The International Journal of Biostatistics

Volume 6, *Issue* 2 2010 *Article* 16

CAUSAL INFERENCE

Cutting Feedback in Bayesian Regression Adjustment for the Propensity Score

Lawrence C. McCandless, *Simon Fraser University*
Ian J. Douglas, *London School of Hygiene & Tropical Medicine*
Stephen J. Evans, *London School of Hygiene & Tropical Medicine*
Liam Smeeth, *London School of Hygiene & Tropical Medicine*

Recommended Citation:
McCandless, Lawrence C.; Douglas, Ian J.; Evans, Stephen J.; and Smeeth, Liam (2010) "Cutting Feedback in Bayesian Regression Adjustment for the Propensity Score," *The International Journal of Biostatistics*: Vol. 6: Iss. 2, Article 16.
DOI: 10.2202/1557-4679.1205

Cutting Feedback in Bayesian Regression Adjustment for the Propensity Score

Lawrence C. McCandless, Ian J. Douglas, Stephen J. Evans, and Liam Smeeth

Abstract

McCandless, Gustafson and Austin (2009) describe a Bayesian approach to regression adjustment for the propensity score to reduce confounding. A unique property of the method is that the treatment and outcome models are combined via Bayes theorem. However, this estimation procedure can be problematic if the outcome model is misspecified. We observe feedback that can bias propensity score estimates. Building on new innovation in Bayesian computation, we propose a technique for cutting feedback in a Bayesian propensity analysis. We use the posterior distribution of the propensity scores as an input in the regression model for the outcome. The method is approximately Bayesian in the sense that it does not use the full likelihood for estimation. Nonetheless, it severs feedback between the treatment and outcome giving propensity score estimates that are free from bias but modeled with uncertainty. We illustrate the method in a matched cohort study investigating the effect of statins on primary stroke prevention.

KEYWORDS: confounding, bias, observational studies, Markov chain Monte Carlo

Author Notes: Funding for McCandless was obtained from a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank EPIC for supplying the data and all the general practices that participate in the THIN database.

1. Introduction

Propensity score (PS) techniques are a class of statistical methods that can be used to reduce confounding in observational studies. They have seen widespread application in recent years, particularly in epidemiologic investigations of the effects of treatments or exposures. The methods involve first estimating the PS for each study subject, defined as the probability of treatment given measured confounders. Then the estimated PS is used to build comparisons of treated and untreated subjects that are balanced with respect to the distribution of confounders. Analysis techniques include stratification on the PS, matching on the PS, inverse probability weighting and other techniques (Lunceford and Davidian, 2004).

PS techniques have seen only modest development from a Bayesian perspective. Estimators that use the PS are typically calculated from estimating equations or likelihood functions. There has been little research on the role of Bayes theorem when using the PS for causal inference. One reason is because there is no consensus on how the PS should be incorporated into a Bayesian analysis. Some scientists argue that when constructing a model for the outcome variable in an observational study, the likelihood function should not depend on the PS (Robins and Ritov, 1997; Tan, 2006). Others argue the PS may play a practical role in building robust Bayesian procedures that have good frequentist properties (Rubin, 1985).

McCandless, Gustafson and Austin (2009) describe a Bayesian approach to regression adjustment for the PS. Their methodology mimics the popular technique of stratifying on quintiles of the estimated PS. Markov chain Monte Carlo (MCMC) is used to sample from the posterior distribution of model parameters. However, a curious finding is that the Bayesian approach allows the models for outcome variable and treatment assignment be estimated simultaneously rather than sequentially. Standard PS techniques proceed in two stages. We first estimate the PS and then substitute it in place of the true PS in an outcome regression model. In contrast, a Bayesian propensity analysis calculates the posterior distribution for the PS and the treatment effect at the same time.

A consequence of the Bayesian approach is that the outcome variable can produce feedback that influences the estimated PS. The method makes a trade-off between fitting the outcome model and the treatment assignment model. Provided that all models are been correctly specified, then this can greatly increase the efficiency of the estimated PS compared to standard frequentist

techniques (McCandless et al. 2009). However, if the regression model for the relationship between the outcome and the PS is misspecified, then the Bayesian approach results in contamination between models that biases the PS estimates. See McCandless et al. (2009) for detailed simulation results comparing the performance of different methods.

In a recent paper, Lunn et al. (2009a) introduce techniques for *cutting feedback* when fitting complex Bayesian models. They describe an example from pharmacokinetics that involves the joint analysis of multiple datasets where there is uncertainty in the model specification. To limit feedback between models, the authors propose an approximate Bayesian technique that uses the posterior distribution from one fitted model as an input when fitting the remaining models. This restricts the flow of information between models during MCMC computation. A similar Bayesian estimation procedure called *modularization* is proposed by Liu, Bayarri and Berger (2009) in the context of analyzing computer models.

The idea of cutting feedback is natural in regression adjustment for the PS because may be desirable to estimate the outcome and treatment models separately. PS techniques allows us to generate unconfounded comparisons between treatment groups. An outcome model is ultimately required for estimating the treatment effect. However, it is typically viewed as a nuisance model and handled with minimal parametric assumptions (e.g. in conditional logistic regression when matching on the PS). Consequently, it may be inappropriate to allow the outcome model to influence the estimated PS. In fact, Rubin (2008) argues that PS estimation should occur without any reference to outcome data. From a purist Bayesian perspective, the best solution is to use a more flexible nonparametric model for the mean response. But a simpler approach is to restrict the flow of information between model components altogether.

In this article we introduce a technique for cutting feedback in Bayesian regression adjustment for the PS. We consider observational studies with a time-fixed dichotomous treatment and several potential confounders. We motivate the methodology using a data example from pharmacoepidemiology. We describe a matched cohort study of the relationship between statin therapy and risk of stroke using data from the United Kingdom Health Improvement Network (THIN) database (Smeeth et al., 2008). To limit feedback in a Bayesian propensity analysis, we use the posterior distribution for the PS as a covariate input in the regression model for the outcome variable. The method is only approximately Bayesian because it does not involve joint estimation of the

DOI: 10.2202/1557-4679.1205

treatment and outcome models. However, we illustrate that the resulting PS estimates are free of feedback from the outcome model and they are compatible with estimates produced in standard frequentist propensity analysis.

2. THIN Data for Studying the Health Effects of Statin Among Elderly UK Patients

Statins have emerged as the most widely prescribed cholesterol lowering medication (Rutishauser, 2006). Their effectiveness in primary prevention of cardiovascular disease has been well documented in several large randomized trials dating back to the 1990's. However there has been renewed interest in the health effects of statins that are unrelated to cardiovascular disease, such as risk of infection, cancer, dementia and other chronic illnesses. In a recent paper, Smeeth et al. (2008) used PS techniques to study the impact statins on a variety of health outcomes in a matched cohort study of UK patients. They used data from the Health Improvement Network database (THIN) database, which contains computerized medical records from general practices in the UK.

In this article we replicate the analysis of Smeeth et al. (2008), but focus on using PS techniques to estimate the effect of statins on risk of stroke in the elderly. Previous research has already demonstrated that statins reduce the risk of stroke. However, we can use this knowledge in the present investigation in order to validate competing PS analyses.

Our target population is UK patients aged 65 or older and registered at one of 303 general practices that contributed data to the THIN database during the period January 1995 and December 2006. Following Smeeth et al. (2008), the treated group of statin users was defined as all patients initiating a statin after 1995 and who had one year of continuous registration at a general practice. We identified a total of 23306 such individuals, of whom 19274 had complete information on important potential confounders such as smoking, alcohol consumption and body mass index. For each treated patient we set an *index date* for initiation of follow-up, which was the date of first receiving a statin.

To limit the size of the untreated group, we followed Smeeth et al. (2008) and used a matched cohort design. Each treated subject was matched with up to five untreated subjects that were randomly sampled from the THIN database. Matching variables included age and gender, and additionally, index date and general practice in order to limit confounding from prescribing

practice and temporal trends. A total of 71050 untreated subjects were selecting giving a total sample size of $n = 19274 + 71050 = 90324$, with an median of 4 untreated subjects in each matched set. The median follow-up time for all patients was 5.3 years.

Let X be an indicator variable for statin treatment for a patient in the study, coded 1 if the subject initiated a statin and 0 otherwise. We conducted a time-to-event analysis of time to stroke following the index date, with censoring occurring at the end of follow-up. Subjects were excluded if they had stroke prior to the index date. Let T denote the time of observation and let δ denote a censoring indicator. If $\delta = 0$ then the subject was censored and T is the time of censoring. If $\delta = 1$ then the subject had a stroke and T is the time of stroke. Following convention in survival analysis, we assume that the censoring mechanism is uninformative (Ibrahim, Chen and Sinha, 2004).

A total of 3713 strokes occurred over the course of follow-up. Of those, 814 occurred in the treated group during 86943 persons-years of follow-up, whereas 2899 strokes occurred in the untreated group during 382217 person-years of follow-up. The crude relative risk of stroke for treated versus untreated is $\frac{814/86943}{2899/382217} = 1.23$ with 95% confidence interval (1.14, 1.33), suggesting that statins are dangerous and increase stroke risk.

In fact, the reverse is true (Rutishauser, 2006) and the association between statins and stroke is likely to be confounded. Table 1 describes the characteristics of the treatment groups upon entry into the study. While the age and gender distributions are roughly balanced between groups owing to the matched sampling, we see that that the statin users are a much sicker group of patients. They have higher rates of cardiovascular related illness such as diabetes and hypertension. Consequently, the statin user group was at greater risk of stroke prior to initiating treatment. While the matching process reduces some of the systematic difference between treatment groups, it is clear that valid comparison cannot be drawn without further adjustment for confounding. See Smeeth et al. (2008) for further discussion of the cohort.

3 Frequentist versus Bayesian Propensity Analysis of the THIN Data

3.1 Estimation of the Propensity Scores

The THIN data contain rich information on potential confounders. We let C denote at 40×1 vector of potential confounders, including the 23 variables

DOI: 10.2202/1557-4679.1205

Table 1: Characteristics of statin users versus non-users on the index date. Each column gives totals with percentages in brackets.

	Statin User n=19274	Non-user n=71050
Demographics		
Age		
$65-74$	12296 (64)	47646 (67)
$75-84$	5776 (30)	19935 (28)
>85	1202 (6)	3469 (5)
Female Sex	10997 (57)	40202 (57)
BMI <25	6632 (34)	29709 (42)
BMI 25-30	8800 (46)	28601 (40)
BMI >30	3842 (20)	12740 (18)
Low SES	3734 (19)	13428 (19)
Current smoker	6023 (31)	21863 (31)
Heavy drinker	278 (1)	573 (1)
Comorbid conditions		
Diabetes	4867 (25)	8953 (13)
Coronary heart disease	10569 (55)	9227 (13)
Athereosclerosis	12046 (62)	13088 (18)
Hepatic disease	38 (0)	256 (0)
Renal disease	286 (1)	564 (1)
Hyperlipidemia	6409 (33)	2467 (3)
Hypertension	10772 (56)	27713 (39)
Medications		
Hormone replacement therapy	513 (3)	2366 (3)
Antidepressent	1750 (9)	5746 (8)
Lipid lowering agent	1084 (6)	713 (1)
Aspirin	11296 (59)	12650 (18)
Beta blocker	9906 (51)	17250 (24)
Calcium channel blocker	8994 (47)	14514 (20)
Antihypertensive	7869 (41)	13149 (19)

listed in Table 1, plus an additional 17 unlisted variables that are measures of illness, medication use and access to health services.

Smeeth at al. (2008) used regression adjustment for the PS to reduce confounding. In the present analysis, we also use PS techniques. We estimate the PS for each patient using a logistic regression model for the association between X and C, given by

$$\text{logit}\{Pr(X = 1|C)\} \quad = \quad \gamma^T C. \tag{1}$$

The regression coefficients γ model the association between the covariates and probability of treatment. We write $C = (C_0, C_1, \ldots, C_p)$, with $p = 40$ and fix $C_0 = 1$ as a regression intercept term. Table 2, under the heading MLE, gives maximum likelihood estimates for selected components of the regression parameter γ computed from the THIN data.

In our analysis, the covariate vector C includes age and gender (see Tables 1 and 2), however it does not include the matching variables for patient index date and primary practice that are described in Section 2. The reason is because these covariates are high dimensional with no ordering. Handling these covariates in a propensity model requires conditional logistic regression or other methods to handle multiple nuisance parameters. But conditional likelihood models have no straightforward interpretation within the Bayesian analysis framework (Rice, 2004). By ignoring these matching variables in a propensity analysis, it is possible that we may induce residual confounding. However, because of the matched sampling on exposure, we expect that most of the association between the matching variables and treatment assignment will be modest. Further details on this point are given in the Discussion of Section 5.

Following Rubin and Thomas (1996), we call $Z = \text{logit}\{Pr(X = 1|C)\}$ the propensity score, which we define as the log odds of treatment given measured covariates. If $\hat{\gamma}$ is an estimate for γ, then a patient with covariate vector C will have an estimated PS equal to $\hat{\gamma}^T C$. Figure 1 illustrates the balance of selected covariates with respect to treatment status, as a function of the estimated PS. The figure gives loess kernel density estimates of the quantities $P(Cj = 1|X = 1, Z = z)$ (solid curve) and $P(Cj = 1|X = 0, Z = z)$ (dashed curve), for selected covariate C_j, as a function of z. If the curves lie on top of one another, then this indicates good balance. For example, the bottom right figure shows the prevalence of aspirin use as a function of the estimated PS. In Table 1, we see that statin users are far more likely to take aspirin than non-user (59% versus 18%). However, Figure 1 shows that conditioning

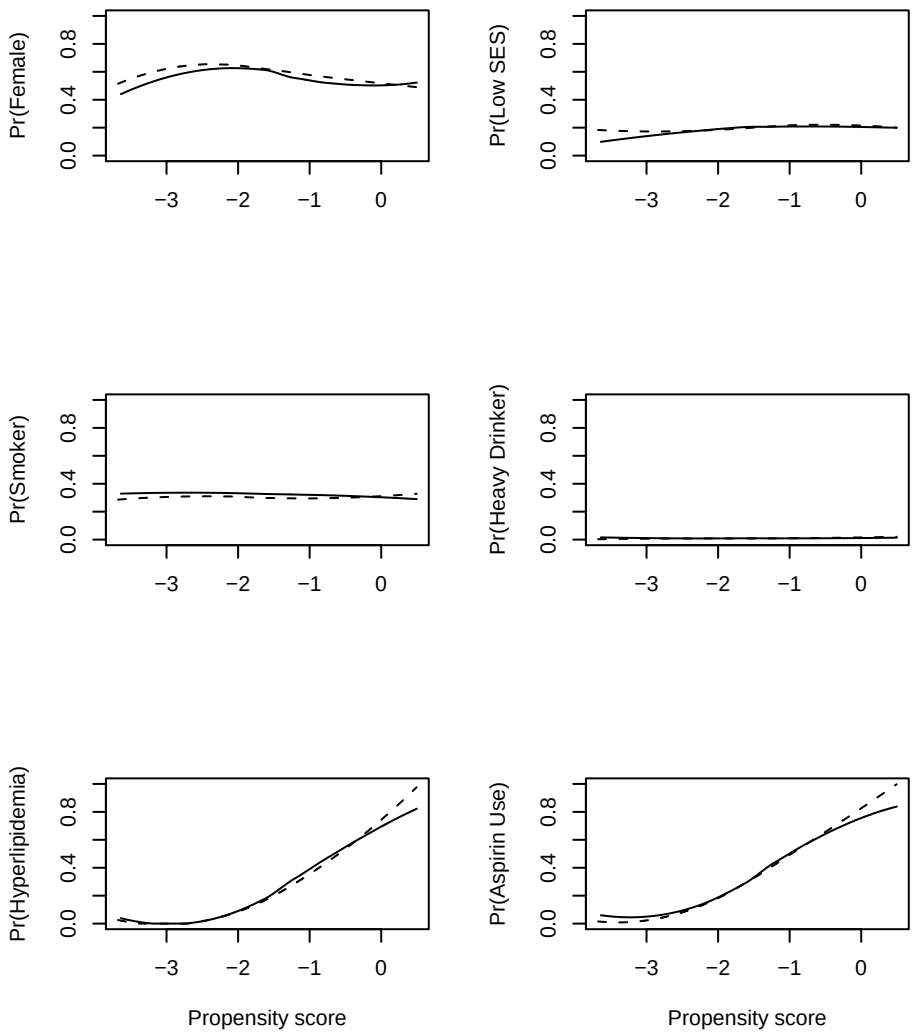
DOI: 10.2202/1557-4679.1205

Table 2: Log odds ratios (standard errors) for regression coefficients γ in the treatment assignment model of equation (1).

	Log Odds Ratio (Standard Error)		
Covariate	MLE	BAYES	SEQUENTIAL BAYES
Demographics			
Age			
$65-74^*$	0	0	0
$75-84$	-0.156 (0.024)	-0.132 (0.024)	-0.157 (0.024)
>85	-0.094 (0.044)	-0.066 (0.044)	-0.095 (0.044)
Female Sex	0.458 (0.024)	0.440 (0.023)	0.458 (0.024)
BMI $<25^*$	0	0	0
BMI 25-30	0.182 (0.023)	0.183 (0.023)	0.182 (0.024)
BMI >30	0.008 (0.030)	0.006 (0.030)	0.007 (0.030)
Low SES	-0.108 (0.026)	-0.101 (0.026)	-0.109 (0.026)
Current smoker	0.147 (0.027)	0.155 (0.026)	0.148 (0.026)
Heavy drinker	0.799 (0.097)	0.801 (0.095)	0.796 (0.098)
Comorbid conditions			
Diabetes	0.753 (0.026)	0.761 (0.025)	0.754 (0.026)
Coronary heart disease	0.880 (0.055)	0.881 (0.053)	0.880 (0.055)
Athereosclerosis	0.553 (0.056)	0.545 (0.053)	0.554 (0.055)
Hepatic disease	-0.883 (0.208)	-0.877 (0.208)	-0.899 (0.210)
Renal disease	0.204 (0.092)	0.202 (0.092)	0.201 (0.091)
Hyperlipidemia	2.523 (0.031)	2.515 (0.031)	2.525 (0.031)
Hypertension	-0.169 (0.024)	-0.159 (0.024)	-0.169 (0.024)
Medications			
Hormone therapy	-0.185 (0.062)	-0.191 (0.063)	-0.184 (0.063)
Antidepressent	-0.070 (0.036)	-0.060 (0.036)	-0.071 (0.037)
Lipid lowering agent	0.138 (0.064)	0.138 (0.064)	0.139 (0.064)
Aspirin	1.090 (0.024)	1.087 (0.024)	1.090 (0.024)
Beta blocker	0.214 (0.024)	0.216 (0.024)	0.214 (0.024)
Calcium channel blocker	0.127 (0.024)	0.128 (0.024)	0.127 (0.024)
Antihypertensive	0.494 (0.025)	0.485 (0.025)	0.494 (0.024)

* Reference category

Figure 1: Balance with respect to treatment status. The plots give kernel density estimates of the prevalence of selected covariates, among treated (solid curve) or untreated (dashed curve), as a function of the estimated PS.



DOI: 10.2202/1557-4679.1205

on Z forces the distribution of aspirin use to be roughly balanced between treatment groups because the solid line and curved line are close together. This implies that if we condition on the estimated PS, then aspirin use is no longer a powerful confounder. In the THIN data, the estimated PS ranges from -5.5 to 4.4. To create Figure 1, we truncated the scores at the 10^{th} and 90^{th} percentiles in order to ensure that there were adequate patients within each treatment group so as to give precise estimates of the distribution of C_j.

Figure 1 indicates that we can use the estimated PS to reduce confounding in the THIN data. Possible analytic strategies include stratifying on quintiles of the PS, matching on the PS, including the PS as a covariate in a regression model, or inverse probability weighting. See Luncefore and Davidian (2004) for a review of analytic techniques.

3.2 Regression Adjustment for the Propensity Score

To adjust for confounding from C, we include Z as a covariate in a Weibull proportional hazards regression model for stroke risk. The hazard function is given by

$$h(T|X,C) = \exp\{\alpha + \beta X + \xi Z\}\lambda T^{\lambda-1} \tag{2}$$

with corresponding survivor function

$$S(T|X,C) = \exp\big[-\exp\{\alpha + \beta X + \xi Z\}T^{\lambda-1}\}\big]$$

(Ibrahim et al., 2004). Equation (2) models the time to stroke as a Weibull distribution with scale parameter $\exp\{\alpha + \beta X + \xi Z\}$ and shape parameter λ. The linear predictor $\alpha + \beta X + \xi Z$ defines the relationship between stroke risk and both treatment and covariates. It includes an intercept α, treatment effect parameter β, and a linear contribution ξZ, which governs the relationship between the Z and risk of stroke. Note that from equation (1) that the quantity Z is a deterministic function of the covariates C (McCandless et al., 2009).

In typical applications, the relationship between the PS and the outcome variable is non-linear and poorly understood. Nonetheless, provided that the outcome model in equation (2) is correct, then the regression parameter β will have a causal interpretation. See Rosenbaum and Rubin (1983) and Lunceford and Davidian (2004) for further discussion of regression adjustment for the PS. A linear relationship between the log hazard and Z may seem overly simplistic in the THIN data. However, it has the advantage that it straightforward to

Table 3: Log odds ratios (standard errors) for the treatment effect β and the association between the propensity score and outcome ξ given in equation (2).

	Log Hazard Ratio (Standard Error)	
Analysis method	Treatment effect β	Propensity slope ξ
MLE	-0.127 (0.049)	0.126 (0.011)
BAYES	-0.149 (0.051)	0.135 (0.012)
SEQUENTIAL BAYES	-0.137 (0.051)	0.131 (0.011)
UADJUSTED	0.209 (0.039)	*

* The unadjusted analysis does not include ξ in the model.

understand and highlights the distinction between Bayesian and frequentist PS techniques. In the discussion that follows, we show that the regression coefficient ξ acts as a throttle that controls feedback between the outcome and PS during estimation. In principle, it is straightforward to modify equation (2) to accommodate more complicated forms of nonparametric dependence, such as regression splines.

Table 3, under row heading "MLE" gives point estimates and standard errors for the treatment effect β and slope parameter ξ when fitting equation (2) by maximum likelihood, while substituting the estimated PS in place of the true PS. For comparison, the row with heading "UNADJUSTED" gives the results from fitting the Weibull proportional hazards model while forcing $\xi = 0$ (i.e. ignoring confounding). In the unadjusted analysis, the hazard ratio is $\exp(0.209) = 1.23$, which suggests that statins increase risk of stroke. The MLE analysis gives a hazard ratio of $\exp(-0.127) = 0.90$ and correctly indicates that statins reduce the risk of stroke (Rutishauser, 2006).

An interesting feature of Table 3 is the MLE estimate of ξ, which is 0.126. Patients with a high propensity score are at greater risk of stroke. This makes sense intuitively. As illustrated in Table 1, statin users are at greater risk of cardiovascular disease upon study entry. Hence ξ is greater than zero. The relationship between the the outcome and propensity score normally receives little attention in data analysis (Kurth et al., 2006) because it is a nuisance parameter with no biological meaning. Nonetheless, the parameter ξ plays an important role in Bayesian analysis because it dictates how the outcome influences the estimation of the PS.

DOI: 10.2202/1557-4679.1205

3.3 A Bayesian Approach

McCandless et al. (2009) recently proposed a Bayesian approach to regression adjustment for the propensity score. The method uses standard models (e.g. equations (1) and (2)). However, it's uniqueness stems from the property that the models for treatment and outcome are fit simultaneously rather than sequentially. To illustrate, write $data = \{(T_i, \delta_i, X_i, C_i); i \in 1 : n\}$ to denote the THIN dataset. The posterior density for model parameters $(\alpha, \beta, \xi, \lambda, \gamma)$ is

$$p(\alpha, \beta, \xi, \lambda, \gamma | data) \propto L(\alpha, \beta, \xi, \lambda, \gamma) p(\alpha, \beta, \xi, \lambda, \gamma), \tag{3}$$

with prior density $p(\alpha, \beta, \xi, \lambda, \gamma)$ and likelihood function

$$\begin{aligned} L(\alpha, \beta, \xi, \lambda, \gamma) &= \prod_{i=1}^{n} p(T_i | X_i, C_i, \delta_i, \alpha, \beta, \xi, \lambda, \gamma) p(X_i | C_i, \gamma) \\ &= \prod_{i=1}^{n} h(T_i | X_i, C_i)^{\delta_i} S(T_i | X_i, C_i) \times p(X_i | C_i, \gamma) \\ &= \prod_{i=1}^{n} \Big[\exp\{\alpha + \beta X_i + \xi Z_i\} \lambda T_i^{\lambda - 1} \Big]^{\delta_i} \times \\ &\quad \exp\big[- \exp\{\alpha + \beta X_i + \xi Z_i\} T_i^{\lambda} \big] \frac{\exp\{X_i (\gamma^T C_i)\}}{1 + \exp\{\gamma^T C_i\}}, \end{aligned}$$

which is the product of the likelihood functions for the outcome and treatment assignment model.

To estimate the treatment effect, we sample from the posterior distribution for model parameters using MCMC. The calculation proceeds in two iterative stages: First, we impute the PS by updating from the conditional distribution of γ, denoted as $p(\gamma | \alpha, \beta, \xi, \lambda, data)$. Second, given the imputed PS, we fit the survival analysis model by updating from the conditional distribution of the outcome model parameters $p(\alpha, \beta, \xi, \lambda | \gamma, data)$. The procedure is conceptually similar to the Monte Carlo EM algorithm for maximum likelihood estimation.

Feedback from the outcome model is a consequence of the conditional distribution for γ, which is given by

$$p(\gamma | \alpha, \beta, \xi, \lambda, data) \propto \prod_{i=1}^{n} p(T_i | X_i, C_i, \delta_i, \alpha, \beta, \xi, \lambda, \gamma) p(X_i | C_i, \gamma) p(\gamma). \tag{4}$$

Stroke risk depends on the PS via the linear predictor ξZ. Therefore, Metropolis updates of γ will to some extent be influenced by the estimated value of

ξ. If ξ is large in magnitude, then this increases feedback. Whereas if $\xi \to 0$ then feedback vanishes, and equation (4) becomes

$$p(\gamma|\alpha, \beta, \xi, \lambda, data) \quad \propto \quad \prod_{i=1}^{n} p(X_i|C_i, \gamma)p(\gamma), \tag{5}$$

which is just the usual posterior distribution used to calculate PS estimates from Bayesian logistic regression of X on C. Full computational details are given in McCandless et al. (2009) for the case of a dichotomous response variable.

We apply the Bayesian propensity analysis to the THIN data. We assign independent diffuse Gaussian priors for the regression coefficients

$$\alpha, \beta, \xi, \gamma_0, \gamma_1, \ldots, \gamma_p \sim N(0, 10^3),$$

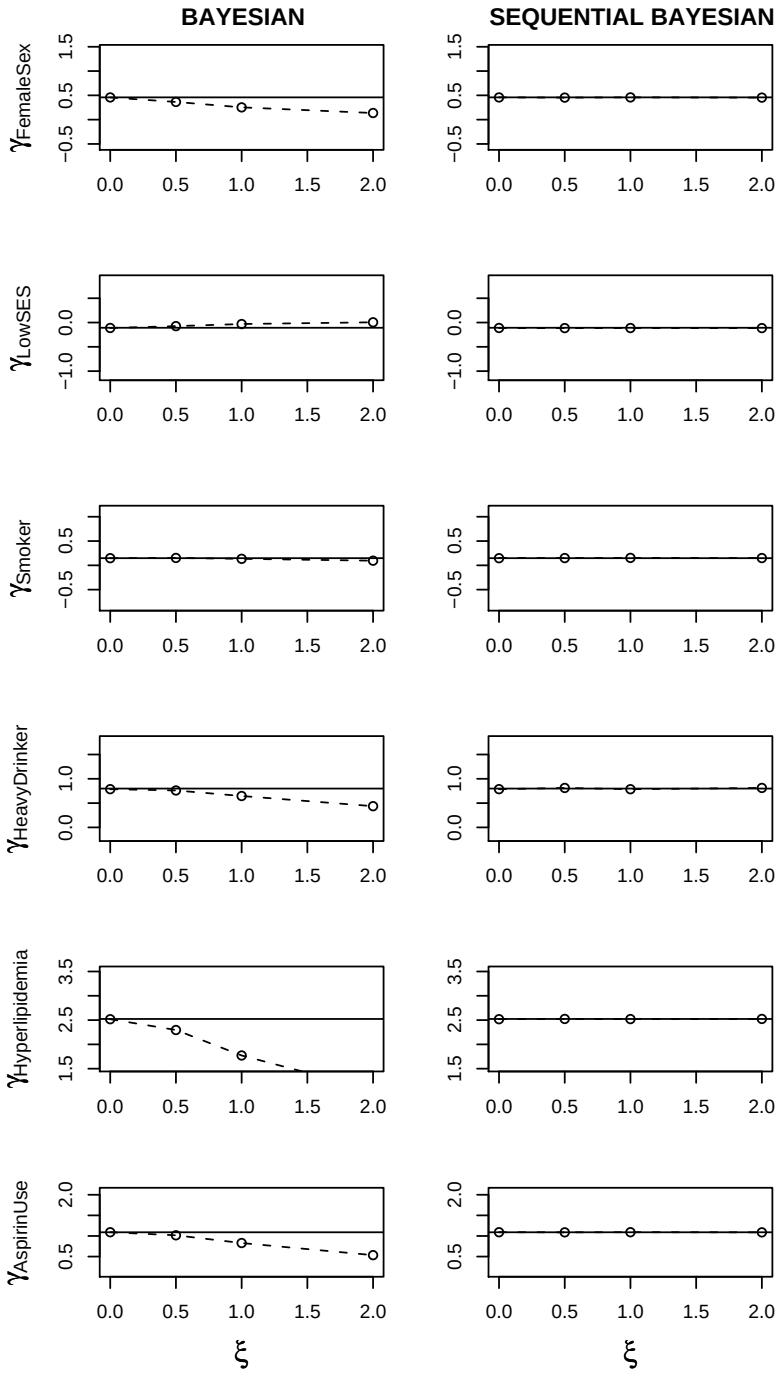
and we assign a diffuse Gamma prior for the shape parameter λ (Ibrahim et al, 2004). Using flat priors raises the possibility that the posterior distribution will be improper. In fact, we see from the likelihood function that if the components of γ are equal to zero, meaning that $Z_i = 0$, then the likelihood function does not depend on ξ and the model is nonidentifiable. Assigning proper priors to model parameters ensures that the posterior is proper, but does not necessarily prevent the sampler from escaping into regions of non-identifiability. Thus care is needed to check for satisfactory MCMC mixing.

We draw a sample from the posterior distribution using an MCMC chain of length 100 000 after burn-in. We assessed sampler convergence by repeating the analysis using different overdispersed starting values and the diagnostic tools in the R CODA package (Plummer et al., 2006). The analysis results are given in Tables 2 and 3 under the headings "BAYES". Table 2 reveals clear albeit small differences between MLE and BAYES estimates of γ. For example, the log odds ratios for the effect of age and gender on statin use are slightly attenuated towards zero in the Bayesian analysis.

In fact, we can easily exaggerate the feedback using sensitivity analysis. Rather than letting the data guide the fitted value for ξ, we can hold ξ fixed during MCMC. This is accomplished by not updating ξ. Figure 2 shows the results of such a sensitivity analysis. The plots on the left under the heading "BAYES" are derived from a Bayesian analysis with ξ held fixed at values ranging between 0.0 and 2.0 (horizontal axis) in a sensitivity analysis. The dashed curves in each plot describes the posterior mean of a component of γ as a function of ξ. For example, in the top left panel the dashed curve describes

DOI: 10.2202/1557-4679.1205

Figure 2: Sensitivity analysis results from applying BAYES or SEQUENTIAL BAYES with the parameter ξ held fixed during MCMC computation. The dashed curves plot the posterior mean of a component of γ as a function of ξ. The solid horizontal lines indicate the MLE of the same component of γ taken from Table 2.



$E[\gamma_{\text{FemaleSex}}|data, \xi]$ as a function of ξ. To caracterize the magnitude of feedback, we also plot the solid horizontal curve, which is the MLE for $\gamma_{\text{FemaleSex}}$ and is equal to 0.458 (Table 2). The extent to which the solid and dashed curves depart from one another indicates the extent that BAYES and MLE give different inferences when ξ is large in magnitude. Large values of ξ boost feedback, whereas when $\xi \to 0$ the feedback vanishes, and BAYES and MLE are equivalent.

4. Cutting Feedback in Bayesian Propensity Analysis

4.1 Review: Cutting Feedback in Bayesian Computation

In a recent paper, Lunn et al. (2009a) introduce a method for *cutting feedback* when fitting complex Bayesian models using MCMC. The approach is similar in spirit to two-stage estimation. We express the posterior distribution as a correct decomposition of conditional distributions for model parameters, but during posterior updating we do not update from the full conditionals. The procedure discounts the likelihood contribution from different data sources to ensure that the model components are fitted separately. Cutting feedback approach has been implemented in the BUGS language (Lunn et al., 2009b) through "cut" command. A related Bayesian computational procedure is proposed by Liu et al. (2009).

The work of Lunn et al. (2009a) is motivated by applications in pharmacokinetics, whereas Liu et al. (2009) describe examples in the analysis of computer models. Each case involves combining inferences from multiple data sources using different models. A full Bayesian approach fits the models simultaneously via Bayes theorem. But this can have undesirable consequences. The datasets may be incompatible with one another in the sense that separate analyses lead to different inferences about the same parameters. The combined fit for any specific model will be poor because of contamination between data sources. Liu et al. (2009) argue that improved modelling is the best solution, however this may not be straightforward in some settings if the data generating mechanism is complex and poorly understood.

Furthermore, simultaneous model fitting may be unappealing based on subject area considerations. For example, the analysis of computer models involves fitting a computer model emulator using simulation runs. Field ob-

DOI: 10.2202/1557-4679.1205

servations are often available, but should not be used to inform the emulator. Joint fitting of all data sources at once is counterintuitive to subject matter experts. Cutting feedback ensures that the flow of information during imputation moves in one direction. See Liu et al. (2009) for further discussion of applications to computer models.

Lunn et al. (2009a) are careful to point out that the resulting inferences does not arise from an underlying probability model. But nor do inferences from other sequential analyses. Further discussion on this point is given in Section 6.

4.2 Application to the THIN data

Following Lunn et al. (2009a) and Liu et al. (2009) we cut feedback between the treatment and outcome models by not updating from the full conditional distribution for γ. The correct conditional densities for $p(\gamma|\alpha, \beta, \xi, \lambda, data)$ is given in equation (4). To cut feedback, we update from the approximate conditional distribution

$$\tilde{p}(\gamma|\alpha, \beta, \xi, \lambda, data) \quad \propto \quad \prod_{i=1}^{n} p(X_i|C_i, \gamma)p(\gamma),$$

which ignores the likelihood contribution from the outcome. Note that this density is exactly equivalent to the posterior distribution from Bayesian logistic regression of X on C given in equation (5). The procedure is trivial to implement in Bayesian computation because it involves only a small modification of the MCMC algorithm. At iteration t, of a random walk Metropolis Hasting algorithm with proposal γ^*, we assign $\gamma^t \leftarrow \gamma^*$ with probability

$$\min\left[\frac{\tilde{p}(\gamma^*|\alpha, \beta, \xi, \lambda, data)}{\tilde{p}(\gamma^{t-1}|\alpha, \beta, \xi, \lambda, data)}, 1\right].$$

Tables 2 and 3 under heading "SEQUENTIAL BAYES" give estimates for γ, β and ξ when cutting feedback in the Bayesian propensity analysis. As in Section 3.3, the estimates are obtained using an MCMC chain of length 100 000 after burn-in, and we apply convergence diagnostics in the software R (Plummer et al., 2006). Table 2 reveals that the sequential Bayes estimates of γ are in close agreement with the MLE. Thus feedback from the outcome has been eliminated. In fact, the point estimates for γ from MLE and SEQUENTIAL BAYES must be asymptotically equivalent because of the well known large sample frequency matching properties of Bayesian and maximum

likelihood estimates. MLE and SEQUENTIAL BAYES are computed using the same likelihood function for γ.

Table 3 illustrates that the estimated treatment effect β is similar for all analyses, but BAYES and SEQUENTIAL BAYES give a small increase in the standard error of β. This increase in posterior uncertainty for the treatment effect reflects the propagation of the (modest) uncertainty in the PS through the analysis. MLE ignores this uncertainty because it substitutes the estimated PS in place of the true PS.

Figure 2 (right hand side) illustrates the effect of cutting feedback on estimation of γ as part of a sensitivity analysis that holds the parameter ξ fixed during MCMC computation. Dashed curves, which are nearly invisible, give the posterior mean of a selected component of γ as a function of ξ. For example, in the top right we plot $E[\gamma_{\text{Female Sex}}|data, \xi]$ as a function of ξ. The solid curves give the MLE of the relevant component of γ taken from Table 2. For SEQUENTIAL BAYES, we see that the dashed and solid curves are overlapping for all values of ξ. This illustrates that the magnitude of ξ (i.e. the association between the outcome and the PS) does not influence feedback on γ. Inferences for γ are driven entirely from the treatment assignment model.

Figure 2 reveals that by not updating from the full conditional distribution of γ we are able to block feedback from the outcome variables when estimating the PS. The resulting estimated PS are in close agreement with those obtained from the standard frequentist approach, but with the commensurate propagation of uncertainty in the PS through the analysis.

Table 4: The effective sample size for MCMC chain segments of length 10 000.

	Effective Sample Size	
	BAYES	SEQUENTIAL BAYES
$\gamma_{FemaleSex}$	283	656
γ_{LowSES}	239	681
γ_{Smoker}	253	661
$\gamma_{HeavyDrinker}$	303	669
$\gamma_{Hyperlipidemia}$	278	724
$\gamma_{AspirinUse}$	293	623

A useful byproduct of cutting feedback is improved MCMC computation. Fitting the combined likelihood can hinder sampler convergence because the chain must move so as to fit the treatment and outcome models simultane-

DOI: 10.2202/1557-4679.1205

ously. Table 4 reports the effective sample size (ESS) of the posterior samples for selected components of γ based on a chain segment of length 10 000. The ESS measures the amount of information in the chains, while taking into consideration the autocorrelations. The ESS equals the length of the chain if the autocorrelation is zero, and is calculated by estimating the spectral density from an autoregressive model (Plummer et al. 2006). In Table 4, the ESS are generally low ($\ll$ 10000), as is typical for the random walk Metropolis Hastings algorithm. However, in each case the ESS is increased by a factor of 2 for the sequential Bayesian analysis. Cutting feedback improves convergence of the sampler.

5. Simulation Study of the Effect of Cutting Feedback

We present brief simulation results to illustrate the consequences of cutting feedback between the PS and outcome variable. Building on McCandless et al. (2009), we consider the scenario of a dichotomous outcome Y, a dichotomous treatment X, and two continuous covariates (C_1, C_2) that are independent Gaussian distributed with mean zero and variance one. We simulate ensembles of 1000 synthetic datasets of sample size $n = 100$, using the data generating mechanism

$$\begin{aligned} \text{logit}\{Pr(Y=1|X,C_1,C_2)\} &= \beta X + \tilde{\xi}_1 C_1 + \tilde{\xi}_2 C_2 \\ \text{logit}\{Pr(X=1|C_1,C_2)\} &= \gamma_1 C_1 + \gamma_2 C_2, \end{aligned}$$

which have y-intercepts equal to zero. When generating the data, we fixed the parameters to be equal to $(\beta, \tilde{\xi}_1, \tilde{\xi}_2, \gamma_1, \gamma_2) = (0, 1, 1, .5, .5)$, which we call *Design A*, or $(\beta, \tilde{\xi}_1, \tilde{\xi}_2, \gamma_1, \gamma_2) = (0, .5, .5, 1, 1)$, which we call *Design B*. In either case, the treatment effect β is equal to zero, but the unadjusted association between X and Y is confounded because (C_1, C_2) are associated with X and Y.

We analyze the synthetic data using five different methods: MLE, BAYES, and SEQUENTIAL BAYES, each of which are fit using the regression models

$$\begin{aligned} \text{logit}\{Pr(Y=1|X,C_1,C_2)\} &= \alpha + \beta X + \xi Z \\ \text{logit}\{Pr(X=1|C_1,C_2)\} &= Z = \gamma_0 + \gamma_1 C_1 + \gamma_2 C_2, \end{aligned}$$

and additionally, method GOLD, which we define as fitting

$$\text{logit}\{Pr(Y=1|X,C)\} = \alpha + \beta X + \tilde{\xi}_1 C_1 + \tilde{\xi}_2 C_2,$$

Table 5: Simulation results that describe the performance of point and 80% interval estimates for the treatment effect $\beta = 0$.

	Design A $(\beta, \tilde{\xi}_1, \tilde{\xi}_2, \gamma_1, \gamma_2) = (0, 1, 1, .5, .5)$			
Method	Bias	SD†	Coverage	Length
GOLD	0.003	0.555	77.2%	1.306
MLE	0.002	0.531	77.4%	1.270
BAYES	-0.006	0.571	76.3%	1.303
SEQUENTIAL BAYES	0.097	0.527	76.7%	1.261
UNADJUSTED	0.686	0.437	35.4%	1.058

	Design B $(\beta, \tilde{\xi}_1, \tilde{\xi}_2, \gamma_1, \gamma_2) = (0, .5, .5, 1, 1)$			
Method	Bias	SD†	Coverage	Length
GOLD	-0.004	0.518	79.4%	1.291
MLE	-0.004	0.510	79.6%	1.281
BAYES	-0.016	0.540	78.4%	1.289
SEQUENTIAL BAYES	0.018	0.524	79.0%	1.280
UNADJUSTED	0.693	0.422	34.3%	1.058

† Standard Deviation (SD)

and method UNADJUSTED, which we define as logistic regression of Y on X ignoring (C_1, C_2) altogether.

Table 5 summarizes the performance of point and 80% interval estimates for the treatment effect parameter $\beta = 0$ when data are simulated using Designs A and B. The columns entitled Bias and Standard Deviation (SD) give the empirical average bias and sample standard deviation of the 1000 point estimates for β calculated using each method. Coverage probability and average length summarize the distribution of 80% interval estimates for β.

As expected, UNADJUSTED performs badly with large bias and poor coverage because it ignores the confounders altogether, whereas GOLD succeeds in eliminating confounding. MLE, BAYES and SEQUENTIAL BAYES all succeed in reducing bias. In terms of efficiency as measured by SD, the performance of MLE is excellent. Note that in large samples, GOLD point

DOI: 10.2202/1557-4679.1205

estimates have the highest possible efficiency because they are maximum likelihood estimates calculated using the correct model for the data. BAYES point estimates of β have lower efficiency than MLE amd GOLD. BAYES interval estimates are wider on average than MLE, but confer no noticeable improvement in coverage probability. These findings are echoed by McCandless et al. (2009) who conducted detailed simulations comparing MLE and BAYES under various data generating scenarios. The bottom of Table 5 summarizes the performance of SEQUENTIAL BAYES and shows an improvement in efficiency of point estimates compared to BAYES. The interval estimates have smaller average width and match more closely with those from MLE.

6 Discussion

In Bayesian analysis, different sources of information are combined via Bayes theorem. If one of the sources of information is incorrect, for example through model misspecification, then this can adversely affect the fitting of other components of the model. We illustrate this reasoning in the context of Bayesian regression adjustment for the PS. Joint estimation of the treatment and outcome models produces feedback that can bias the PS estimates. We note that the feedback is not inherently Bayesian and would also emerge from any analysis that uses the combined likelihood.

If we are confident about modelling assumptions, then the likelihood principle dictates that we ought to combine the likelihoods for the treatment and outcome when calculating inferences (Robins and Ritov, 1997; Tan, 2006). Consequently, the best analytic approach to reducing feedback in the PS is improved modelling of the outcome variable. For example, we could use a more flexible specification for the relationship between the PS and stroke risk in equation (2).

However this is can be a challenging exercise. In the THIN data, the PS for each patients is to some extent an artifact of the matching process. The relationship between the outcome and PS has little biological meaning and can vary from one study to the next. It can also be argued that careful modelling of the outcome is not in the spirit PS techniques. The outcome model is of lesser importance, and PS techniques focus on careful modelling of the treatment assignment mechanism. Thus cutting feedback seems like a sensible strategy for obtaining robust inferences that restrict the flow of information between models during estimation. It also incorporate stochastic uncertainty in the PS while simplifying Bayesian computation.

A difficulty with cutting feedback in a Bayesian propensity analysis is that there is no guarantee that the Markov chain is converging to a sensible equilibrium distribution. By not updating from the full conditional distribution for γ in equation (4) it means that there is no underlying full probability model for the data and parameters. Note however that, as argued in Section 4.2, the MCMC chain for γ is guaranteed to converge to the posterior distribution for γ from Bayesian logistic regression of X on C. See Lunn et al. (2009a) for further discussion of sampler convergence in sequential Bayesian analysis.

A possible remedy is to build an MCMC chain that updates from the full conditionals of the desired target distribution. To illustrate how one might accomplish this, first suppose that the parameter γ is known. Then the posterior distribution for the outcome model parameters under a Weibull proportional hazards regression is

$$p(\alpha,\beta,\xi,\lambda|data,\gamma) \;=\; \Big[\prod_{i=1}^{n} p(T_i|X_i,C_i,\delta_i,\alpha,\beta,\xi,\lambda,\gamma)\Big]p(\alpha,\beta,\xi,\lambda)/Q(\gamma)$$

where $Q(\gamma) = \int \Big[\prod_{i=1}^{n} p(T_i|X_i,C_i,\delta_i,\alpha,\beta,\xi,\lambda,\gamma)\Big]p(\alpha,\beta,\xi,\lambda)d(\alpha,\beta,\xi,\lambda)$ is the constant of normalization. To prevent feedback from the outcome when fitting a Bayesian propensity analysis, we can construct the desired target density as

$$\begin{aligned} p(\alpha,\beta,\xi,\lambda,\gamma|data) &= p(\alpha,\beta,\xi,\lambda|data,\gamma)\times p(\gamma|data) \\ &\propto \left\{\frac{\prod_{i=1}^{n} p(T_i|X_i,C_i,\delta_i,\alpha,\beta,\xi,\lambda,\gamma)p(X_i|C_i,\gamma)}{Q(\gamma)}\right\} p(\alpha,\beta,\xi,\lambda,\gamma). \end{aligned}$$

This posterior distribution is identical to the posterior density used by BAYES in equation (3), except that it reweights the density using the normalizing constant $Q(\gamma)$. This formulation incorporates the posterior distribution for γ from Bayesian logistic regression of X on C into the analysis, but without joint fitting of the treatment and outcome model. Unfortunately, MCMC computation requires knowledge of $Q(\gamma)$. For Gaussian response models, one could work out $Q(\gamma)$ analytically. Chen et al. (1999) describes computation of normalizing constants for logistic regression. Other more general computational strategies are discussed by Robert and Casella (2004).

DOI: 10.2202/1557-4679.1205

References

Chen, M., Ibrahim, J. G., Yiannoutsos, C., 1999. Prior elicitation, variable selection and Bayesian computation for logistic regression models, Journal of the Royal Statistical Society, Series B 61, 223–242.

Ibrahim, J. G., Chen, M. H., and Sinha, D., 2004. Bayesian survival analysis, New York: Springer Verlag.

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K. and Robins, J. M., 2006. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect, American Journal of Epidemiology 163, 262–70.

Liu, F., Bayarri, M. and Berger, J., 2009. Modularization in Bayesian analysis, with emphasis on analysis of computer models, Bayesian Analysis 4, 119–50.

Lunceford, J. K. and Davidian, M., 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study, Statistics Medicine 23, 2937–60.

Lunn, D., Best, N., Spiegelhalter, D., Graham, G., and Neuenschwander, B., 2009a. Combining MCMC with "sequential" PKPD modelling, Journal of Pharmacokinetics and Pharmacodynamics 36, 19–38.

Lunn, D., Spiegelhalter D. Thomas, A., Best, N., and Lunn, D., 2009b. The BUGS project: Evolution, critique and future directions, Statistics in Medicine (in press).

McCandless, L. C., Gustafson, P. and Austin, P. C., 2009. Bayesian propensity score analysis for observational data, Statistics in Medicine 28, 94–112.

Plummer, M., Best, N., Cowles, K., and Vines K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC, R News 6, 7–11.

Robert, C. P., and Casella, G., 2004. Monte Carlo statistical methods, New York: Springer Verlag.

Rice, K. M., 2004. Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications, Journal of the American Statistical Association 99, 510–522.

Robins, J. M., and Ritov, Y., 1997. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models, Statistics in Medicine 16, 285–319.

Rosenbaum, P. R. and Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects, Biometrika 70, 41–57.

Rubin, D. B., 1985. The use of propensity scores in applied Bayesian inference. In Bayesian Statistics 2, eds. Bernardo J. M., De Groot M. H., Lindley D. V., Smith A. F. M. Valencia University Press: Valencia, 463–72.

Rubin, D. B., 2008. For objective causal inference, design trumps analysis, Annals of Applied Statistics 2, 808–40.

Rubin, D. B., Thomas, N., 1996. Matching using estimated propensity scores: Relating theory to practice. Biometrics 52, 249–64.

Rutishauser, J., 2006. The role of statins in clinical medicine LDL-cholesterol lowering and beyond, Swiss medical weekly 136, 41–9.

Smeeth, L., Douglas, I., Hall, A.J., Hubbard, R., and Evans, S., 2008. Effect of statins on a wide range of health outcomes: A cohort study validated by comparison with randomized trials, British Journal of Clinical Pharmacology 67, 99–109.

Tan, Z., 2006. A distributional approach for causal inference using propensity scores, Journal of the American Statistical Association 101, 1619–37.

DOI: 10.2202/1557-4679.1205