



Augmenting Both Arms of a Randomized Controlled Trial Using External Data: An Application of the Propensity Score-Integrated Approaches

Heng Li¹ · Wei-Chen Chen¹ · Chenguang Wang² · Nelson Lu¹ · Changhong Song¹ · Ram Tiwari¹ · Yunling Xu¹ · Lilly Q. Yue¹

Received: 24 February 2021 / Revised: 12 May 2021 / Accepted: 7 June 2021 /
Published online: 19 June 2021

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Abstract

Leveraging external data is a topic that have recently received much attention. The propensity score-integrated approaches are a methodological innovation for this purpose. In this paper we adapt these approaches, originally introduced to augment single-arm studies with external data, for the augmentation of both arms of a randomized controlled trial (RCT) with external data. After recapitulating the basic ideas, we provide a step-by-step tutorial of how to implement the propensity score-integrated approaches, from study design to outcome analysis, in the RCT setting in such a way that the study integrity and objectivity are maintained. Both the Bayesian (power prior) approach and the frequentist (composite likelihood) approach are included. Some extensions and variations of these approaches are also outlined at the end of this paper.

Keywords Power prior · Composite likelihood · Outcome-free design

1 Introduction

Augmenting a prospective clinical study with external data in order to reduce the study length and cost is a topic that is receiving increasing attention, since it aligns with FDA CDRH vision that patients in the US have access to high-quality, safe and effective medical devices of public health importance first in the world

✉ Lilly Q. Yue
Lilly.Yue@fda.hhs.gov

¹ Division of Biostatistics, Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA

² Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA

[1]. Recent methodological advancements for such leveraging of external data in a single arm study, under the name of propensity score-integrated approaches, include Wang et al. [2, 3]. In the current paper we extend these approaches so that they can be used to augment both arms of an RCT. As a motivating example, let us consider the following scenario. A medical device company plans to conduct an RCT to evaluate the safety and effectiveness of a device in order to seek approval for its marketing in the US, with control therapy being optimal medical management, and primary endpoint being a one-year adverse event rate. Based on the projected enrollment rate, it is anticipated that enrollment will take five years. With an additional one year of follow-up, it is going to take six years for the trial to complete. The company wants to explore the possibility of leveraging external data to speed up the trial, so that this new technology can reach patients sooner. The company thinks that such a study design may be viable because a significant amount of clinical data has already been accumulated in the EU, where the device was CE marked, and a high-quality registry has been established in the EU for patients treated with the device. This registry can potentially be used as a source of external data that can be leveraged to augment the treated arm of the RCT. To augment the control arm, a high-quality disease registry in the US may be used. A proposal is thus put forward to conduct a study consisting of an RCT with both arms supplemented by data from these two external sources, respectively. The objective of this paper is to describe an appropriate statistical procedure that can be used to design the study being proposed, and subsequently analyze the study data.

In general, there are two statistical considerations when using external data to augment a prospective clinical study (to be called the current study): (1) the similarity, in terms of baseline covariates, between patients from the external data source and patients from the current study, and (2) the amount of information that the external data contribute to the statistical inference. If the external patients and the current study patients are not similar, then it may be difficult to justify incorporating the external patients into the current study. Since the external data source, especially when it is a real-world data (RWD) source, may contribute much more patients than the current study does, we need the option of down-weighting the external patients so that they do not dominate study results if we decide to leverage them. These two considerations are addressed in Wang et al. [2, 3] in the statistical procedures they developed for leveraging external data in single arm studies. In these procedures, the propensity score methodology is utilized to deal with the issue of patient similarity, by forming strata of patients according to their propensity scores and then leveraging the external patients within each propensity score stratum. The rationale for doing this is that within each propensity score stratum the external patients and the current study patients tend to be more similar in terms of baseline covariates than they are overall, which makes within-stratum leveraging of the external data more justified. The issue of down-weighting external patients can be handled in at least two ways. In the frequentist framework, one can use the methodology of composite likelihood, as is the case with the propensity score-integrated composite likelihood approach (PSCL) developed in Wang et al. [3]. In the Bayesian framework, the technique of

power prior can be utilized, as is the case with the propensity score-integrated power prior approach (PSPP) developed in Wang et al. [2].

To leverage external data to augment both arms of an RCT, PSPP and PSCL can be applied to each arm of the RCT (the current study) as if it were a single-arm study, and the results of the outcome analysis can then be combined. More specifically, suppose the current study randomizes patients between treatment A and treatment B, and let the parameter of interest be the treatment effect $\mu = \theta_A - \theta_B$. Our strategy is to first apply PSCL (or PSPP) to treatment A and obtain a point estimate and its standard error (or a posterior distribution) for θ_A , then do the same for θ_B , and finally conduct statistical inference on $\theta_A - \theta_B$ based on the independence between the two point estimates (or posterior distributions). This paper is intended not only to detail the above methodological extension of PSCL and PSPP, but also to serve as a user guide for its proper implementation in practice. As such, it can be a handy reference for the planning of an RCT for which augmenting both arms with external data is a viable option, and for the subsequent design, conduct, and analysis of such an RCT if this option is taken. We believe that the implementation of our strategy can be best described by means of an illustrative example, which is provided in Sect. 3. In Sect. 2 we review the basics of PSPP and PSCL. Section 4 concludes with some discussion.

1.1 Propensity Score-Integrated Approaches

The propensity score-integrated approaches are essentially statistical procedures that make borrowing external patients more justified, by using propensity score to form strata in such a way that within each stratum external patients are similar with those in the current study and carrying out the borrowing within the propensity score strata (in a sense that will become clear by the end of the hypothetical example in Sect. 3). The rationale is that borrowing would be more reasonable if the external patients being borrowed are similar with the corresponding patients in the current study. Here, “similar” means alike in terms of observed covariates. Thus, two groups of patients are similar if the distribution of observed covariates in one group is close to that in the other, in which case we say that the two groups are *balanced* in these covariates. Therefore, what propensity score-integrated approaches do can also be summarized as “borrowing after balancing.” To delineate how this works, let us first review the original definition of propensity score for observational studies [4, 5], and then adapt this definition to suit our goal. In an observational study, the propensity score $e(X)$ for a patient with a vector X of observed baseline covariates is the conditional probability of being in the treated group ($T=1$) rather than the control group ($T=0$) given the vector of baseline covariates X :

$$e(X) = \Pr(T = 1|X) \tag{1}$$

Propensity score (PS) is a balancing score in the sense that conditional on the PS, the distribution of observed baseline covariates is the same between the treated and control patients. This implies that when the PSs are balanced across the two treatment groups, all the observed covariates are balanced in expectation across the

two groups. In practice, patients' PSs are estimated by modeling the probability of treatment group membership T as a function of the observed covariates, typically via logistic regression. Estimated PSs can then be used for matching, weighting, or stratification, in order to balance the treated group and the control group to reduce bias in the statistical inference for treatment effects.

Since our objective is to create strata within which the observed covariates are balanced between the external patients and the current study patients, PS is defined accordingly as the conditional probability of a patient coming from the external data source rather than the current study given the value of covariates. If the current study is a single-arm study, then only one such PS is needed (see Wang et al. [2, 3]). Since in this paper the statistical procedures in Wang et al. [2, 3] will be applied to each arm of an RCT randomizing between treatment A and treatment B, we need to define two PSs, one for treatment A and one for treatment B. Specifically, for patients receiving treatment g ($g = A$ or B), let those from the current study be labeled $Z^{(g)} = 0$ and patients from the external data source be labeled $Z^{(g)} = 1$, we define PS for treatment g as

$$e^{(g)}(X) = \Pr(Z^{(g)} = 1|X) \quad (2)$$

where X is the vector of observed covariates. The two PSs so defined are both balancing scores, which means that, for patients receiving treatment g ($g = A$ or B), when the PS $e^{(g)}(X)$ is balanced between the external patients and the current study patients, all the observed covariates are balanced in expectation between the current study patients and the external patients. To take advantage of this balancing property one can, separately for treatment A and treatment B, form strata in which the estimated PSs are relatively homogeneous, so that within each PS stratum the distribution of observed covariates among the external patients is close to that among patients in the current study, and balance for all covariates within each stratum is therefore expected. In practice, balance is assessed, separately for treatment A and treatment B, between the external patients and the current study patients for each covariate, and, if not satisfactory for some covariates then PSs may be re-estimated by modifying the corresponding PS model. This makes the entire process of PS estimation (including modeling), PS stratification, and balance assessment, which is called PS design, an iterative one.

Considering its iterative nature, PS design needs to be outcome-free, i.e., performed with no outcome data in sight, in order to ensure the integrity of study design and interpretability of the subsequent statistical analysis. This is because, as has been mentioned, the goal of PS design is to adequately balance covariates, and, to improve balance, the PS model may be modified multiple times. In other words, PS models are not pre-specified. This raises the question of how to maintain study objectivity when outcome data such as those from external sources already exist, which presents an opportunity for data dredging. Outcome-free design is essentially a matter of blinding or masking, which can also be referred to as building a fire-wall in the biopharmaceutical arena. Various schemes have been devised for this purpose. The scheme that we propose is for the investigator of the study to identify an independent statistician to perform the PS design to whom no outcome data are

provided. The independent statistician shares with the investigator the responsibility of upholding the outcome-free principle [6–11]. In practice, this blinding scheme is embedded in a design paradigm called “two-stage design,” first introduced in Yue et al. [10]. To carry out PSCL or PSPP, the two-stage design is to be implemented, and a large part of Sect. 3 is to describe how to do it through an example.

Having explained how balance can be achieved, let us now introduce two ways to borrow external patients while down-weighting them, one Bayesian, using power prior, and the other frequentist, using composite likelihood. Down-weighting is needed when the number of patients available from the external source is larger than the number of patients needed to be borrowed for the current study (which is usually the case), and we want to limit the influence these external patients have on the study results.

The power prior [12] is originally intended to be an informative prior constructed from historical data [13]. If we substitute external data for historical data, the method fits our purpose perfectly. A power prior π for a parameter θ based on external data D_0 is constructed as follows:

$$\pi(\theta) \propto [L(\theta|D_0)]^\alpha \pi_0(\theta) \tag{3}$$

where $L(\theta|D_0)$ is the likelihood function of θ given the external data, $\pi_0(\theta)$ is the initial prior distribution for θ , and α ($0 \leq \alpha \leq 1$) is called the power parameter. This prior is multiplied to the likelihood function of θ given the current study data D_1 , $L(\theta|D_1)$, to obtain the posterior distribution of θ ,

$$\pi(\theta|D_1) \propto [L(\theta|D_1)]\pi(\theta), \tag{4}$$

which is then used for the statistical inference of θ . From this construction, α can evidently be interpreted as the fraction of information external patients contribute to the inference for θ . For example, if $\alpha = 0.1$, each external patient contributes 10% of their information, and the total amount of information the external patients bring to the statistical inference is equivalent to the information contributed by 0.1 times the total number of external patients, which can be interpreted as the (nominal) number of patients being borrowed for some common distributions such as normal and binomial. If $\alpha = 1$ then the number of patients borrowed is equal to the number of all the patients constituting D_0 . At the other extreme, if $\alpha = 0$ then no patients are borrowed.

The composite likelihood [14] for the parameter of interest θ is a weighted product of probability density functions:

$$L(\theta|Y) = \prod_i f(y_i|\theta)^{\lambda_i} \tag{5}$$

where each i represents a patient and λ_i is a nonnegative weight. Clearly, when all the λ_i 's equal to 1, composite likelihood reduces to ordinary likelihood. To use composite likelihood to serve our purpose, we let $\lambda_i = 1$ for patients from the current study and $0 < \lambda_i \leq 1$ for patients from the external data source. If statistical inference for θ is conducted based on the composite likelihood after giving λ_i 's numerical values in this way, then we are essentially down-weighting the external patients relative to the current study patients. For example, if $\lambda_i = 0.1$ for all external patients, then

each external patient contributes roughly 10% of their information, and the (nominal) number of patients borrowed is 0.1 times the total number of external patients.

Having provided all the elements of PS-integrated approaches (PSCL and PSPP), we next demonstrate how they can be implemented to augment both arms of an RCT with external data.

2 Implementation: An Illustrative Example

Consider the scenario described in the introduction section, where a medical device company is planning to conduct an RCT, and, in order to shorten the duration of the study, intends to augment both arms of the RCT with registry data. The investigational device arm (A) is augmented by patients extracted from a high-quality device registry in Europe (where the investigational device has been CE marked and therefore commercially available), and the optimal medical management arm (B) is augmented by patients extracted from a high-quality disease registry in the US. For the purpose of demonstration, both PSCL and PSPP are presented. The study design part, however, is identical for these two methods.

As mentioned in Sect. 2, the two-stage design [10] is to be adopted to carry out PSCL or PSPP. For our example, the main elements of the first stage of the two-stage design are summarized in Table 1. After the primary endpoint (one-year adverse event rate) and the study hypotheses (superiority) are agreed upon, sample size is calculated. Assuming an adverse event rate of 29% for the control arm (B) and 20% for the investigational device arm (A), under a one-sided significance level of 0.025, 500 patients per arm is needed to demonstrate superiority of A over B with 90% power. Suppose it is decided based on clinical and regulatory considerations that the nominal number of external patients to be borrowed is 100 per arm, so that 400 patients per arm are to be prospectively enrolled into the current study (RCT). Here, the word “nominal” means that the amount of information contributed by the external patients is equal to that of 100 patients, not to imply that 100 “actual” external patients will be selected to augment the current study. Since PS-integrated approaches is used to borrow external data, all the baseline covariates to be balanced (included in the PS model) need to be specified, based on clinical judgment, and the

Table 1 Main elements of the first design stage

Primary outcome: probability of adverse event within 1 year (θ)
Hypotheses: $H_0: \theta_A - \theta_B \geq 0$, $H_a: \theta_A - \theta_B < 0$
Significance level: 0.025 one-sided / posterior probability threshold: 0.975
Baseline covariates specified for PS model (17)
PSCL/PSPP planned
Independent statistician identified
Sample size for the current study: 400 per arm
Nominal number of external patients: 100 per arm

Table 2 Sample size in each PS stratum for arm A

	PS stratum					Total
	1	2	3	4	5	
Current study	80	80	80	80	80	400
EU registry	269	218	160	196	98	941

Table 3 Sample size in each PS stratum for arm B

	PS stratum					Total
	1	2	3	4	5	
Current study	80	80	80	80	80	400
US registry	320	279	253	210	130	1192

independent statistician who performs PS design needs to be identified. The first design stage concludes when all the above elements are completed.

The second design stage consists of PS design, and the determination of the power parameter α or the exponent λ for each PS stratum (recall that borrowing is to be carried out within strata), depending on whether Bayesian or frequentist inference is used. We will only illustrate how to accomplish these tasks for the investigational device arm (A), because the exact same steps are taken for the control arm (B). After the enrollment of all 800 patients into the current study (with 400 in arm A and 400 in arm B) and the collection of all registry patients meeting the inclusion exclusion criteria, the PS $e^{(A)}(X)$, as defined in (2), is estimated for each patient in the investigational device arm (A) of the current study and from the European device registry. Then 941 external patients for the European device registry are selected by excluding those external patients whose PSs are not in the range of that of the patients in the current study. All patients (400 + 941) are grouped into 5 PS strata in such a way that the same number of current study patients ($80 = 400/5$) are in each PS stratum (i.e., using propensity quintiles among the 400 current study patients as boundaries of the strata). This guarantees that each stratum contains current study patients. Since within each PS stratum the current study patients and external patients are expected to be more similar than they are overall, the borrowing of external patients within stratum is more justified. The number of European registry patients and current study patients in the investigational device arm in each PS stratum are displayed in Table 2. The corresponding numbers for the control arm, after the PS stratification using $e^{(B)}(X)$ is conducted for patients in the control arm of the current study and in the US registry, are displayed in Table 3. Although the number of PS strata is five for both treated and the control patients in this example, in practice this number may be different from five and may be different between the two arms for the purpose of covariate balance. For the covariate balancing assessment, readers are referred to [15].

Recall that it was decided based on clinical and regulatory considerations that the total amount of information to be borrowed is equivalent to 100 external patients.

Since borrowing takes place within each stratum, we need to figure out how to allocate the 100 patients to the 5 PS strata.

There are many possible ways to do so. One may allocate the nominal number of patients to be borrowed evenly to each stratum, so that each stratum gets $100/5 = 20$. Our strategy is to make the nominal number of patients to be borrowed in each stratum proportional to the similarity of external patients and the current study patients in terms of baseline covariates in that stratum. And we propose to measure this similarity by an overlapping coefficient [16], the overlapping area of propensity score distributions of the two groups of patients (you may use other reasonable measures). The overlapping coefficients are then standardized so that they add up to 1. The standardized overlapping coefficients time the total nominal number of patients to be borrowed (100) determine the nominal number of patients to be borrowed in each stratum. In this example, the nominal number of patients allocated to each stratum for the investigational device arm using our strategy is close to that using equal allocation, as shown in Table 4.

Once the nominal number of patients to be borrowed in each stratum is obtained, the power parameter α for the Bayesian approach or the exponent λ for the composite likelihood in the frequentist approach in each PS stratum can be calculated by dividing the nominal number of external patients to be borrowed by the total number of external patients in that stratum. The overlapping coefficient, the standardized overlapping coefficient, the nominal number of patients to be borrowed, and the power parameter (or exponent) in each stratum for the investigational device arm (A) are presented in Table 4. The same steps are applied to the control arm (B) and the corresponding numbers are displayed in Table 5. All the above design activities are performed by an independent statistician who is blinded to the outcome data. This completes the second stage of the two-stage design.

After clinical outcomes are observed on all the patients, statistical inference can be conducted. For the Bayesian approach, first use the power prior methodology within each PS stratum to get stratum-specific posterior distributions for the parameter of interest for treatment A based on information in Table 4, which are then combined to obtain the posterior distribution for θ_A , then apply the same steps to obtain the posterior distribution for θ_B based on information in Table 5,

Table 4 Overlapping coefficient, standardized overlapping coefficient, nominal number of patients to be borrowed, and power parameter (or exponent) in each stratum for arm A

	PS stratum					Total
	1	2	3	4	5	
Current study (n)	80	80	80	80	80	400
EU registry (n)	269	218	160	196	98	941
Overlapping coefficient	0.79	0.79	0.83	0.84	0.90	
Standardized overlapping coefficient (%)	19	19	20	20	22	100
Patients borrowed (= $100 \times$ Std. Overlap. Coef.)	19	19	20	20	22	100
α (or λ) [= Patients Borrowed/EU registry (n)]	0.07	0.09	0.13	0.10	0.22	

Table 5 Overlapping coefficient, standardized overlapping coefficient, nominal number of patients to be borrowed, and power parameter (or exponent) in each stratum for arm B

	PS stratum					Total
	1	2	3	4	5	
Current study (<i>n</i>)	80	80	80	80	80	400
US registry (<i>n</i>)	320	279	253	210	130	1192
Overlapping coefficient	0.81	0.79	0.79	0.84	0.77	
Standardized overlapping coefficient (%)	20	20	20	21	19	100
Patients borrowed (= 100 × Std. Overlap. Coef.)	20	20	20	21	19	100
α (or λ) [=Patients Borrowed/US Registry (<i>n</i>)]	0.06	0.07	0.08	0.10	0.15	

and finally obtain the posterior distribution for $\theta_A - \theta_B$. Note that the flat prior Beta(1, 1) is set as the initial prior $\pi_0(\theta)$ for each arm. In this example, the posterior probability of $\theta_A - \theta_B < 0$ is greater than 98.0% > 97.5%, which meets the study success criterion. For the frequentist approach, first construct the composite likelihood to get stratum-specific maximum likelihood estimates and standard errors for the parameter of interest for treatment A (Table 4), which are then combined to obtain the point estimate and standard error for θ_A , then apply the same steps to obtain the point estimate and standard error for θ_B (Table 5), and finally obtain the point estimate and standard error for $\theta_A - \theta_B$. In this example the maximum likelihood estimate of $\theta_A - \theta_B$ is -5.7%, with a one-sided *p*-value = 0.013 < 0.025, which meets the study success criterion. In this example, the combination of stratum-specific estimates (or posterior distributions) across the PS strata consists simply in taking the average, because each stratum has equal number of current study patients and usually the target patient population is represented by the current study. The stratum-specific and overall point estimates (for PSCL) and posterior means (for PSPP) are displayed in Table 6 for treatment A and B. For more details on how to obtain the point estimate and standard error (or posterior distribution) for θ_A and θ_B , see Wang et al. [3] (or Wang et al. [2]). Note that the discount parameters are the same for both PSPP and PSCL approaches, similar results are expected in the outcome analysis when the PSPP approach with flat priors is applied. Both PSPP and PSCL approaches are implemented in the R package “psrwe” [17].

Table 6 Point estimates of adverse event rates and treatment effect based on PSCL and PSPP

Method	Arm	PS stratum					Overall	Treatment effect
		1	2	3	4	5		
PSCL	A	0.315	0.255	0.193	0.232	0.254	0.250	-0.0574
	B	0.426	0.368	0.242	0.244	0.257	0.307	
PSPP	A	0.318	0.260	0.200	0.238	0.259	0.255	-0.0569
	B	0.429	0.370	0.248	0.249	0.263	0.312	

3 Discussion

In this paper we described how PS-integrated approaches can be used to augment both arms of a two-arm RCT with external data, possibly from two different sources, with respect to study design and outcome analysis. It is straightforward to adapt this procedure for the augmentation of L arms with external data in a K -arm RCT for any $L \leq K$. All we need to do is to define L propensity scores for the L arms that borrow external data, possibly from L different sources, and then follow essentially the same steps as in the previous section to estimate the parameters of interest for these L arms. The parameters of interest for the arms that do not borrow external data can be estimated in the usual way. Statistical inference for any comparison between the K arms may then be conducted. Chen et al. [18] describes a procedure for the special case of $K=2$ and $L=1$, but their method does not generalize to arbitrary K and L . Another extension is to the scenario where there are multiple external data sources to borrow from for each arm. A simplistic way to do this is to pool data from all the external sources and proceed as if they were from a single source. More elaborate approaches are discussed elsewhere.

For the PSPP approach, the power parameters are treated as constants in this paper. Alternatively, they can be treated as random variables in what is called a full Bayesian strategy [2]. For more details about the full Bayesian strategy see Wang et al. [2].

In our illustrative example, the borrowing of external data is planned before prospectively enrolling any patients. However, the same statistical procedure can be applied where the borrowing is unplanned. Suppose a prospective RCT with a treated and a control arm is actively enrolling patients, and an unforeseen event, such as the COVID-19 pandemic, occurs which stops the enrollment before it is completed. While sometimes the enrollment may restart later, this is not always possible or practical. Using the current data to test study hypotheses would result in an underpowered study as the planned sample size is not reached. In such a case, PS-integrated approaches can be used to augment one or both arms of the RCT with external data (if it is appropriate to do so) to recover the lost power, thereby salvaging the study.

The above are just some examples of extensions, variations, and areas of application of the PS-integrated approaches. We believe that, due to their flexibility and adaptability, the PS-integrated approaches are a viable statistical innovation that may be utilized in a variety of situations as a tool for the leveraging of external data to further support regulatory purposes.

Declarations

Conflict of interest No potential conflicts of interest were disclosed.

References

1. FDA CDRH, "CDRH Mission, Vision and Shared Values.," 1 February 2017. [Online]. Available: <https://www.fda.gov/about-fda/center-devices-and-radiological-health/cdrh-mission-vision-and-shared-values>. Accessed 2020.

2. Wang C, Li H, Chen W-C, Lu N, Tiwari R, Xu Y, Yue LQ (2019) Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat* 29(5):731–748
3. Wang C, Lu N, Chen W-C, Li H, Tiwari R, Xu Y, Yue LQ (2020) Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat* 30(3):495–507
4. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55
5. Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 79(387):516–524
6. Li H, Mukhi V, Lu N, Xu Y-L, Yue LQ (2016) A note on good practice of objective propensity score design for premarket nonrandomized medical device studies with an example. *Stat Biopharm Res* 8(3):282–286
7. Lu N, Xu Y, Yue LQ (2019) Good statistical practice in utilizing real-world data in a comparative study for premarket evaluation of medical devices. *J Biopharm Stat* 29(4):580–591
8. Lu N, Xu Y, Yue LQ (2020) Some considerations on design and analysis plan on a nonrandomized comparative study using propensity score methodology for medical device premarket evaluation. *Stat Biopharm Res* 12(2):155–163
9. Xu Y, Lu N, Yue LQ, Tiwari R (2020) A study design for augmenting the control group in a randomized controlled trial: a quality process for interaction among stakeholders. *Ther Innov Regul Sci* 54:269–274
10. Yue LQ, Lu N, Xu Y (2014) Designing premarket observational comparative studies using existing data as controls: challenges and opportunities. *J Biopharm Stat* 24(5):994–1010
11. Yue LQ, Campbell G, Lu N, Xu Y, Zuckerman B (2016) Utilizing national and international registries to enhance pre-market medical device regulatory evaluation. *J Biopharm Stat* 26(6):1136–1145
12. Chen M-H, Ibrahim JG (2000) Power prior distributions for regression models. *Stat Sci* 15(1):46–60
13. Ibrahim JG, Chen M, Gwon Y, Chen F (2015) The power prior: theory and applications. *Statist Med* 34:3724–3749
14. Varin C, Reid N (2011) An overview of composite likelihood methods. *Stat Sin* 21(1):5–42
15. Austin PC (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 28:3083–3107
16. Inman HF, Bradley EL Jr (1989) The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun Stat* 18(10):3851–3874
17. C. Wang, “psrwe: PS-Integrated Methods for Incorporating RWE in Clinical Studies,” R Package, 2020. [Online]. <https://CRAN.R-project.org/package=psrwe>.
18. Chen W-C, Wang C, Li H, Lu N, Tiwari R, Xu Y, Yue LQ (2020) Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *J Biopharm Stat* 30(3):508–520