A Propensity-Score Integrated Approach to Bayesian Dynamic Power Prior Borrowing

Jixian Wang, Hongtao Zhang, and Ram Tiwari

ABSTRACT

Use of historical control data to augment a small internal control arm in a randomized control trial (RCT) can lead to significant improvement of the efficiency of the trial. It introduces the risk of potential bias, since the historical control population is often rather different from the RCT. Power prior approaches have been introduced to discount the historical data to mitigate the impact of the population difference. However, even with a Bayesian dynamic borrowing which can discount the historical data based on the outcome similarity of the two populations, a considerable population difference may still lead to a moderate bias. Hence, a robust adjustment for the population difference using approaches such as the inverse probability weighting or matching, can make the borrowing more efficient and robust. In this article, we propose a novel approach integrating the propensity score for the covariate adjustment and Bayesian dynamic borrowing using power prior. The proposed approach uses Bayesian bootstrap in combination with the empirical Bayes (EB) method using quasi-likelihood for determining the power prior. The performance of our approach is examined by a simulation study. We apply the approach to two Acute Myeloid Leukemia (AML) studies for illustration.

1. Introduction

In situations where the control arm of a randomized clinical trial (RCT) is smaller than the test arm to treat more patients with the test treatment, in order that statistical inference for treatment comparison is not compromised due to the small control arm, the use of external data, which may be from another clinical trial or real-world data (RWD), has become a valuable source for augmenting the internal control of the RCT. This approach is often referred to as borrowing external controls. Regulatory guidance documents on the use of RWD, to aid drug development, have been published (European Medicines Agency 2020; Food and Drug Administration 2018). However, the use of external data also introduces the risk of potential bias, as the historical control population may be rather different from the RCT.

There are several approaches to eliminate or reduce the bias due to population difference. In particular, robust adjustment for the population difference using propensity score (PS) based approaches such as the inverse probability weighting (IPW) or matching can make the borrowing more efficient and robust to model misspecification to some extent (Rosenbaum and Rubin 1983; Robins, Rotnitzky, and Zhao 1994). If we assume that the adjustment can eliminate the bias, one may be tempted to use a single arm trial with adjusted external control completely. However, some assumptions such as no unobserved confounders can not be verified based on the data. Therefore, using adjusted external control has the potential risk of introducing confounding bias.

An RCT, with a small control arm, may provide an internal reference for evaluating the difference between the internal and external control populations. To deal with the population difference, power prior approaches (Ibrahim and Chen 2000; Ibrahim, Chen, and Sinha 2003; Hobbs et al. 2011; Hobbs, Carlin, and Sargent 2013; Neuenenschwander, Branson, and Spiegelhalter 2009) can be used to discount the historical data to mitigate the impact of the bias. The amount of borrowing can be either fixed or determined by discounting the historical data, based on the similarity of the outcomes of the two populations, known as Bayesian dynamic borrowing. However, there may still be considerable differences that lead to moderate bias. To mitigate the impact of population difference, some recently developed approaches used PS matching or stratification to reduce the difference within matched pairs or strata, and then applied the power prior within them (Wang and Rosner 2019; Wang et al. 2019; Sachdeva, Tiwari, and Guha 2021). An alternative approach used a linear outcome model assuming exchangeability after covariate adjustment (Kotalik et al. 2021).

As a further development based on the above-mentioned work, we propose a novel approach integrating PS based approaches for the covariate adjustment and Bayesian dynamic borrowing using the power prior (Ibrahim and Chen 2000; Ibrahim, Chen, and Sinha 2003; Hobbs et al. 2011; Hobbs, Carlin, and Sargent 2013; Chen et al. 2011; Gravestock and Held 2017, 2018; Wang and Rosner 2019; Wang et al. 2019). The proposed approach combines the advantages of propensity score based approaches for adjusting confounding bias without specifying the outcome model, and the power prior that down-weights the information from the historical data if, after adjustment, it is still considerably different from the internal control. Our approach is an approximate full Bayesian that takes into account the uncertainty of model fitting and weighting.
The major challenge is that the PS based methods are frequentist approaches with minimum model specification, while the power prior methods are built in the Bayesian framework. Our approach is partially built on the work of approximate PS-based Bayesian approaches (Graham, McCoy, and Stephens 2016; Saarela, Belzile, and Stephens 2016; Capistrano, Moodie, and Schmidt 2019), in which IPW and doubly robust (DR) estimation approaches were placed in the Bayesian framework and the posterior distribution was approximated by Bayesian bootstrap (BB) (Rubin 1981). Our approach is considerably simpler than the outcome based full Bayesian approach using MCMC such as Kotalik et al. (2021).

2. A Review of Relevant Approaches

2.1. Borrowing Historical Controls to Augment an Internal Control Arm

First, we state our task of borrowing historical controls to augment an internal control arm formally. Let $D_i = (y_i, X_i, H_i), i = 1, \ldots, n$, be the outcome, covariates, and indicator of historical control for the $i$th subject in the combination of the internal trial population and the historical control population. The sample sizes and means of the internal and historical control populations are: $n_i = \sum_{i=1}^{n} H_i, n_0 = n - n_i, \bar{y}_0 = \sum_{i=1}^{n} (1 - H_i)y_i/n_0$ and $\bar{y}_h = \sum_{i=1}^{n} H_iy_i/n_i$, respectively. We also denote the whole dataset as $D = (D_1, \ldots, D_n)$, and those of the internal and external controls as $D^0 = D|H_i = 0$ and $D^h = D|H_i = 1$, respectively. Although our final goal is to evaluate the treatment effect of the treatment applied in the treated arm in the trial population, the key issue we concentrate on is the evaluation of the outcome under the control: $\mu = E(y_i|H_i = 0)$, borrowing historical control data with confounding adjustment. Here, $y_i$ could be either a continuous or binary variable, in the latter case, $\mu$ is the rate or proportion of the outcome. Due to the population difference between the trial and the historical control, $E(y_h) = E(y_i|H_i = 1)$ is likely different from $\mu$, therefore, an adjustment for the population difference is often necessary.

2.2. Propensity-Score Based Adjustment

A commonly used approach for population adjustment is based on PS defined as the probability of belonging to the historical control, given the covariates $X_i$:

$$e_i = P(H_i = 1|X_i) \quad (1)$$

which is often modeled by a logistic regression as $P(H_i|X_i, \gamma)$ with parameters $\gamma$ in the logistic model. Key assumptions for PS based adjustment include $0 < P(H = 1|X) < 1$ for any $X$, and that potential outcomes are independent of $H_i$ conditional on $X_i$. Details can be found in Rosenbaum and Rubin (1983). Under these assumptions, $y_i \perp H_i|e_i$, hence, we can use the IPW estimator

$$\hat{\mu}_{ipw} = \left( \sum_{i=1}^{n} H_iw_i \right)^{-1} \left( \sum_{i=1}^{n} H_iw_iy_i \right) \quad (2)$$

where the weight $w_i = (1 - e_i)/e_i$ is the odd of being in the trial population to transported historical data to the trial population, to estimate $\mu$. Using the property $y_i \perp H_i|e_i$, it is straightforward to show that $E(\hat{\mu}_{ipw}) = \mu$. Therefore, when the PS model is correctly specified, one can combine $\hat{\mu}_{ipw}$ with the internal control mean $\bar{y}_0$ for a more accurate estimation of $\mu$. This IPW estimator (2) is slightly different from the standard IPW estimator for the average treatment effect (ATE) for the whole population, since we aim at estimating the control effect in the trial population.

2.3. Propensity-Score based Bayesian Methods

Although PS based approaches were proposed from frequentist aspect, effort has been made to use PS in the Bayesian framework to provide a robust Bayesian approach for population adjustment. Early work (McCandless, Gustafson, and Austin 2009) used joint modeling for the PS and outcome, with the PS as a covariate covariate in the outcome model. Multiple improvements have been proposed on the estimation of PS model parameters for “cutting feedback” (McCandless et al. 2010; Zigler et al. 2013; Zigler 2016).

An alternative to the modeling approach uses the IPW estimator and resampling to obtain the posterior distribution of $\mu$ (Graham, McCoy, and Stephens 2016; Capistrano, Moodie, and Schmidt 2019). Capistrano et al. proposed an approach using BB (Rubin 1981; Newton and Raftery 1994):

1. Repeat Steps 2–4 (below), for $s = 1, \ldots, S$, times.
2. Generate $\xi_{is}, i = 1, \ldots, n$, independently, from the uniform Dirichlet distribution.
3. Fit the PS model (1) for $H_i$ using weights $\xi_{is}$.
4. Obtain

$$\hat{\mu}_{ipws} = \left( \sum_{i=1}^{n} H_iw_i\xi_{is} \right)^{-1} \sum_{i=1}^{n} H_i\xi_{is}w_iy_i \quad (3)$$

The $S$ estimates $\hat{\mu}_{ipws}$ can be considered as posterior samples of $\mu$. An in-depth consideration of their properties, including formulating this approach in the Bayesian framework using de Finetti’s representation (De Finetti 1974) can be found in Saarela, Belzile, and Stephens (2016).

2.4. Power Prior and Bayesian Dynamic Borrowing

Although the previous approaches can provide adjusted estimates of $\mu$ in a frequentist or Bayesian framework, combining them with $\bar{y}_0$ remains a challenge, especially when IPW adjustment may not completely eliminate confounding bias due to population difference. For mitigating the potential bias, Bayesian borrowing with the power prior is a powerful tool for this purpose (Ibrahim and Chen 2000; Ibrahim, Chen, and Sinha 2003; Hobbs et al. 2011; Hobbs, Carlin, and Sargent 2013; Chen et al. 2011; Gravestock and Held 2017, 2018; Wang and Rosner 2019; Wang et al. 2019). Let $D^0$ and $D^h$ denote data from the internal and historical control data, respectively, $\theta$ denote model parameters, and $L(\theta|D)$ denote the likelihood function given data $D$. The power prior, conditional on $D^h$, is formulated as

$$\pi(\theta|D^h, a_0) \propto L(\theta|D^h)^{a_0} \pi_0(\theta) \quad (4)$$
where \( 0 \leq a_0 \leq 1 \) is the power prior (discounting) parameter in the likelihood of historical data, and \( \pi_0(\theta) \) is the initial prior for \( \theta \). The corresponding posterior distribution is

\[
\pi(\theta|D^h, D^0, a_0) \propto L(\theta|D^0)L(\theta|D^h)^{a_0}\pi_0(\theta)
\]

The parameter \( a_0 \) allows one to control the weight of historical data in the current study. One can choose \( a_0 \) close to zero when historical and current data are highly heterogeneous, and choose \( a_0 \) close to one when they are congruent.

With the power prior model, our goal is to estimate \( \mu \), the mean response given control data \( D^0 \) and \( D^h \). Assume that \( y_h \) and \( y_0 \) are normally distributed with common mean \( \mu \), and variances \( \sigma_h^2 \) and \( \sigma_0^2 \) in the historical and internal control populations, respectively. With fixed \( a_0 \), the posterior distribution of \( \mu \) is

\[
\mu \sim N(\tilde{\mu}, \tilde{\sigma}^2)
\]

where

\[
\tilde{\mu} = a_0 \sigma_h^{-2} \bar{y}_h + \sigma_0^{-2} \bar{y}_0
\]

\[
\tilde{\sigma}^2 = (a_0 \sigma_h^{-2} + \sigma_0^{-2})^{-1}.
\]

In applications, \( \sigma_h^2 \) and \( \sigma_0^2 \) can be replaced by the corresponding sample variances.

One can also consider \( a_0 \) as random, with a noninformative prior. Then, its distribution, consequently the amount of borrowing, will depend on the similarity between \( D^h \) and \( D^0 \). A simple approach to estimate it uses empirical Bayesian (EB) approach (Gravestock and Held 2017, 2018). This approach estimates \( a_0 \) by maximizing the marginal likelihood \( L(a_0|D^0, D^h) \). For the normal distribution case above, it has a closed form:

\[
a_0 = \frac{\sigma_h^2}{\max[(\bar{y}_h - \bar{y}_0)^2, \sigma_h^2 + \sigma_0^2] - \sigma_0^2}.
\]

For binomial outcomes, with a Beta(1, 1) prior and a fixed \( a_0 \), the posterior distribution of \( \mu \) can be written as

\[
\mu \sim \text{Beta}(a_0 y_h + y_0 + 1, n_0 + a_0(n_0 - y_h) - y_0 + 1)
\]

where \( y_h = \sum_{i=1}^{n_h} H_i y_i \) and \( y_0 = \sum_{i=1}^{n_0} (1 - H_i) y_i \).

Assuming a random \( a_0 \), its marginal likelihood is (eq. (4), Gravestock and Held 2018)

\[
\pi(a_0|D^h, D^0) \propto \frac{\text{Beta}(a_0 y_h + y_0 + 1, n_0 + a_0(n_0 - y_h) + n_0 - y_0 + 1)}{\text{Beta}(a_0 y_h + 1, a_0(n_0 - y_h) + 1)}.
\]

To determine \( a_0 \), we can use the approach of Gravestock and Held (2017) and find

\[
a_0^* = \text{argmax}_{a_0} \pi(a_0|D^h, D^0)
\]

within the range of \([0,1]\). In practice, this can be done with a grid search.

### 2.5. Bayesian Borrowing with Covariate Adjustment

The extent of borrowing in the dynamic approach (Gravestock and Held 2017, 2018) depends on the difference in observed mean responses between internal and external control groups. A part of the difference may be due to the difference in \( X_i \) between internal and external controls. In order to reduce such a difference, some approaches based on PS have been proposed. Wang and Rosner (2019) and Wang et al. (2019) proposed a stratification-based on the PS and using the overlapping area between the PS distributions of subjects of historical and internal controls. This approach uses all subjects from the historical control who meet inclusion/exclusion criteria and discounts those with a large difference to mitigate their impact. The amount of borrowing from individuals is determined by the similarity in the PS distributions, rather than the outcomes, between the internal and external controls.

### 3. A PS Integrated Approximate Bayesian Approach

Our integrated approach combines posterior sampling using BB, PS adjustment with the IPW estimator, and the EB estimator for the power prior parameter \( a_0 \). One advantage is that \( a_0 \) is not fixed, and a posterior distribution can be obtained with BB. We use BB to generate posterior samples of the mean and variance for both internal and historical controls. For the latter, we adapt the IPW estimator (2) as did in Capistrano, Moodie, and Schmidt (2019) to adjust for confounding bias, and the justification of combining the BB weights with the IPW weights as described in sec. 6 of Saarela, Belzile, and Stephens (2016). We start with introducing some general notation here. Let \( l(\theta, D_i) \) be the log-likelihood function for an hypothetical new observation \( D_i \) with parameters \( \theta \). Following Saarela, Belzile, and Stephens (2016), we maximize \( E(l(\theta, D_i)|D) \), which can be estimated by BB as

\[
E(l(\theta, D_i)|D) \approx \sum_{i=1}^{n} w_i(\xi) \xi_i l(\theta, D_i)
\]

where \( \xi_i \sim \text{Dirichlet}(1, \ldots, 1) \) are the BB weights, \( \xi = (\xi_1, \ldots, \xi_n) \) is a full set of the weights, and \( w_i(\xi) \) is the IPW weight, depending on the BB weights as well. The expectation in \( E(l(\theta, D_i)|D) \) is not taken over \( D_i \), but the random weights, or, in the Bayesian interpretation of Rubin (1981), over a non-parametric prior on the skeleton of \( D \). Then \( \theta \) is estimated as

\[
\hat{\theta} = \text{argmax}_\theta \left( \sum_{i=1}^{n} w_i(\xi) \xi_i l(\theta, D_i) \right).
\]
internal controls, respectively, as posterior samples for \( \mu \). For the latter, we weight \( y_i \)'s of internal control by BB weights

\[
\hat{y}_0 = \left( \sum_{i=1}^{n} (1 - H_i) \xi_i \right)^{-1} \sum_{i=1}^{n} (1 - H_i) \xi_i y_i,
\]

(14)
hence, \( \hat{y}_0 \) can be considered as a posterior sample for \( \mu \) given data \( D^0 \). The BB weighting is necessary for the internal controls, even adjustment is not needed, to obtain the approximate posterior samples for \( \mu \). To obtain a posterior sample \( \hat{y}_h \) based on \( D^h \) adjusted for population difference, we can use (3), or another adjustment approach. Although both \( \hat{y}_h \) and \( \hat{y}_0 \) depend on \( \xi \), we will suppress \( \xi \) in the notation for simplicity.

The second step is to combine \( \hat{y}_h \) and \( \hat{y}_0 \) based on a quasi log-likelihood conditional on a BB realization \( \xi \):

\[
l(\mu, a_o | \xi) \propto -\left[ (\hat{y}_0 - \mu)^2/\hat{\sigma}_0^2 + a_0(\hat{y}_h - \mu)^2/\hat{\sigma}_h^2 \right]
\]

(15)
where \( \hat{\sigma}_h^2 \) and \( \hat{\sigma}_0^2 \) are the weighted sample variances of \( y_0 \) and \( y_h \), respectively, and we denote it as \( l(\mu, a_o | \xi) \) to emphasize its dependence on the BB weights via \( \hat{\sigma}_h^2, \hat{\sigma}_0^2, \hat{y}_0 \) and \( \hat{y}_h \). For example, \( \hat{\sigma}_0^2 = \sum_{i=1}^{n} (1 - H_i) \xi_i^2 (y_i - \hat{y}_0)^2/(n_0 - 1) \) and \( \hat{\sigma}_h^2 \) is normalized such that \( \sum_{i=1}^{n} (1 - H_i) \xi_i^2 \hat{\sigma}_h^2 = n_0 \hat{\sigma}_h^2 \) is calculated in the same way as for \( \hat{\sigma}_0^2 \) using normalized \( \xi_i wi(\xi_i) \).

In our simulation and analysis below, we use R-function 

\texttt{wtd.var(.)} in the Hmisc package (Harrell 2021). Equation (15) has the same form as the likelihood part of eq. (12) of Gravestock and Held (2017), leading to (8) and (6). Therefore, we have

\[
\hat{\mu}(a_0) = \hat{\sigma}_h^2 (a_0 \hat{\sigma}_0^2 \hat{y}_h + \hat{\sigma}_h^2 \hat{y}_0) /
\]

\[
(2) \quad (a_0 \hat{\sigma}_0^2 + \hat{\sigma}_h^2)^{-1}.
\]

(16)

For a fixed \( a_0 \), \( \hat{\mu}(a_0) \) follows the posterior distribution given \( D^h \) and \( D^0 \).

To use \( \hat{\mu}(a_0) \) in practice, we use EB to estimate \( a_0 \) following Gravestock and Held (2017). Replacing \( \hat{y}_h, \hat{y}_0 \) with \( y_h, y_0, \hat{\sigma}_0^2, \hat{\sigma}_h^2 \) with \( \hat{\sigma}_0^2 \) and \( \hat{\sigma}_h^2 \) in (8), we have

\[
\hat{a}_0 = \frac{\hat{\sigma}_h^2}{\max[(\hat{y}_h - \hat{y}_0)^2, \hat{\sigma}_h^2 + \hat{\sigma}_0^2] - \hat{\sigma}_0^2}.
\]

(17)

Replacing \( a_0 \) with \( \hat{a}_0 \) in (6) and (7), we obtain a BB sample of \( \mu \). The following Algorithm 1 gives steps of the integrated algorithm for normally distributed outcomes, in which we omit the dependency on \( \hat{a}_0 \) for notation simplicity. It repeats the above steps \( s = 1, \ldots, S \) times to obtain a posterior sample of \( \mu \). Then, \( \hat{\mu}, s = 1, \ldots, S \), can be considered as approximate posterior samples of \( \mu \). Steps 2–4 of the algorithm are similar to those in Capistrano, Moodie, and Schmidt (2019). Steps 5–7 are for Bayesian borrowing with \( a_0 \) determined by the EB approach, incorporated into the full BB steps. Note that \( \hat{\mu}_s \) is weighted by both the PS and the BB weights, although the PS is already a bootstrapped version. This is necessary for using BB for approximate Bayesian inference. See Saarela, Belzile, and Stephens (2016) for technical details.

However, as a referee rightly pointed out, using the EB estimate \( \hat{a}_0 \) to replace \( a_0 \) is not included in the framework of Saarela, Belzile, and Stephens (2016) and Capistrano, Moodie, and Schmidt (2019). But this can be justified as follows. First we show that when \( n \to \infty, \hat{a}_0 \to a_0 \), which is either 1 or 0, depending the difference between \( \hat{y}_h \) and \( \hat{y}_0 \), which in turn depends on the adjustment for population difference. Here, we implicitly assume that \( n \to \infty \) leads to samples sizes of both trial and historical data to infinity (not necessarily at the same rate).

With \( n \to \infty, (\hat{y}_h - \hat{y}_0)^2 \to \hat{\sigma}_h^2 + \hat{\sigma}_0^2 + \delta^2 \), where \( \delta \) is the true difference between the internal and adjusted external means. When there is no adjustment or the adjustment is invalid, \( \delta^2 > 0 \), but \( \hat{\sigma}_h^2 \) and \( \hat{\sigma}_0^2 \) converge to zero. This leads to \( \hat{a}_0 \to 0 \), which means no borrowing to avoid biases. In contrast, with a valid adjustment, \( \delta^2 = 0 \), hence by (18) \( \hat{a}_0 \to \hat{\sigma}_h^2/((\hat{\sigma}_h^2 + \hat{\sigma}_0^2) - \delta^2) \to 1 \), leading to full borrowing. To obtain asymptotic properties of \( \hat{\mu}(a_0) \), we replace \( a_0 \) with \( \hat{a}_0 \) in (17) and (16) and apply Slutsky's Lemma (van der Vaart 2000; Lemma 2.8) repeatedly on them, and note that (17) is greater than zero almost surely. This leads to \( \hat{\mu}(\hat{a}_0) \to \hat{\mu}(a_0) \) in distribution, where \( a_0 = 0 \) or \( a_0 = 1 \). Therefore, \( \hat{\mu}_s \) is a posterior sample of \( \mu \), either conditional on \( D^0 \) only (where \( a_0 = 0 \)) or on both \( D^h \) and \( D^0 \) (when \( a_0 = 1 \)).

The above algorithm can be easily adapted for binomial outcomes. The major difference is in Steps 5–7, which should be replaced by:

5: In (10) replace \( y_h, \) with \( n_h y_h + y_0, \) with \( n_0 \hat{y}_0 \).
6: Obtain \( \hat{a}_0 \) from (11) with a grid search. In this article, we search using a 0.02 grid in the range [0,1].
7: Obtain

\[
\hat{\mu}_s = \frac{\hat{a}_0 n_h y_h + n_0 \hat{y}_0 + 1}{\hat{a}_0 n_h + n_0 + 2}.
\]

(19)

In the last step, the posterior mean is taken as \( \hat{\mu}_s \).

4. A Simulation Study

A simulation study is conducted to examine the performance of the integrated approach. The difference between the two control populations is represented by the mean difference in \( X_i \). The simulation data generating models and parameters are:

- \( X_i \sim N(0, I_p) \) when \( H_i = 0 \) and \( X_i \sim N(-b, I_p) \) when \( H_i = 1 \), where \( I_p \) is a \( p \)-dim identity matrix, and varying \( b \) represents the population difference between internal and historical controls. This leads to a logistic PS model.
The outcome model is \( Y_i = \beta^T X_i + \varepsilon_i \), where \( \varepsilon \sim N(0, 1) \) hence the population difference leads to a difference \( \beta^T b \).

- \( \beta = \beta_1 \) and \( b = b_1 \), where \( 1_p \) is a \( p \)-vector of ones. We set \( p = 5 \) or \( p = 10 \), \( \beta = 0.3 \) and vary \( b \). The last is the key parameter as a measure of population difference.

- We run 1000 simulations. For each, 100 bootstrap runs are used to estimate the posterior distribution of \( \mu \).

The simulation examines four estimators: Average differences with no and full borrowing, and dynamic borrowing with and without IPW adjustment. Without IPW adjustment, the above settings would lead to a negative bias of different sizes in the combined estimation of \( \mu \).

Figure 1 shows the posterior distributions of \( \mu \) of dynamic borrowing with and without IPW adjustment by mean difference of covariates between internal and historical controls \( E(X_i|H_i = 0) = 0 \) and \( E(X_i|H_i = 1) = -b \), compared with those of no borrowing and full borrowing, with normally distributed outcomes. The correct mean outcome under the control treatment is \( E(Y_i|H_i = 0) = 0 \).

- Figure 1 shows the posterior distributions of \( \mu_{ob} \) with and without IPW adjustment by mean difference of covariates between internal and historical controls \( E(X_i|H_i = 0) = 0 \) and \( E(X_i|H_i = 1) = -b \), compared with those of no borrowing and full borrowing, with normally distributed outcomes. The correct mean outcome under the control treatment is \( E(Y_i|H_i = 0) = 0 \).

Figure 1 shows the posterior distributions of \( \mu \) of dynamic borrowing with and without IPW adjustment by mean difference of covariates between internal and historical controls \( E(X_i|H_i = 0) = 0 \) and \( E(X_i|H_i = 1) = -b \), compared with those of no borrowing and full borrowing, with normally distributed outcomes. The correct mean outcome under the control treatment is \( E(Y_i|H_i = 0) = 0 \).

Table 1 gives the mean bias, variance, MSE and variance ratio (relative to no borrowing) of the four estimators for normal outcomes with \( p = 5 \) and \( p = 10 \). As expected, when there is no difference between the internal and external control populations, using IPW adjustment increases variability. However, when there is a difference, using IPW reduces not only the bias, but also the variability for \( p = 5 \). The results with \( p = 10 \) have a similar pattern, except that all estimators perform worse than those with \( p = 5 \). Particularly, when \( b = 0.6 \), the variance ratios of the two dynamic borrowing estimators are almost 1. In general, the bias of dynamic borrowing without IPW is greater than that with IPW, but still much smaller than the full borrowing bias when \( b > 0 \). This result is consistent with the properties we claim at the end of the section above, because dynamic borrowing without IPW can be considered as a surrogate scenario of no proper population adjustment.
Table 1. Mean bias, variance, MSE and variance ratio (vs. no borrowing) of four estimators with no and full borrowing, and dynamic borrowing with and without IPW adjustment for continuous outcome.

| p | b | Method        | Bias    | Variance | MSE     | Var ratio |
|---|---|---------------|---------|----------|---------|-----------|
| 5 | 0 | No borrowing  | 0.003   | 0.028    | 0.029   | 1.000     |
|   |   | Full borrowing| 0.001   | 0.014    | 0.014   | 0.497     |
|   |   | Dynamic + IPW | 0.003   | 0.026    | 0.026   | 0.901     |
|   |   | Dynamic       | 0.003   | 0.022    | 0.022   | 0.770     |
| 0.15 | 0 | No borrowing  | 0.006   | 0.029    | 0.029   | 1.000     |
|   |   | Full borrowing| −0.106  | 0.014    | 0.025   | 0.492     |
|   |   | Dynamic + IPW | 0.029   | 0.024    | 0.023   | 0.843     |
|   |   | Dynamic       | −0.022  | 0.026    | 0.027   | 0.898     |
| 0.3 | 0 | No borrowing  | −0.007  | 0.030    | 0.030   | 1.000     |
|   |   | Full borrowing| −0.232  | 0.015    | 0.069   | 0.502     |
|   |   | Dynamic + IPW | 0.013   | 0.024    | 0.024   | 0.814     |
|   |   | Dynamic       | −0.041  | 0.032    | 0.034   | 1.070     |
| 0.6 | 0 | No borrowing  | −0.001  | 0.029    | 0.029   | 1.000     |
|   |   | Full borrowing| −0.451  | 0.015    | 0.218   | 0.527     |
|   |   | Dynamic + IPW | 0.001   | 0.027    | 0.027   | 0.946     |
|   |   | Dynamic       | −0.018  | 0.030    | 0.031   | 1.048     |
| 10 | 0 | No borrowing  | −0.004  | 0.037    | 0.037   | 1.000     |
|   |   | Full borrowing| −0.003  | 0.019    | 0.019   | 0.512     |
|   |   | Dynamic + IPW | −0.004  | 0.035    | 0.035   | 0.949     |
|   |   | Dynamic       | −0.004  | 0.028    | 0.029   | 0.770     |
| 0.15 | 0 | No borrowing  | 0.001   | 0.038    | 0.038   | 1.000     |
|   |   | Full borrowing| −0.223  | 0.020    | 0.069   | 0.520     |
|   |   | Dynamic + IPW | 0.020   | 0.033    | 0.032   | 0.861     |
|   |   | Dynamic       | −0.039  | 0.040    | 0.041   | 1.042     |
| 0.3 | 0 | No borrowing  | −0.003  | 0.038    | 0.038   | 1.000     |
|   |   | Full borrowing| −0.449  | 0.019    | 0.221   | 0.494     |
|   |   | Dynamic + IPW | 0.001   | 0.035    | 0.035   | 0.930     |
|   |   | Dynamic       | −0.026  | 0.040    | 0.041   | 1.067     |
| 0.6 | 0 | No borrowing  | 0.002   | 0.037    | 0.037   | 1.000     |
|   |   | Full borrowing| −0.899  | 0.019    | 0.826   | 0.506     |
|   |   | Dynamic + IPW | 0.001   | 0.037    | 0.037   | 0.981     |
|   |   | Dynamic       | −0.009  | 0.037    | 0.037   | 1.013     |

A similar simulation has also been performed for binary outcomes. The simulation setting is similar to those above, except that $Y_i \sim \text{Bin}(p_i)$ with $p_i = 1/(1 + \exp(-\beta^T X_i))$. The results with $p = 5$ and different $b$ values are presented in Figure 2, showing a similar pattern to that in Figure 1. Table 2 gives mean bias, variance, MSE and variance ratio (vs. no borrowing) of four estimators with no and full borrowing, and dynamic borrowing with and without IPW adjustment for binary outcome with $p = 5$ and $p = 10$. Similar patterns as those of normally distributed outcomes are found.

5. An Illustrative Example

We illustrate our approach using publicly available datasets from two trials evaluating gemtuzumab ozogamicin (GO, a CD-33 targeted therapy) with chemotherapy for children and adolescents with acute myeloid leukemia (AML): AML03P1 (Cooper et al. 2012) and AML0531 (Gamis et al. 2014). Both trials had a GO with chemotherapy arm which consisted of a remission induction phase (course 1) followed by an intensification phase (course 2). We use patients who had the status of complete remission (CR) ascertained at the end of course 2. The number of patients are 59 and 234, respectively from trial AML03P1 and AML0531.

Our aim is to borrow data from AML0531 to strengthen the small arm in AML03P1. We use IPW to control the difference in the following baseline factors: log-age, log-bone marrow leukemic blast percentage (log-BM), central nervous system (CNS) disease, race, risk group and white blood count (WBC) count at diagnosis. Table 3 presents the fitted PS model together with the mean/percentage differences of the above covariates. There is a substantial difference in log-WBC and also smaller differences in race and high-risk. The weighted differences are generally much smaller, in particular in log-BM, showing the effect of covariate balancing with IPW. To apply the proposed approach, the algorithm presented in Section 3 is implemented in R with the code given in the Appendix. Figure 3 shows the posterior density, median and 95% credible interval of the CR rate at the end of course 2 for AML03P1 with Bayesian dynamic borrowing with and without IPW adjustment, compared to those with full and no borrowing. The density with full borrowing is rather different from the original AML03P1 one (no borrowing), while the dynamic borrowing one is in between the two. The IPW adjusted one is more similar to the original one, but with less variability. The median(SE) of the four estimator(Bayesian dynamic borrowing without and with IPW, full and no borrowing) are 0.94(0.024), 0.95(0.017), 0.91(0.017) and 0.97(0.025), respectively. These results show the advantage of IPW adjustment before applying Bayesian dynamic borrowing when the difference between the trials is significant.

6. Discussion

We have proposed a novel approach integrating propensity score for the covariate adjustment and Bayesian dynamic borrowing using power prior. The approach combines the advantages of a propensity score based approach for adjusting confounding bias without specifying the outcome model, and the power prior that down-weights the historical data if, after adjustment, it is still considerably different from the internal control. Our approach is an approximate full Bayesian that takes into account the uncertainty of model fitting and weighting. Our approach uses Bayesian bootstrap, in combination with the empirical Bayesian method for determining the power prior, and is easier to implement than a full Bayesian approach using MCMC. The simulation results showed robust performance of our approach under different scenarios and is generally a better approach than dynamic borrowing without adjustment.

One advantage of our approach for drug development is that the IPW step does not depend on the outcome data. Therefore, both IPW and BB weights can be determined without access to outcome data and can be locked before the unlock of outcome database. For IPW weights, one may change the PS model, outcome data and can be locked before the unlock of outcome database. For BB weights, it would be sufficient to lock the seed of the random number generator.

We have focused on using IPW adjustment in our approach here for simplicity. Nevertheless, our approaches can also use the DR adjustment. The DR estimator combines $\hat{\mu}_{ipw}$ with a
J. WANG, H. ZHANG, AND R. TIWARI

Figure 2. Distribution of $\hat{\mu}_{cb}$ with and without IPW adjustment by mean difference of covariates between internal and historical controls $E(X_i|H_i = 0) = 0$ and $E(X_i|H_i = 1) = -b$, compared with those of no borrowing and full borrowing, with binary outcomes. The correct mean outcome under the control treatment is $\mu = 0.5$.

The prediction of $\mu$ using covariates:

$$\hat{\mu}_{dr} = \left( \sum_{i=1}^{n} H_i(1 - e_i)e_i^{-1} \right)^{-1} \times \sum_{i=1}^{n} e_i^{-1}[(1 - e_i)H_iY_i - (H_i - e_i)m_0(X_i, \hat{\beta})]$$

where $m_0(X_i, \hat{\beta})$ is an outcome model for $y_i$ with estimated parameters $\hat{\beta}$ such that

$$E\left( \sum_{i=1}^{n} (1 - H_i)m_0(X_i, \hat{\beta}) \right)/n_0 = \mu.$$

This estimator is DR, as it is consistent if either the PS model (1) or $m_0(X_i, \hat{\beta})$ is correctly specified. These frequentist approaches do not need specification of a full model, and hence are more robust. The use of BB approach based on $\hat{\mu}_{dr}$ has been proposed by Graham, McCoy, and Stephens (2016). It is clear that the DR estimator (20) uses two components: IPW and outcome prediction. BB approach based on outcome prediction has been well studied; therefore, we have omitted the DR approach. To use our approach, we apply the BB weights to fitting $m_0(X_i, \hat{\beta})$ as well as (20).

Our approach is flexible enough to use other adjustment approaches such as other (not PS based) covariate balancing weights; for example, the calibration estimation, including the so called match adjusted indirect comparison in health economics. These approaches weight historical control patients to balance the covariates that are potential prognostic factors, then use the same weights to obtain a weighted mean of the historical controls. As discussed above, our approach takes $\hat{\mu}_{ipw}$ as an adjusted estimator for $\mu$, then determines $\hat{\alpha}_0$ according to its similarity to the mean of internal controls. $\hat{\mu}_{ipw}$ can be replaced by the weighted mean with calibration estimation. Our approach can also be adapted to borrow from multiple sources of external controls. For example, for binary outcome, the approach can be combined with the method of Gravestock and Held (2018).
Table 2. Mean bias, variance, MSE, and variance ratio (vs. no borrowing) of four estimators with no and full borrowing, and dynamic borrowing with and without IPW adjustment for binary outcome.

| p | b  | Method      | Bias  | Variance | MSE  | Var ratio |
|---|----|-------------|-------|----------|------|-----------|
| 5 | 0  | No borrowing| 0.000 | 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.001| 0.002    | 0.002| 0.494     |
|  |  | Dynamic + IPW| 0.000| 0.004    | 0.004| 0.778     |
|  |  | Dynamic     | 0.000| 0.004    | 0.004| 0.765     |
|  | 0.3| No borrowing| −0.001| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.051| 0.002    | 0.002| 0.482     |
|  |  | Dynamic + IPW| −0.001| 0.004    | 0.004| 0.830     |
|  |  | Dynamic     | −0.012| 0.004    | 0.005| 0.911     |
|  | 0.6| No borrowing| −0.001| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.096| 0.002    | 0.012| 0.454     |
|  |  | Dynamic + IPW| −0.001| 0.005    | 0.005| 0.922     |
|  |  | Dynamic     | −0.012| 0.005    | 0.005| 1.059     |
|  | 1  | No borrowing| 0.000| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.147| 0.002    | 0.024| 0.415     |
|  |  | Dynamic + IPW| −0.001| 0.005    | 0.005| 0.994     |
|  |  | Dynamic     | −0.006| 0.005    | 0.005| 1.047     |
|  | 10 | No borrowing| 0.002| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| 0.000| 0.002    | 0.002| 0.486     |
|  |  | Dynamic + IPW| 0.001| 0.004    | 0.004| 0.801     |
|  |  | Dynamic     | 0.001| 0.004    | 0.004| 0.765     |
|  | 0.3| No borrowing| −0.001| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.091| 0.002    | 0.010| 0.458     |
|  |  | Dynamic + IPW| −0.001| 0.004    | 0.004| 0.888     |
|  |  | Dynamic     | −0.013| 0.005    | 0.005| 1.051     |
|  | 0.6| No borrowing| −0.001| 0.005    | 0.005| 1.000     |
|  |  | Full borrowing| −0.160| 0.002    | 0.028| 0.377     |
|  |  | Dynamic + IPW| −0.002| 0.005    | 0.005| 0.988     |
|  |  | Dynamic     | −0.006| 0.005    | 0.005| 1.041     |

Table 3. Summary of fitted logistic PS model for probability of being in study AML0531, with raw and IPW weighted covariate differences (AML0531-AML03P1).

| Estimate      | Std.Error | z value | Pr(>|z|) | Raw Diff. | Weighted Diff. |
|---------------|-----------|---------|---------|-----------|---------------|
| CNSdisease    | 0.189     | 0.693   | 0.273   | −0.042    | −0.011        |
| Race          | −0.783    | 0.453   | −1.731  | 0.084     | 0.036         |
| log_MRD       | 0.059     | 0.132   | 0.449   | 0.653     | −0.044        |
| log_age       | −0.082    | 0.153   | −0.538  | 0.591     | 0.057         |
| log_WBC       | −0.282    | 0.119   | −2.371  | 0.018     | 0.553         |
| low_risk      | −0.187    | 0.349   | −0.535  | 0.592     | −0.035        |
| high_risk     | 1.199     | 0.693   | 1.728   | 0.084     | 0.048         |

Direct adjustment using outcome models is another approach that we have not mentioned, but can also be used together with the algorithm we proposed. With a fitted outcome model \( m_0(X, \hat{\beta}) \) to the internal control data with BB weights, one can use the predicted mean \( \hat{y}_h = \sum_{i=1}^{n} H_i \xi^* m_0(X_i, \hat{\beta}) / n_h \) in Section 3. This approach can be considered as a special case of the DR approach described above. We mention it separately, as it can also be used with a standard Bayesian method using a (generalized) linear model (Ibrahim et al. 2015). While the standard Bayesian approach may be more accurate, our approach relies on less assumptions, is more flexible and easy to use.
muc=mean(yc*wi)
muh=mean(yh/odd)
sigc=wtd.var(yc,wi)/nc
sigh=wtd.var(yh,odd)/nh
aps=sigh/\max((muc-muh)^2, sigc+sigh)-sigc)
sig0=1/(1/sigc+ a0/sigh)
mups=rbind(Bout,c(b,a0,aps,muc,muf,mu0, mups))

library(ggplot2) ss=dim(Qout)[1]/4 bv=unique(Qout[,1])
Estimates=as.numeric(Qout[,4:7]) Estimator=rep(c("No borrowing","Full borrowing","Dynamic","IPW+Dynamic"), rep(ss*4,4)) Estimator=factor(Estimator,levels=c("Dynamic","IPW+Dynamic","No borrowing","Full borrowing")) vB=paste("b=",rep(Qout[,1],4)) Pout=data.frame(Estimates,b=vB,Estimator) ggplot(data=Pout,aes(x=Estimator, y=Estimates)) + geom_boxplot() +facet_wrap(\~b)

########################################################################### AML example

aml2 <- aml2[complete.cases(aml2),]

# create binary variables
aml2$'CNS disease' <- ifelse(aml2$'CNS disease'=="No", 0, 1) aml2$'Race' <- ifelse(aml2$'Race'=="Black or African American", 1, 0)

modelmat=as.matrix(aml2[-c(1:3,6,8,16,17)])
nn=table(aml2$curr) nc=nn[1] nh=nn[2] nbb=1000
Out=matrix(ncol=4,nrow=nbb)
for(BB in 1:nbb){
  wi=rexp(nc)
  vi=rexp(nh)
  wi=wi/mean(wi)
  vi=vi/mean(vi)
  yc=aml2$cr1[aml2$curr==0]
  yh=aml2$cr1[aml2$curr==1]
  curr=c(rep(0,nc),rep(1,nh))
  modelmat2=rbind(modelmat[,aml2$curr==0], modelmat[,aml2$curr==1])
  muc=sum(yc*wi)
muh=sum(yh/odd)
  ll=lbeta(va0*muh+muc+1,va0*(nh-muh)+nc
  -muc+1)-lbeta(va0*muh+1,va0*(nh-muh)+1)
  a00=max(va0[ll==max(ll)])
  mu0=(a00*muh+muc+1)/(nc+a00*nh+2)
  fitps=glm(curr~modelmat2,family = 'binomial', weights=c(wi,vi))
  ps=predict(fitps,type="response")[curr==1]
  odd=(1-ps)/ps*vi
  odd=odd/mean(odd)
  muc=mean(wi)
muh=mean(vi)
  ll=lbeta(va0*muh+muc+1,va0*(nh-muh)+nc
  -muc+1)-lbeta(va0*muh+1,va0*(nh-muh)+1)
  aps=max(va0[ll==max(ll)])
  mups=(aps*muh+muc+1)/(nc+aps*nh+2)
  Out=rbind(Out,c(mu0,mups,muf,aps,muc))
}
apply(Out,2,mean)
apply(Out,2,var)

plot(density(Out[,1],from=0.8,to=1), ylim=c(0,25),main="",xlab="CR rate at course 2",lwd=1.5)
lines(density(Out[,2],from=0.8,to=1), lty=2,lwd=1.5)
Acknowledgments

The authors would like to thank the associate editor and two referees for detailed comments and suggestions, which led to substantial improvement of the manuscript.

Disclosure Statement

The authors confirm that there are no competing interests to declare.

ORCID

Jixian Wang http://orcid.org/0000-0001-9963-6022

References

Capistrano, E. S. M., Moodie, E. E. M., and Schmidt, A. M. (2019), “Bayesian Estimation of the Average Treatment Effect on the Treated Using Inverse Weighting,” *Statistics in Medicine*, 38, 2447–2466. [183,184,185]

Chen, M. H., Ibrahim, J. G., Lam, P., Yu, A., and Zhang, Y. (2011), “Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data,” *Biometrics*, 67, 1163–1170. [182,183]

Cooper, T. M., Franklin, J., Gerbing, R. B., Alonzo, T. A., Hurwitz, C., Raimondi, S. C., Hirsch, B., Smith, F. O., Mathew, P., Arceci, R. J., Feusner, J., Iannone, R., Lavey, R. S., Meshinchi, S., and Gamis, A. (2012), “AAML0531, A Pilot Study of the Safety of Gemtuzumab Ozogamicin in Combination with Chemotherapy for Newly Diagnosed Childhood Acute Myeloid Leukemia: A Report from the Children’s Oncology Group,” *Cancer*, 118, 761–769. [187]

De Finetti, B. (1974), *Theory of Probability: A Critical Introductory Treatment*, trans. by A. Machi and J. Adrian Smith, Chichester: Wiley. [183]

European Medicines Agency (2020), “Guideline on Registry-based Studies.” [182]

Food and Drug Administration (2018), “Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices.” [182]

Gamis, A. S., Alonzo, T. A., Meshinchi, S., Sung, L., Gerbing, R. B., Raimondi, S. C., Hirsch, B. A., Kahwashi, S. B., Heerema-McKenney, A., Winter, L., Glick, K., Davies, S. M., Byron, P., Smith, F. O., and Aplenc, R. (2014), “Gemtuzumab ozogamicin in Children and Adolescents with De Novo Acute Myeloid Leukemia Improves Event-Free Survival by Reducing Relapse Risk: Results from the Randomized Phase III Children's Oncology Group Trial AAML0531,” *Journal of Clinical Oncology*, 32, 3021–3032. [187]

Graham, D. J., McCoy, E. J., and Stephens, D. A. (2016), “Approximate Bayesian Inference for Doubly Robust Estimation,” *Bayesian Analysis*, 11, 47–69. DOI:10.1214/14-BA928. [183,188]

Gravestock, I., and Held, L. (2017), “Adaptive Power Priors with Empirical Bayes for Clinical Trials,” *Pharmaceutical Statistics*, 16(5), 349–360. DOI:10.1002/pst.1814. [182,183,184,185]

Gravestock, I., and Held, L. (2018), “Power Priors based on Multiple Historical Studies for Binary Outcomes,” *Biometrical Journal*, 61, 1201–1218. [182,183,184,188]

Harrell, F. E. (2021), “Package Hmisc,” Available at https://github.com/cran/Hmisc. [185]

Hobbs, B. P., Carlin B. P., Mandrekar S. I., and Sargent D. J. (2011), “Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials,” *Biometrics*, 67, 1047–1056. [182,183]

Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013), “Adaptive Adjustment of the Randomization Ratio Using Historical Control Data,” *Clinical Trials*, 10, 430–440. [182,183]

Ibrahim, J. G., and Chen M.-H. (2000), “Power Prior Distributions for Regression Models,” *Statistical Science*, 15, 46–60. [182,183]

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2003), “On Optimality Properties of the Power Prior,” *Journal of the American Statistical Association*, 98, 204–213. [182,183]

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015), “The Power Prior: Theory and Applications,” *Statistics in Medicine*, 34, 3724–3749. [8] [189]

Katalik, A., Vock, D. M., Donny, E. C., Hatsukami, D. K., and Koopeiners, J. S. (2021), “Dynamic Borrowing in the Presence of Treatment Effect Heterogeneity,” *Biostatistics*, 22, 789–804. [182,183]

McCandless, L. C., Douglas, I. J., Evans, S. I., and Smeeth, L. (2010), “Cutting Feedback in Bayesian Regression Adjustment for the Propensity Score,” *The International Journal of Biostatistics*, 6, 1–24. [183]

McCandless, L. C., Gustafson, P., and Austin, P. C. (2009), “Bayesian Propensity Score Analysis for Observational Data,” *Statistics in Medicine*, 28, 94–112. [183]

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009), “A Note on the Power Prior,” *Statistics in Medicine*, 28, 3562–3566. DOI:10.1002/sim.3722. [182]

Newton, M. A., and Raftery, A. E. (1994), “Approximate Bayesian Inference with the Weighted Likelihood Bootstrap,” *Journal of the Royal Statistical Society*, Series B, 56, 3–48. [183]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), “Estimation of Regression Coefficients When Some Regressors Are Not Always Observed,” *Journal of the American Statistical Association*, 89, 846–866. [182,183]

Rosenbaum, P. R., and Rubin, D. B. (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55. [182,183]

Rubin, D. B. (1981), “The Bayesian Bootstrap,” *Annals of Statistics*, 9, 130–134. [183,184]

Saarela, O., Belzile, L. R., and Stephens, D. A. (2016), “A Bayesian View of Doubly Robust Causal Inference,” *Biometrika*, 103, 667–681. [183,184,185]

Sachdeva, A., Tiwari, R. C., and Guha, S. (2021), “A Novel Approach to Augment Single-Arm Clinical Studies with Real-World Data,” *Journal of Biopharmaceutical Statistics*, 27, 1–17. DOI:10.1080/10543406.2021.2011902. [182]

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge: Cambridge University Press. [185]

Wang, C., and Rosner, G. L. (2019), “A Bayesian Nonparametric Causal Inference Model for Synthesizing Randomized Clinical Trial and Real-World Evidence,” *Statistics in Medicine*, 38, 2573–2588. [182,183,184]

Wang, C., Li, H., Chen, W.-C., Lu, N., Tiwari, R., Xu, Y., and Yue, L. (2019), “Propensity Score- Integrated Power Prior Approach for Incorporating Real-World Evidence in Single-Arm Clinical Studies,” *Journal of Biopharmaceutical Statistics*, 29, 1–18. DOI:10.1080/10543406.2019.1657133. [182,183,184]

Zigler, C. M. (2016), “The Central Role of Bayes’ Theorem for Joint Estimation of causal Effects and Propensity Scores,” *The American Statistician*, 70, 47–54. [183]

Zigler, C. M., Watts, K., Yeh, R. W., Wang, Y., Coull, B. A., and Dominici, F. (2013), “Model Feedback in Bayesian Propensity Score Estimation,” *Biometrics*, 69, 263–273. [183]