3]=e1, ce[4]=e2
ce[1]<-mu.c[1]
ce[2]<-mu.c[2]
ce[3]<-mu.e[1]
ce[4]<-mu.e[2]

dc<-ce[2]-ce[1]       #differential cost
de<-ce[4]-ce[3]       #differential log-odds ratio
dp<-p.e[2]-p.e[1]   #differential probability of survival

}
```

A.3. Calling WinBUGS from within STATA to run the model in A.2 and import the output for analysis. You will need to run the package winbugsfromstata.pkg from the web at: http://www2.le.ac.uk/departments/health-sciences/extranet/BGE/genetic-epidemiology/gedownload/winbugsfromstata/

```
************************************
* write the data into WinBUGS format *
************************************
quietly wbarray arm cost surv, format(%9.0g) ///
     saving(filename)

************************
* run the WinBUGS script *
************************
wbrun, script(filename) ///
        w("c:/Program Files/WinBUGS14/WinBUGS14")

************************************
* read WinBUGS coda output into STATA *
************************************
wbcoda,  root(filename) clear

************************************
* calculate the descriptive statistics *
************************************
wbstats dc de


*********************************************************
* Plot the Kernel density of the parameters of interest *
*********************************************************
wbdensity dc
wbdensity de

*****************************************************
* Plot the McMC traces of the parameters of interest *
*****************************************************
wbtrace de dc , gopt(scheme(s2mono)) cgopt(row(2) scheme(s2mono))

*******************************************************
* Autocorrelation plots for the parameters of interest *
*******************************************************
gen num=_n
tsset order
ac dc , lag(200)
ac de , lag(200)
```

Table 1 Baseline characteristics of the study cohort

Variable	PTCA (N=24,088)	CABG (N=19,835)	*p*-values
Age	60±12	63±10	<0.001
Female	28%	23%	<0.001
Cardiogenic shock	1%	0.7%	0.001
Acute renal failure	0.6%	0.7%	0.639
Chronic renal failure	1.5%	1.6%	0.394
Diabetes with complications	1.4%	2.5%	<0.001
Congestive heart failure	7.8%	14%	<0.001
Cerebrovascular disease	1%	2%	<0.001
Malignancies	0.8%	0.9%	0.204
Pulmonary Oedema	0.4%	0.9%	<0.001
Cardiac dysrhythmia	9.7%	10%	0.010
Charlson score	0.3±0.7	0.5±0.8	<0.001
Household median income (CAN$)	21096±4265	20894±3972	<0.001

Values are Mean ± SD for continuous variables and proportion for the binary variables

Table 2 Standardised differences

Variable	Unadjusted	Matched cohort	*Quintile*				
			1st	2nd	3rd	4th	5th
Age	26.7	0.3	2.7	1.3	3.7	-0.3	-9.8
Female	-11.3	1.0	3.2	2.0	2.4	-1.0	-1.9
Cardiogenic shock	-3.3	-0.2	-1.7	0.9	2.0	-1.6	-1.8
Acute renal failure	0.4	0.7	5.9	3.4	-2.5	1.6	-2.7
Chronic renal failure	0.8	0.5	6.6	5.1	-1.2	-0.3	-4.5
Diabetes with complications	7.3	0.8	8.1	7.0	1.6	0.0	-3.2
Congestive heart failure	20.4	1.1	4.4	-0.7	0.2	2.8	1.7
Cerebrovascular disease	7.0	0.6	-2.0	1.4	2.9	-3.3	2.0
Malignancies	1.2	0.5	1.6	-3.6	0.0	0.7	2.6
Pulmonary Oedema	6.0	0.8	1.3	-0.7	2.2	-5.5	4.3
Cardiac dysrhythmia	2.5	1.9	3.3	3.9	1.6	4.0	-2.7
Charlson score	29.5	1.7	6.3	3.7	3.2	3.0	-1.1
Household median income (CAN$)	-4.9	0.7	-0.5	1.4	-2.8	2.4	3.5

Note: Values are standardised % differences. For continuous variables these are obtained as $\frac{100 \cdot (\bar{x}_{treatment} - \bar{x}_{control})}{\frac{\sqrt{(s^2_{treatment} + s^2_{control})}}{2}}$ and for proportions are obtained as $\frac{100 \cdot (p_{treatment} - p_{control})}{\sqrt{p_T(1-p_T) + p_C(1-p_C)} / 2}$

Table 3 Comparison of treated and untreated in matched sample

Variable	PTCA (N=15,943)	CABG (N=15,943)
Age	62±12	62±10
Female	24%	24%
Cardiogenic shock	0.7%	0.7%
Acute renal failure	0.6%	0.6%
Chronic renal failure	1.6%	1.6%
Diabetes with complications	1.7%	1.8%
Congestive heart failure	9.3%	9.7%
Cerebrovascular disease	1.3%	1.4%
Malignancies	0.8%	0.9%
Pulmonary Oedema	0.5%	0.5%
Cardiac dysrhythmia	9.4%	9.9%
Charlson score	0.4±0.7	0.4±0.8
Household median income (CAN$)	20,924±4,265	20,953±3,972

Values are Mean ± SD for continuous variables and proportion for the binary variables

Table 4 Cost-effectiveness results (unmatched, matched, regression adjusted, sub-classification)

	Cost difference * (CAN$)	Odds Ratio* (Survival)	
Unadjusted analysis	2243 (2072 – 2413)	.743 (.678 - .814)	PTCA dominates CABG
Regression adjusted	1679 (1505 – 1852)	.920 (.837 - 1.011)	PTCA dominates CABG
Propensity Score Matching	1667 (1143 – 2175)	.838 (.597 - 1.144)	PTCA dominates CABG
Sub-classification	1693 (1521 – 1864)	.847 (.765 - .937)	PTCA dominates CABG

* Difference in mean cost (CABG-PTCA) and 95% credibility intervals (CrI).

** Odds ratio of survival at after one year following CABG vs PTCA.
Values less than 1 indicate survival advantage in favour of PTCA;
values greater than 1 indicate survival advantage in favour of CABG.

Figure 1: Distribution of the propensity score in the CABG and PTCA patients (initial cohort)

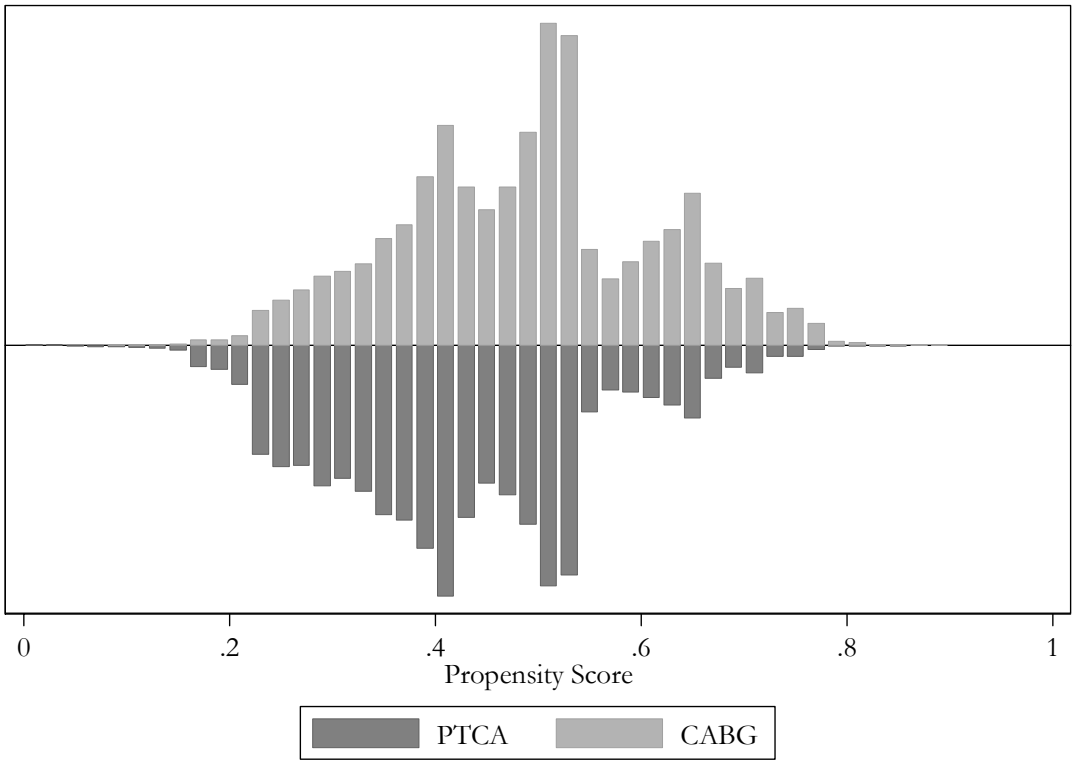
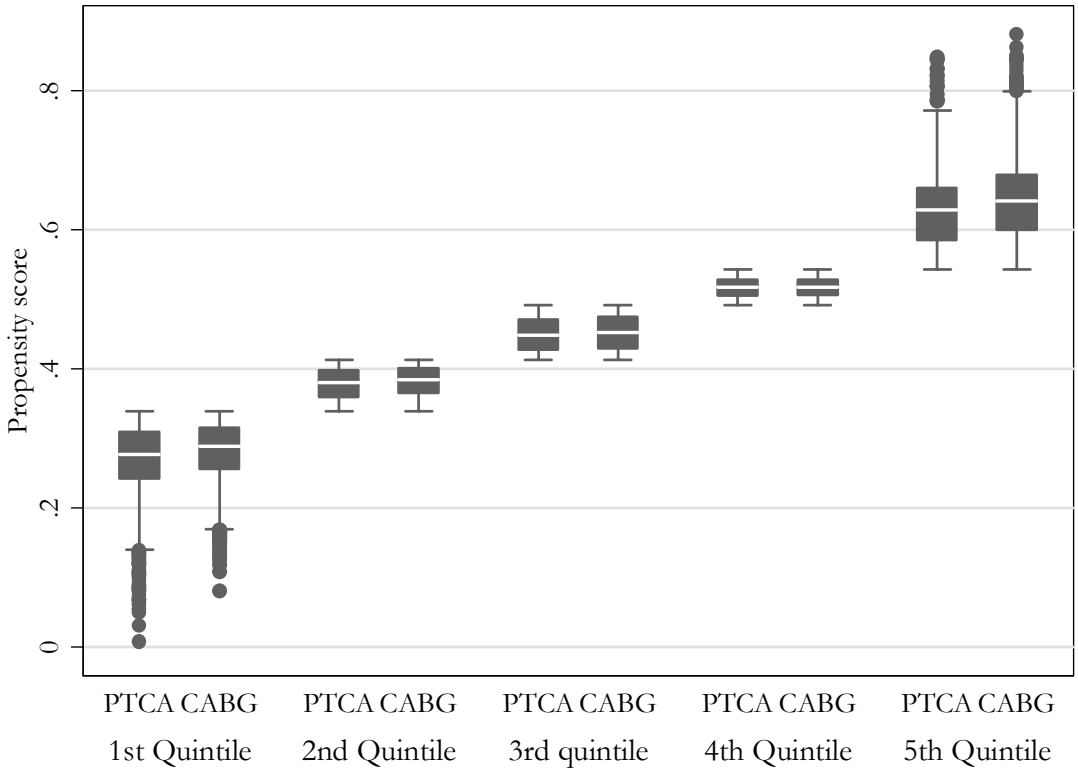
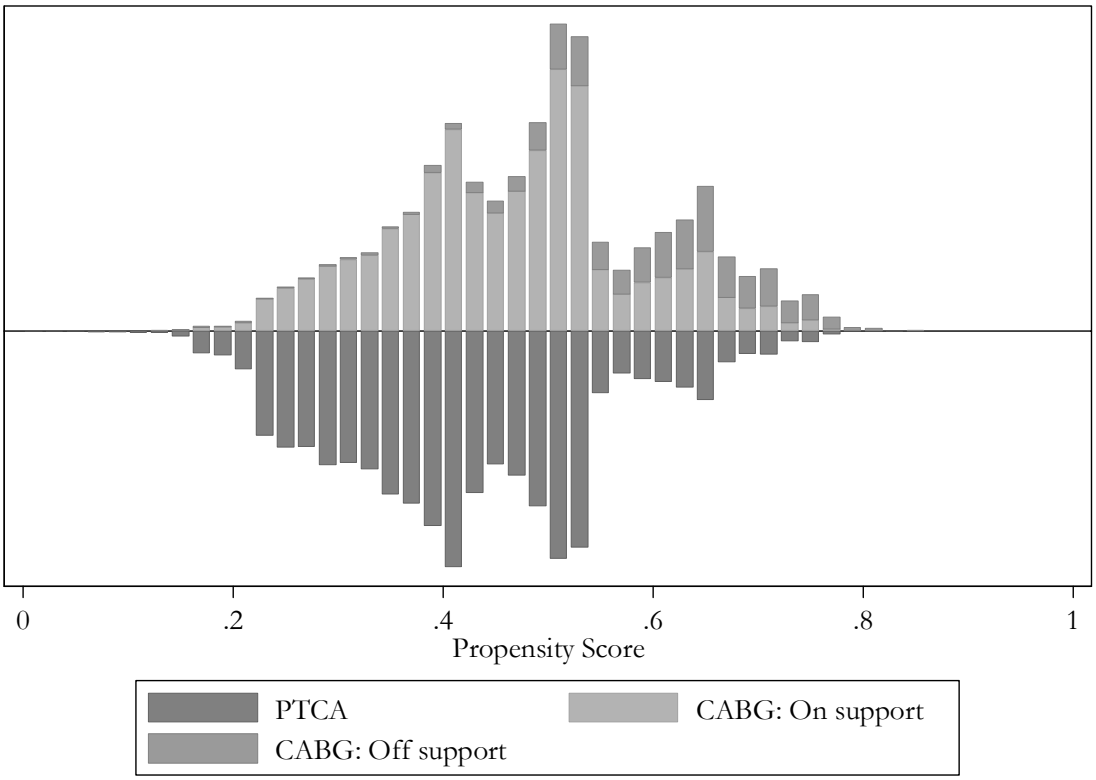


Figure 2: Comparison of propensity score between CABG and PTCA group by quintile of the propensity score



Note: The shaded grey vertical box at the centre of each box represents the middle 50 per cent of the distribution of the data within each quintile. The lower and upper ends of the box represent the 25th and 75th percentile of the distribution, respectively. The solid horizontal lines through each shaded box denote the median of the distribution. The vertical lines (i.e. 'whiskers') extend out to 1.5 x the interquartile range. The dots beyond the whiskers identify extreme observations.

Figure 3: Distribution of the propensity score in the CABG and PTCA group after 1-to-1 matching based on individuals' propensity score



References

1. Briggs AH. Handling uncertainty in economic evaluation and presenting the results. In: Drummond MF, McGuire A, editors. Economic evaluation in health care. Merging theory with practice. Oxford: Oxford University Press; 2001.
2. Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. Statistics in Medicine 1999;18(23):3245-62.
3. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps; a non-parametric approach to confidence interval estimation. Health Economics 1997;6:327-340.
4. Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fiellers' theorem. Health Economics 1996;5:297-305.
5. Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and c/e-ratios alongside a clinical trial. Health Economics 1994;3:309-319.
6. O'Brien BJ, Drummond MF, Labelle RJ, Willan AR. In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in heath care. Medical Care 1994;32:150-163.
7. Briggs AH, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Economics 1998;7:723-740.
8. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. Report: Health Technology Assessment; 1999 February 1999.
9. Tambour M, Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results [comment]. Health Economics 1998;7(2):143-7.
10. Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. International Journal of Health Technology Assessment 1998;14:467-471.
11. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 1998;18:S68-S80.
12. Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost-benefit analysis. International Journal of Technology Assessment in Health Care 1991;7:12-21.

13. Fenwick E, Byford S. A guide to cost-effectiveness acceptability curves. Br J Psychiatry 2005;187:106-8.

14. Fenwick E, Claxton K, Sculpher MJ. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 2001;10:779-89.

15. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Economics 2004;13(5):405-15.

16. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Economics 2002;11(5):415-430.

17. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Economics 2004;13:461-475.

18. Willan AR, Lin DY, Manca A. Regression methods for cost-effectiveness analysis with censored data. Statistics in Medicine 2005;24(1):131-45.

19. Thompson SG, Nixon R, Grieve R. Addressing the issues that arise in multicentre cost data, with application to a multinational study. Journal of Health Economics 2006;25(6):1015-1028.

20. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? British Medical Journal 2000:1197-1200.

21. Thompson SG, Nixon RM. How sensitive are cost-effectiveness analyses to choice of parametric distributions? Medical Decision Making 2005;25(4):416-423.

22. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trials data. In; 2000. p. 27.

23. O'Hagan A, Stevens JW. Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. Stat Methods Med Res 2002;11(6):469-90.

24. O'Hagan A, Stevens JW. On estimators of medical costs with censored data. Journal of Health Economics 2004;23(3):615-625.

25. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost-effectiveness analysis from clinical trials data. Statistics in Medicine 2001;20:733-753.

26. Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. Statistics in Medicine 2004;23:1311-31.

27. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness analysis. Health Econ 2005;14(12):1217-29.

28. Manca A, Lambert PC, Sculpher MJ, Rice N. Cost effectiveness analysis using data from multinational trials: The use of bivariate hierarchical modelling. Medical Decision Making 2007;27(4):471-490.

29. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing Generalisability by Location in Trial-Based Cost-Effectiveness Analysis: the Use of Multilevel Models. Health Economics 2005;14(5):471-85.

30. Cooper NJ, Sutton AJ, Mugford M, Abrams KR. Use of Bayesian Markov Chain Monte Carlo methods to model cost-of-illness data. Medical Decision Making 2003;23(1):38-53.

31. Winship C, Morgan SL. The estimation of casual effects from observational data. Annual Review of Sociology 1999;25:659-707.

32. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Economics 1996;5:513-524.

33. Claxton K, Sculpher MJ, Drummond MF. A rational framework for decision making by the National Institute for Clinical Excellence. Lancet 2002;360:711-715.

34. Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: Administrative versus clinical data. Statistics in Medicine 2005;24(10):1563-1578.

35. Austin PC, Mamdani MM. A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. Statistics in Medicine 2006;25(12):2084-2106.

36. Coyte PC, Young W, Croxford R. Costs and outcomes associated with alternative discharge strategies following joint replacement surgery: Analysis of an observational study using a propensity score. Journal of Health Economics 2000;19(6):907-929.

37. D'Agostino Jr. R. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1999;17(19):2265 - 2281.

38. Ichimura H, Taber C. Propensity-score matching with instrumental variables. American Economic Review 2001;91(2):119-124.

39. Indurkhya A, Mitra N, Schrag D. Using propensity scores to estimate the cost-effectiveness of medical therapies. Statistics in Medicine 2006;25(9):1561-1576.

40. Johnson A, Berg G, Fleegler E, Lehn J. Clinical and utilization outcomes for a heart failure care support program: A matched-cohort study. Disease Management and Health Outcomes 2005;13(5):327-335.

41. Kravitz R L, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Quarterly 2004;82(4):661-687.

42. Landrum MB, Ayanian JZ. Causal effect of ambulatory specialty care on mortality following myocardial infarction: A comparison of propensity score and instrumental variable analyses. Health Services and Outcomes Research Methodology 2001;2(3-4):221-245.

43. Merito M, Pezzotti P. Comparing costs and effectiveness of different starting points for highly active antiretroviral therapy in HIV-positive patients: Evidence from the ICONA cohort. European Journal of Health Economics 2006;7(1):30-36.

44. Mitra N, Indurkhya A. A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data. Health Economics 2005;14(8):805-815.

45. Reed SD, Dillingham PW, Briggs AH, Veenstra DL, Sullivan SD. A Bayesian approach to aid in formulary decision making: Incorporating institution-specific cost-effectiveness data with clinical trial results. Medical Decision Making 2003;23(3):252-264.

46. Roux AVD. The study of group-level factors in epidemiology: Rethinking variables, study designs, and analytical approaches. Epidemiologic Reviews 2004;26:104-111.

47. Sarkar S, Watts S, Ohashi Y, Carroll RJ, Uesaka H, Mason TM, et al. Bridging data between two ethnic populations. A new application of matched case-control methodology. Drug Information Journal 2002;36(2):349-356.

48. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: A systematic review. Journal of Clinical Epidemiology 2005;58(6):550-559.

49. Zhou X-H, Li C, Gao S, Tierney WM. Methods for testing equality of means of health care costs in a paired design study. Statistics in Medicine 2001;20(11):1703-1720.

50. Austin PC. A critical appraisal of propensity score matching in the medical literature from 1996 to 2003. Statistics in Medicine 2008;in press.

51. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for

improvement. Journal of Thoracic and Cardiovascular Surgery 2007;134:1128-1135.

52. Stürmer T, M. J, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yelded increasing use, advantages in specific savings, but not substantially different estimates compared with conventional multivariable methods. Journal of Clinical Epidemiology 2006;59:437-447.
53. Weiss JP, Saynina O, McDonald KM, McClellan MB, Hlatky MA. Effectiveness and cost-effectiveness of implantable cardioverter defibrillators in the treatment of ventricular arrhythmias among medicare beneficiaries. Am J Med 2002;112(7):519-27.
54. Polsky D, Mandelblatt JS, Weeks JC, Venditti L, Hwang YT, Glick HA, et al. Economic evaluation of breast cancer treatment: considering the value of patient choice. J Clin Oncol 2003;21(6):1139-46.
55. Hornberger J, Cosler LE, Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. Am J Manag Care 2005;11(5):313-24.
56. Coleman CI, McKay RG, Boden WE, Mather JF, White CM. Effectiveness and cost-effectiveness of facilitated percutaneous coronary intervention compared with primary percutaneous coronary intervention in patients with ST-segment elevation myocardial infarction transferred from community hospitals. Clin Ther 2006;28(7):1054-62.
57. Dhainaut JF, Payet S, Vallet B, Franca LR, Annane D, Bollaert PE, et al. Cost-effectiveness of activated protein C in real-life clinical practice. Crit Care 2007;11(5):R99 [Epub ahead of print].
58. Rosenbaum PR, Rubin DB. The central role of propensity score in observational studies for causal effect. Biometrika 1983;70:41-55.
59. Guo S, Barth RP, Gibbons C. Propensity score matching strategies for evaluating substance abuse services for child welfare clients. Children and Youth Services Review 2006;28:357-383.
60. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician 1985;39:33-38.
61. Black WC. The CE plane: a graphic representation of cost-effectiveness. Medical Decision Making 1999;10(3):212-4.
62. Williams JI, Young W. A Summary of Studies on the Quality of Health Care Administrative Databases in Canada. In: Goel V, Williams JI, Anderson GM, Blackstien-Hirsch P, Fooks C, Naylor CD, editors. Patterns

of Health Care in Ontario, The ICES Practice Atlas. 2nd edition ed. Ottawa: Canadian Medical Association; 1996. p. 339-345.

63. Statistics Canada. Consumer Price Index, health and personal care, by province. In: Statistics Canada; 2007.
64. StataCorp. Stata Statistical Software: Release 9.0. In. College Station, TX: StataCorp LP; 2005.
65. Leuven E, Sianesi B. psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. In. 2.0.5 (12 April 2004) ed; 2003.
66. Lunn DJ, Thomas T, Best N, Spiegelhalter DJ. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 2000;10(4):325-37.
67. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 1998;7:434-455.
68. Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. Journal of the Americal College of Cardiology 2001;37:992-997.
69. Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. Journal of Clinical Epidemiology 2001;54(4):387-98.
70. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med 2007;26(4):734-53.
71. Sculpher MJ, Claxton K, Drummond MF, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Economics 2006;15(7):677-87.
72. Rubin DB. Using propensity scores to help design observational studies: an application to the tobacco litigation. Health Services & Outcomes Research Methodology 2001;2:169-188.
73. Heckman JJ, Vytlacil E. Structural equations, treatment effects, and econometric policy. Econometrica 2005;73(3):669-738.
74. Newhouse JP, McClellan MB. Econometrics in outcomes research: the use of intrumental variables. Annual Review of Public Health 1998;19:17-34.

75. Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fiscal Studies 2000;21(4):427-468.