Research Article
Volume 14:01, 2023

Journal of Biometrics & Biostatistics

ISSN: 2155-6180

Open Access

Statistical Evaluation of the Validity of Real World Data and Real World Evidence

Yuankang Zhao* and Shein-Chung Chow

Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, USA

Abstract

Real-world data (RWD) often consist of positive or negative studies and the data may be structured or unstructured. In this case, the validity of real-world evidence (RWE) that derived from RWD is a concern for providing substantial evidence regarding the safety and efficacy of the test treatment under investigation. The validity of RWD/RWE is essential, especially when it is intended to support regulatory submission. In practice, studies with positive results are more likely accepted in RWD, which may cause substantial selection bias. In this article, a quantitative form of selection bias is defined and studied. Based on the form of bias, three reproducibility probability-based approaches are proposed to estimate the true proportion of positive studies in the structural and unstructured data. The reproducibility probability-based approach provides effective bias adjustment when the proportion of significant studies in RWD is different as designed power based on the result of simulation study. The Estimated Power approach and Bayesian approach provide robust and effective bias adjustment in most cases and the Confidence Bound approach provide huge and effective adjustment only when bias is larger than 10%. The proposed adjustment method in conjunction with other treatment effect specification method is useful in estimating the treatment effect based on RWD.

Keywords: Real-world data (RWD) • Real-world evidence (RWE) • Pragmatic clinical trial • Selection bias • Reproducibility probability

Introduction

Real-world data (RWD) refers to the data relating to patient health status or the delivery of health care routinely collected from various sources (FDA) [1]. The source of RWD includes electronic health record (EHR), medical claims databases, products and diseases registries, data from randomized clinical trials, and so on. As indicated in the US Food and Drug Administration (FDA) draft guidance on Framework for FDA's Real-World Evidence Program, real-world evidence (RWE) refers to the clinical evidence about the usage and potential benefits or risks of a medical product derived from an analysis of RWD (FDA) [2]. In practice, RWE is often generated by different designs or investigations, such as pragmatic clinical trial (PCT) and prospective or retrospective observational studies.

Although randomized controlled trials (RCTs) are the gold standard for evaluating the safety and efficacy of pharmaceutical drugs, RCTs are conducted under specific or controlled environment, which do not reflect real clinical practice [3]. In practice, RCTs limit generalizability due to strict inclusion and exclusion criteria, as well as high costs and long duration, causes people to consider RWE as alternative clinical evidence [4]. Not only RWE can make up for the drawbacks, but it also can provide treatment effects evidence in more diverse applied settings due to the massive volume of data, as well as provide evidence of some rare disease drug development due to the data availability in multicenter trial and EHR. Therefore, the challenge faced by biostatisticians is how to generate robust RWE from RWD and integrate it into drug development and regulatory review; in other words, map RWE to Substantial Evidence [5].

In order to map RWE to substantial evidence (current regulatory standard and can only be obtained through the conduct of RCTs), researchers often focus on the following aspects. First, it is to determine the difference in evidence provided by RWD and data collected by RCTs (i.e., gap analysis between RWE and substantial evidence). Second, it is to evaluate whether the RWD is robust and representative of the target population (i.e., data relevancy and selection bias) [6]. In addition, data quality or data reliability is the most crucial part of evaluating RWD because of the volume and multiple sources nature of RWD. Finally, it is to assess whether historical data is suitable for data borrowing (e.g., the use of Bayesian inference) for efficient quantitative analysis in order to meet the current regulatory standard [7]. In this article, we will focus on the validity of RWD/RWE (i.e., the presence of selection bias and information bias).

As indicated by Chow SC [7], a mathematical model with statistical analysis was performed, which considers the biased positive studies proportion, providing a systematic way to evaluate the validity of RWD from the regulatory perspective [7]. The most recent guidance on RWD clearly indicates that the "relevant impacts of unstructured data on data quality should be documented in the protocol and analysis plan". As indicated in FDA (2021), unstructured data refer to the data within EHR, either as free text data fields (such as physician notes) or as other non-standardized information in computer documents (such as PDF-based radiology reports). All these data need further processing (such as the deep learning algorithm with a significant amount of human aid) to extract valid clinical information, which inevitably brings bias. In this article, we study the validity (in terms of selection bias) of unstructured data and divide datasets into four classifications as shown in (Figure 1). Therefore, we will give a new model with further statistical analysis and simulation, as well as apply our model in RWD setting to prove the valuable insight from regulatory perspectives. In the Section 2, the statistical method for studying the validity of real-world data is briefly outlined. This section includes the estimation of the selection bias using three proposed reproducibility probability-based approaches. In the Section 3, simulation studies were conducted to examine the robustness and efficiency of the proposed methods. In the Section 4, the limitations and relative advantages of the proposed methods are compared with other bias adjustment methods. Section 5 provides some concluding remarks.

Statistical methods

The validity of real world data: Let μ_{rwd} be the true mean of the target population study's target patient group, μ_s and μ_{ns} be the true means of data

***Address for Correspondence:** Yuankang Zhao, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, USA, E-mail: yuankang.zhao@duke.edu*

Received: 08 December 2022, Manuscript No. jbmbs-22-82892; **Editor assigned:** 09 December, 2022, Pre QC No. P-82892; **Reviewed:** 22 December 2022, QC No. Q-82892; **Revised:** 02 January 2023, Manuscript No. R-82892; **Published:** 09 January 2023,, DOI: 10.37421/2155-6180.2023.14.150

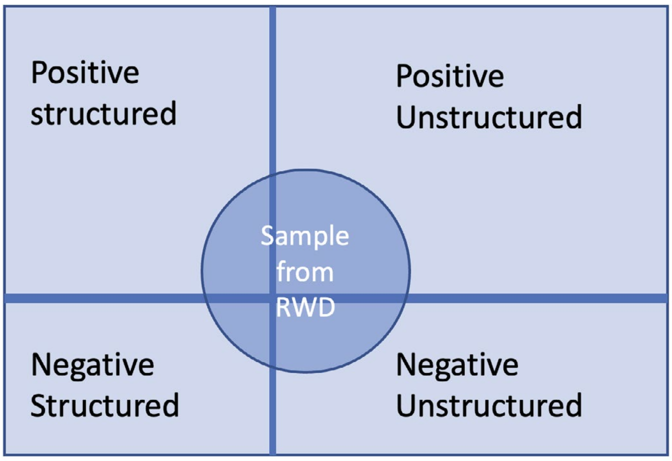

Figure 1. Model that consider unstructured data.

sets from positive and negative studies done in the same target population. Also, let μ_s and μ_{ns} be the true means of data sets of structured data and unstructured data. In addition, let ρ be the true proportion of structured data conducted on the target population, and let r_1 and r_2 be the true proportions of positive result within the structured data and unstructured data, respectively. Assume that there is no treatment-by-centre and treatment-by-study interaction in multicentre studies for illustrating. Following (Figure 1), we have

$$\mu_{rwd} = \rho\mu_s + (1-\rho)\mu_{ns}, \quad (1)$$

$$\mu_s = r_1\mu_P + (1-r_1)\mu_N, \quad (2)$$

$$\mu_{ns} = r_2\mu_P + (1-r_2)\mu_N, \quad (3)$$

Usually, $\mu_N < \delta < \mu_P$, where δ is the clinical meaningful treatment effect. In other words, the estimate effect of a study larger than δ would be considers a positive study. Besides, usually in practice, we would expect $\frac{1}{2} < r_i \le 1$. In particular, if $r_i = 1$, then μ would degrade to $\mu = \mu_P$.

Then, for illustration purpose, we assume that all the studies included in the RWD are parallel design studies, and all positive and no positive studies share same sample size, denoting $n_P\ and\ n_N . x_{ij}$ denote the clinical response of the patient i in the jth positive study, i=1,…,n_P and j=1,…,sn, where n is the number of studies included in the RWD. Besides, denote the clinical response of the patient i in the jth no positive study, i=1,…,n_N and j=1,…,(1-s)n. If ρ and r_i are given, the bias of mean of RWD, μ_{rwd} is

$$Bias(\mu_{rwd}) = E(\hat{\mu}_{rwd}) - \mu$$

$$= E(\rho\hat{\mu}_s + (1-\rho)\hat{\mu}_{ns}) - \mu$$

$$= E[\rho(r_1\hat{\mu}_P + (1-r_1)\hat{\mu}_N) + (1-\rho)(r_2\hat{\mu}_P + (1-r_2)\hat{\mu}_N)] - \mu$$

$$= [\rho r_1 + (1-\rho)r_2](E(\hat{\mu}_P) - \mu_P)) + [\rho(1-r_1) + (1-\rho)(1-r_2)](E(\hat{\mu}_N) - \mu_N) \quad (4)$$

Where

$$\hat{\mu}_P = \overline{x} = \frac{1}{snn_P}\sum_{j=1}^{sn}\sum_{i=1}^{n_P} x_{ij}$$

$$\hat{\mu}_N = \overline{y} = \frac{1}{(1-s)nn_N}\sum_{j=1}^{(1-s)\mathrm{n}}\sum_{i=1}^{n_N} y_{ij}$$

And

$$s = \rho r_1 + (1-\rho)r_2$$

Based on the equation 4, if we assign ρ and r_i certain value, we can gain the bias as shown in Table 1. From Table 1, we can find that the different ρ and r_i correspond different selection bias.

In practice, however, r_i are unknown. For a given RWD, estimation of r_i ($ie.\hat{\rho},$,i=1,2) is based on the number of positive studies in RWD. Similarly, the estimation of ρ (ie. $\hat{P},$) is based on the number of structural data in RWD. However, r_i usually be overestimated, since positive data are more likely published and included in historical data (S.C Chow, 2020). Therefore, we have

$$E(\hat{r}_i) = r_i + \Delta_i$$

Therefore, the bias of mean of RWD, μ_{rwd} is given by

$$Bias(\mu_{rwd}) = E(\hat{\mu}_{rwd}) - \mu = E(\hat{\rho}\hat{\mu}_s + (1-\hat{\rho})\hat{\mu}_{ns}) - \mu$$

$$= E[\hat{\rho}(\hat{r}_1\hat{\mu}_p + (1-\hat{r}_1)\hat{\mu}_N) + (1-\hat{\rho})(\hat{r}_2\hat{\mu}_p + (1-\hat{r}_2)\hat{\mu}_N)] - \mu$$

For simplicity, if we neglect $\Delta i.\Delta j$, where i≠j, then the bias of mean of RWD can be simplified to

$$Bias(\mu_{rwd}) \approx [\Delta_1\rho + \Delta_2(1-\rho)](\hat{\mu}_P - \hat{\mu}_N) \quad (5)$$

Considering the power calculation, by (1), the variance of μ_{rwd} is given by

$$Var(\hat{\mu}_{rwd}) = \frac{1}{n}\hat{\sigma}^2 = s^2\left(\frac{\sigma_P^2}{Snn_P}\right) + (1-s)^2\left(\frac{\sigma_N^2}{(1-S)nn_N}\right) \quad (6)$$

Where

$$\sigma_P^2 = \frac{1}{snn_P}\sum_{j=1}^{sn}\sum_{i=1}^{n_P}(x_{ij} - \overline{x})^2$$

$$\sigma_N^2 = \frac{1}{(1-s)nn_N}\sum_{j=1}^{(1-\ )\mathrm{n}}\sum_{i=1}(y_{ij} - \overline{y})^2$$

Table 1. Selection bias when ρ and r_1 are given.

ρ	r1	r2	Selection bias
0	0	0	$E(\hat{\mu}_N) - \mu_N$
0	0.5	0.5	$\frac{1}{2}(E(\hat{\mu}_p) - \mu_p) + \frac{1}{2}E(\hat{\mu}_N) - \mu_N$
0	1	1	$E(\hat{\mu}_p) - \mu_p$
0.5	0	0	$E(\hat{\mu}_N) - \mu_N$
0.5	0.5	0.5	$\frac{1}{2}(E(\hat{\mu}_p) - \mu_p) + \frac{1}{2}E(\hat{\mu}_N) - \mu_N$
0.5	1	1	$E(\hat{\mu}_p) - \mu_p$
1	0	0	$E(\hat{\mu}_N) - \mu_N$
1	0.5	0.5	$\frac{1}{2}(E(\hat{\mu}_P) - \mu_P) + \frac{1}{2}(E(\hat{\mu}_N) - \mu_N)$
1	1	1	$E(\hat{\mu}_p) - \mu_p$

If we take derivative above, then

$$\frac{\partial}{\partial r_1}\left[\operatorname{Var}\left(\hat{\mu}_{rwd}\right)\right]=\frac{\rho}{n}\left(\frac{\sigma_P^2}{n_P}-\frac{\sigma_N^2}{n_N}\right) \quad (7)$$

$$\frac{\partial}{\partial r_2}\left[\operatorname{Var}\left(\hat{\mu}_{rwd}\right)\right]=\frac{1-\rho}{n}\left(\frac{\sigma_P^2}{n_P}-\frac{\sigma_N^2}{n_N}\right) \quad (8)$$

In practice, the variance in positive studies is larger than in negative studies, since difference between controlled group and referred group is larger in positive study. If the gap of sizes of positive and no positive studies is not huge, we have $\frac{\sigma_P^2}{n_P}>\frac{\sigma_N^2}{n_N}$. Thus, $Var\left(\hat{\mu}_{rwd}\right)$ is an increasing function of $r_1\, and\, r_2$. Furthermore, the power of the RWD can be calculated by the probability:

$$P\{\frac{\sigma_P^2}{n_P}>\frac{\sigma_N^2}{n_N} \mid \mu_P,\mu_N,\sigma_P^2,\sigma_N^2,r_i, and\ \rho\}$$

Based on the data availability of RWD.

Estimation of u_P – u_N

Let (L_P,U_P) and (L_N,U_N) be the $(1-a)100\%$ for u_P and u_N. Under the assumption of normality, we have

$$(L_P,U_P)=\hat{u}_P \pm Z_{1-a/2}\frac{\hat{\sigma}_p}{\sqrt{snn_p}}$$

$$(L_N,U_N)=\hat{u}_N \pm Z_{1-a/2}\frac{\hat{\sigma}_N}{\sqrt{(1-s)nn_N}}$$

According to Chow's assumption (2020), when selection bias does exist, it is reasonable to assume the positive studies and no positive studies are different. In this case, we assume $u_P > u_N$ *and* (L_P,U_P) *and* (L_N,U_N) would not have intersection. At some extreme case [7], L_P is close to U_N. Then, we have

$$\hat{u}_P - Z_{1-a/2}\frac{\hat{\sigma}_p}{\sqrt{snn_p}} \approx \hat{u}_N + Z_{1-a/2}\frac{\hat{\sigma}_N}{\sqrt{(1-s)nn_N}}$$

Therefore, the distance of u_P and u_N can be calculated as the estimation of $u_P - u_N$, we have

$$\hat{u}_P - \hat{u}_N \approx Z_{1-\frac{a}{2}}\frac{\hat{\sigma}_p}{\sqrt{snn_p}} + Z_{1-\frac{a}{2}}\frac{\hat{\sigma}_N}{\sqrt{(1-s)nn_N}} \quad (9)$$

Reproducibility probability: The definition of reproducibility probability is the estimated power (EP) of a future trial using the information from previous trials. In theory, different trials are independent; the probability of achieving a statistically significant result from the new trial would be identical with previous studies if these trials apply the same study design and hypothesis, regardless of the outcome of previous trials. When H_A is true, the probability of gaining a significant result is the power of the test [8,9]

$$P\left(|T|>c|H_A\right)=P\left(|T|>c|\theta\right)$$

Where H_A is the alternative hypothesis and ϴ is an unknown parameter or a vector of parameter (Shao & Chow, 2002). Even though trials are independent to each other, it reasonable to use the previous trials information in RWD to refer a later trial. Reproducibility probability could be estimated through (i) the estimated power (EP) approach, (ii) the confidence bound (CB) approach, which is more a conservative approach than EP approach and (iii) Bayesian approach, a more sensible approach to obtain reproducibility [9].

To illustrate EP approach, consider a control group and reference group trial with unequal variances. Let x_{ij} be the jth subject in the ith group (i = 1,2) and distributed as $N(u_i,\sigma_i^2)$ respectively, in which $\sigma_1^2 \neq \sigma_2^2$. Therefore, assuming n_1 and n_2 are large, statistics T is shown as

$$T=\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

Then, T asymptotically has normal distribution N(ϴ,1), where

$$\theta=\frac{u_1-u_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Therefore, the reproducibility probability can be calculated by replacing ϴ by its estimate statistics T

$$\hat{P}=\Phi\left[T(x)-z_{0.975}\right]+\Phi\left[-T(x)-z_{0.975}\right] \ (10)$$

The EP approach would provide an optimistic result that the adjustment of bias might be inadequate when applied in our proposed approach. The CB method would yield a more cautious estimation of reproducibility probability.

Considering the trial setting mentioned before, the CB approach considers a 95th percent lower confidence bound as the reproducibility probability. Therefore, CB approach provides the estimation of reproducibility probability as follow equation.

$$\hat{P}_{-}=\Phi\left[\left|T(x)\right|-2z_{0.975}\right] \qquad (11)$$

Furthermore, the Bayesian approach provides a clear definition of reproducibility. To be more specific, we assume the unknown parameter ϴ is a random vector with priori distribution π(ϴ), which is known. Then, we can define reproducibility probability is the conditional probability of |T|>c in the future trial, which can be shown as

$$p\left(|T|>c|x\right)=\int p\left(|T|>c|\theta\right)\pi\left(\theta|x\right)d\theta \quad (12)$$

where T is T statistics based on future data set and π(ϴ|x) is the posterior density of ϴ given x.

Considering the same setting in EP approach part, if the variance, σ^2, is known, then reproducibility probability is

$$\int\left[\Phi\left(\theta-z_{0.975}\right)+\Phi\left(-\theta-z_{0.975}\right)\right]\pi\left(\theta|x\right)d\theta=\Phi\left(\frac{T(x)-z_{0.975}}{\sqrt{2}}\right)+\Phi\left(\frac{-T(x)-z_{0.975}}{\sqrt{2}}\right)\# \ 13 \quad (13)$$

Where

$$T=\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$$

when $|T(x)|>z_{0.975}$, the probability close to

$$\Phi(\frac{|T(x)|-z_{0.975}}{\sqrt{2}})$$

Estimation of $\Delta_i\,(i=1,2)$

The proportion of positive studies in RWD, $\hat{r}$, is highly likely overestimated, since positive studies are more likely collected. In other words, $\hat{r}$ is likely larger than the true proportion r.

According to the definition of reproducibility probability, r can be estimated by reproducibility probability of observing a significant result based on the mean and variance of response in RWD. The probability can be represented as follow equation:

$$p=P\{future\,study\,is\,positive|u\equiv\hat{\mu}_B\ and\,\sigma\equiv\hat{\sigma}_B\}$$

The interpretation of the reproducibility probability under this setting is that we expect to gain p × 100 significant studies if the similar experimental setting trial conduct 100 times based on the observed mean response $(\hat{\mu}_B)$ and standard deviation $(\hat{\sigma}_B)$ in RWD. To be more specific, with RWD, we have

$$\hat{u}_B = \hat{u}_1 - \hat{u}_0 \; and \; \hat{\sigma}_B = \sqrt{\frac{n_1\hat{\sigma}_1^2 + n_0\hat{\sigma}_2^2}{n_1 + n_0}}$$

where $\hat{u}_1$ and $\hat{u}_0$ are the pooled means of controlled and reference group in RWD, n_1 and n_0 are the sizes of controlled and reference group of pooled RWD, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_0^2$ are the estimated variances of the controlled and reference group, respectively. Then, the T statistic is shown as

$$T = \sqrt{\frac{n_{m1}n_{m0}}{n_{m1} + n_{m0}}}\frac{\hat{u}_B}{\hat{\sigma}_B}$$

where n_{m1} and n_{m0} are the median sample sizes of controlled and reference group in studies included in RWD. Therefore, the bias of positive proportion, $\Delta i.$ can be represent as

$$\Delta_i = \hat{r}_i - \hat{p}_i$$

Simulation study

In this part, we used the method proposed in chapter 2 to adjust bias in simulation studies of different scenarios. The performance of methods is evaluated by comparing the adjusted mean and the true mean of the target population. For simplicity, we assume that the structural data is from equal sample size randomized clinical trials and the unstructured data is from equal sample size observational cohort studies. All these data are generated from the same population. As for the target population, we assume that each trial or study contains two groups of patients, the treatment group and the control group. Each group's responses follow the normal distribution N(2,25) and N(0,25). In order to achieve 80% power to detect a clinically meaningful difference of 2, 100 subjects in each group are needed based on sample size calculation. Furthermore, to achieve 50%, 60%, 70%, 80%, 90% power in the trial, we need sample size of 49, 63, 79, 100, 133 per group respectively. We generated 1000 studies to simulate the target population for each sample size. To evaluate the performance of the proposed methods, we define relative bias as $(\hat{\mu}_{rwd} - \mu)/\mu$, where $\hat{\mu}_{rwd}$ is the estimated treatment effect based on real-world data and μ is the true treatment effect, which is 2 in this simulation. We also define the relative adjusted bias as $(\hat{\mu}_{rwd} - \mu - \varepsilon)/\mu$, where ε is the estimated bias based on the proposed method. Four scenarios are considered to assess the robustness of the adjustment method. The first scenario compares performance between EP (estimated power) method, CB (confidence bound) method, and Bayesian method in structural data under the previous setting when only structural data presented. The second scenario compares the performance of the three methods when negative data are absent in structural data. The third scenario tests the robustness when structural data and unstructured data presents inconsistent positive proportion. Finally, we test the robustness when inconsistent positive proportion presents and imbalance structural data proportion occurs.

In the first scenario, we assessed the relative bias and adjusted bias based on three adjustment methods by different power and r, the proportion of positive studies, which is summarized by (Table 2) and (Figure 2). As expected, when the gap between the proportion of positive studies and power, the theoretical proportion of positive studies, increases, the selection bias increases. As the most conservative adjusted method, EP method can slightly adjust selection bias compared with the other two methods. When the bias larger than 15%, the bias can reduce 20% by the EP method. On the other hand, when the bias less than 5%, the adjustment is limited, causing the adjusted bias is nearly the same as original bias. As for CB approach, the adjustment is significantly larger than the adjustment made by EP approach. When the selection bias is larger than 10%, a significant adjustment in right direction can be made by the EP approach. However, when the bias is smaller than 10%, the CB approach

Table 2. Scenario 1: Performance of proposed approaches at different power.

Power	*r*	Bias	EP approach	CB approach	Bayesian Approach
0.508	0.7	0.1584	0.1449	0.0313	0.1378
	0.75	0.1986	0.1808	0.0577	0.1716
	0.8	0.2409	0.2155	0.0677	0.2029
	0.85	0.2727	0.2387	0.0786	0.224
	0.9	0.3117	0.2662	0.0809	0.2478
	0.95	0.3507	0.2796	0.0275	0.2531
0.6122	0.7	0.0427	0.0332	-0.0804	0.0252
	0.75	0.0815	0.0677	-0.0568	0.0576
	0.8	0.1167	0.0972	-0.0394	0.0849
	0.85	0.1527	0.1263	-0.0242	0.1115
	0.9	0.1854	0.1445	-0.0501	0.1241
	0.95	0.2285	0.1763	-0.0424	0.152
0.7102	0.7	0.0115	0.014	-0.069	0.0063
	0.75	0.043	0.0424	-0.0442	0.0337
	0.8	0.0669	0.0622	-0.0288	0.0526
	0.85	0.1041	0.0953	-0.0078	0.0838
	0.9	0.1318	0.1168	-0.0002	0.1033
	0.95	0.1687	0.1476	0.0218	0.1326
0.8074	0.7	-0.0741	-0.0691	-0.1395	-0.0762
	0.75	-0.0394	-0.0366	-0.1102	-0.0445
	0.8	-0.0075	-0.0073	-0.0866	-0.0163
	0.85	0.0217	0.0183	-0.0638	0.0087
	0.9	0.052	0.0442	-0.0514	0.0327
	0.95	0.0835	0.0701	-0.0363	0.057
0.9035	0.7	-0.1138	-0.1029	-0.1628	-0.1099
	0.75	-0.0827	-0.0738	-0.1339	-0.0811
	0.8	-0.0502	-0.0434	-0.1049	-0.0511
	0.85	-0.0188	-0.0147	-0.0781	-0.0227
	0.9	0.0125	0.0136	-0.0539	0.0051
	0.95	0.043	0.0401	-0.0319	0.031

can lead the bias toward the wrong direction and even show an absolute bias larger than the original bias. Finally, the Bayesian approach give us a more effective adjustment result. The overall adjustment pattern is similar to the EP approach but more effective. To be more specific, when the bias larger than 15%, the bias could reduce 40% by the EP method. When the bias is less than 5%, the Bayesian approach significantly leads to more adjustment compared with EP approach. In addition, compared with the CB approach, the Bayesian approach always makes adjustment toward the right direction.

Second scenario, we consider the situation that only positive studies exist in the structural data. In other words, the r_1, proportion of positive study is fixed to 1 in this scenario. As we can see in Table 3 and (Figure 3), bias decrease when power increase, and the result of the EP approach and Bayesian approach is similar to the pattern in scenario 1. More specifically, when bias is more negligible, the EP and Bayesian approaches lead less adjustment. In addition, the EP approach provides a more conservative adjustment compared with the Bayesian approach. As for CB approach, similar as scenario 1, when bias is larger than 10%, the bias can be significantly reduced. However, the CB

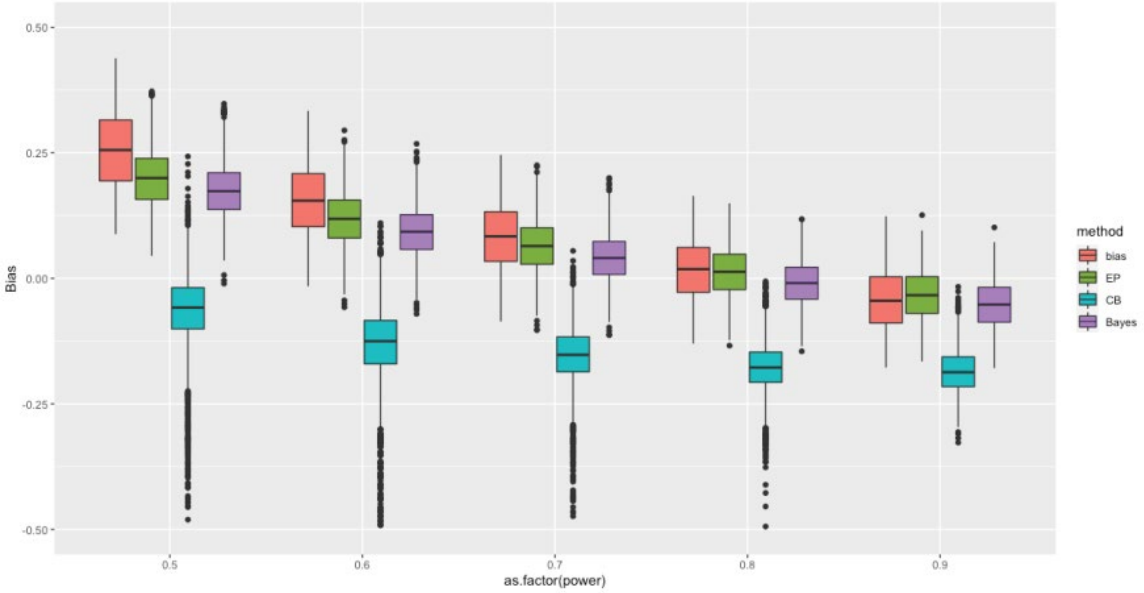

Figure 2. Box plot of performance of proposed approaches at different power.

Table 3. Performance of proposed approaches at different power when negative studies absent.

Power	Bias	EP approach	CB approach	Bayesian Approach
0.508	0.4170	0.3675	0.1910	0.3472
0.6122	0.2915	0.2525	0.0986	0.2343
0.7102	0.2100	0.1790	0.0293	0.1606
0.8074	0.1402	0.1178	-0.0246	0.0999
0.9035	0.0455	0.0310	-0.0934	0.0152

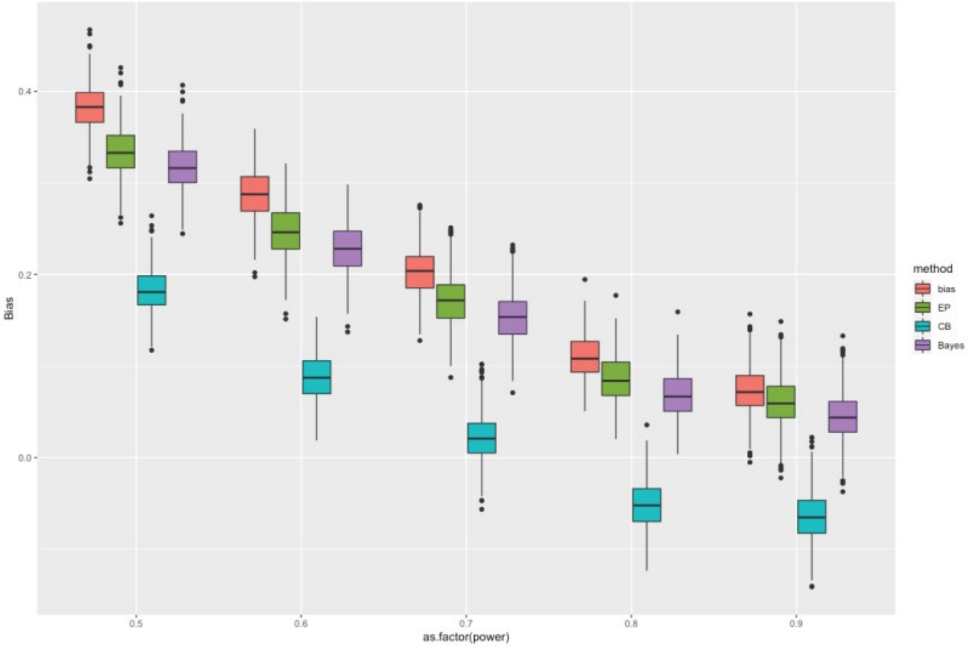

Figure 3. Box plot of performance of proposed approaches at different power when negative studies absent.

approach could lead adjustment toward the wrong direction when bias smaller than 10%.

In the third scenario, the power set to 0.6, ρ set to 0.5, and the rest of setting is the same as previous scenario, and we consider the situation that the structural data and unstructured data provide inconsistent positive proportion. To be more specific, we consider that the structural data provide high proportion of positive result and the unstructured data provided low proportion of positive result. In this context, according to Table 4, when the two proportion are away from each other and the mean of two proportion away from the setting power, 0.6, the bias decreases. Similar situation observed in scenario 1 and 2. The EP approach and Bayesian approach provide effective bias adjustments toward the right direction, and the Bayesian approach provide slightly more adjustment compared with EP approach for most of the case. As for CB approach, most of the adjustment towards wrong direction, which in line with the observation about CB approach under 10% bias from scenario 1 and 2. However, we do notice that the adjustment based on Bayesian method also toward to wrong direction slightly when bias is less than 5%. In conclusion, the EP approach and Bayesian approach provide effective adjustment under the inconsistent positive study proportion situation, but CB approach lead the adjustment toward wrong direction in this scenario.

In the final scenario, we consider both imbalance of structural data and inconsistent positive proportion occurs. The r_1 set to 0.4, and r_2 set to 0.8, the power set to 0.6, and the rest of setting is the same as scenario 1. As we can see in Table 5, when the ρ, the structural data proportion, increase (i.e., Balance of structural and unstructured data), the bias reduced. Under the imbalanced structural data proportion, both EP approach and Bayesian Approach shows capability to reduce bias. But for CB approach, it adjusts the bias toward the wrong direction and even increase the bias when the bias less than 5%.

In summary, three proposed methods can adjust the bias in real world data. EP approach provides the most conservative and stable adjustment. The Bayesian method provides more adjustment toward the right direction with robustness in most the extreme cases. In addition, the CB approach leads to the most aggressive adjustment of these three methods. However, it shows less stability and robustness when the power is larger than the proportion of positive studies in structural data or the bias of real-world data is less than 10%.

Discussion

Several limitations were found in the simulation study. First, when the proportion of positive studies is close to the designed power, the adjustment led by the EP approach and Bayesian approach is limited. In some cases, the Bayesian method even enlarges the bias. This is because the reproducibility probability-based approaches can make over-adjustments when bias is not obvious. Another reason is that the estimation of $\hat{\mu}_P - \hat{\mu}_N$ is over-estimated when bias is negligible. The second limitation is that CB approach provides adjustment towards the wrong direction when bias is less than 10%. One explanation is that the lower confidence bound provided by CB approach is too conservative in estimating the real positive proportion, r, when bias is less than 10%. In addition, the over-estimation of $\hat{\mu}_P - \hat{\mu}_N$ can also lead adjustment failure of CB approach.

Previously, compared the substantial evidence and real-world evidence and pointed out that the bias in the substantial evidence is minimized but selection bias exists in the real-world evidence, and proposed a reproducibility probability based bias adjustment approach. Compared to the method in our paper, we added the Bayesian method to adjust bias, which has been proved as an effective and robust adjustment method and applied the method in a new architecture of real-world data, which divided the data into structural and unstructured data. In reality, meta-analysis has been used to estimate non-inferiority margin or treatment effect based on historical data, which is similar to use real-world data to estimate the treatment effect. The parameter, $\hat{\mu}_B, \sigma_B^2, \hat{\mu}_P, \hat{\mu}_N, \sigma_P^2 \; and \; \sigma_N^2$ can also be estimated by meta-analysis even the individual data is unavailable [10]. Therefore, the combination of meta-analysis and our method can also adjust the bias in real world data. However, either our proposed method or meta-analysis method has the assumption that there is no treatment-by-study interaction and treatment-by-center interaction. Therefore, further investigation of these two interactions can be included in future work.

Table 4. Performance of proposed approaches when inconsistent r presents.

r_1	r_2	Bias	EP Approach	CB Approach	Bayesian Approach
0.75	0.1	-0.1406	-0.1267	-0.227	-0.1254
	0.2	-0.1043	-0.0942	-0.1995	-0.095
	0.3	-0.0685	-0.0621	-0.1728	-0.0651
	0.4	-0.0298	-0.0266	-0.1426	-0.0318
	0.5	0.0059	0.005	-0.1173	-0.0021
0.8	0.1	-0.1236	-0.1119	-0.2139	-0.1116
	0.2	-0.0868	-0.0787	-0.1863	-0.0806
	0.3	-0.0496	-0.0449	-0.1584	-0.049
	0.4	-0.0136	-0.0128	-0.1319	-0.0188
	0.5	0.0255	0.0228	-0.1035	0.0146
0.85	0.1	-0.1049	-0.0949	-0.2001	-0.0958
	0.2	-0.0665	-0.0598	-0.1702	-0.0629
	0.3	-0.0303	-0.0273	-0.1432	-0.0324
	0.4	0.0074	0.0067	-0.1154	-0.0005
	0.5	0.0435	0.0387	-0.0907	0.0295
0.9	0.1	-0.0868	-0.0788	-0.1853	-0.0807
	0.2	-0.0496	-0.045	-0.1577	-0.049
	0.3	-0.0121	-0.011	-0.1303	-0.0171
	0.4	0.0239	0.021	-0.1048	0.0129
	0.5	0.0609	0.0537	-0.0803	0.0435

Table 5. Imbalance proportion of structural data with inconsistent positive study proportion.

ρ	Bias	EP Approach	CB Approach	Bayesian Approach
0.1	0.1057	0.0925	-0.0423	0.0808
0.2	0.0756	0.0659	-0.0657	0.0555
0.3	0.0467	0.0406	-0.0875	0.0315
0.4	0.0182	0.0155	-0.1097	0.0079

Regarding the bias in RWD or historical data, various methods have been proposed to identify and quantify the bias. An adjusted rank correlation test has been proposed to identify the bias in a meta-analysis based on historical data [11]. The test shows fairly high power for meta-analysis when studies number are large, which also suitable for RWD bias detection. Egger proposed a test of asymmetry funnel plot that can predict the discordance of outcomes in meta-analysis and evaluated the prevalence of bias in meta-analyses [12]. Furthermore, several methods are proposed to control the bias to acquire inferiority margin or treatment effect from historical data. Chow and Shao proposed a method for selecting non-inferiority margin with statistical assurance. However, these methods only detected the bias but did not adjust the bias in historical data. As for bias adjustment method in RWD proposed a structure that generates robust RWE from RWD *via* adjusting confounding bias based on causal inference [13]. In addition, several suggestions study designs were proposed to adjust the selection bias in RWD. In our method, not only do we define and quantify the selection bias in RWD, but also we give adjustment methods quantitively [14,15].

Conclusion

Real-World data often consist of positive or negative studies and the data may be structured or unstructured. In this case, the validity of RWD is a concern for providing evaluation of safety and efficacy of the test treatment under investigation. The validity of RWD or Real-World Evidence is important, especially when it is intended to support a regulatory submission. In this paper, we discussed the selection bias adjustment of real-world data based on reproducibility probability approaches. We defined the selection bias in real-world data under the structural and unstructured data setting. Based on the form of bias, three reproducibility probability-based approaches have been introduced to estimate the real proportion of positive studies in the structural and unstructured data. The reproducibility probability-based approach provides effective bias adjustment when the proportion of positive studies is different as designed power. EP approach and Bayesian approach provide robust and effective bias adjustment in most of cases, and the CB approach provide huge and effective adjustment only when bias is larger than 10%.

References

1. FDA. "Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products." *Pharmacoepidemiol Drug Saf* (2021).
2. FDA. "Framework for FDA's real-world evidence program." *Food and Drug Administration* (2018).
3. Kabisch, Maria, Christian Ruckes, Monika Seibert-Grafe and Maria Blettner. "Randomized controlled trials: Part 17 of a series on evaluation of scientific publications." *Dtsch Arztebl Int* 108 (2011): 663.
4. Franklin, Jessica M. and Sebastian Schneeweiss. "When and how can real world data analyses substitute for randomized controlled trials?." *Clin Pharm Therap* 102 (2017): 924-933.
5. Fang, Yixin, Hongwei Wang and Weili He. "A statistical roadmap for journey from real-world data to real-world evidence." *Ther Innov Regul Sci* 54 (2020): 749-757.
6. Song, Fuyu, Chenxuan Zang, Xinyi Ma and Sifan Hu, et al. "The use of real-world data/evidence in regulatory submissions." *Contemp Clin Trials* 109 (2021): 106521.
7. Chow, Shein-Chung. *Innovative methods for rare disease drug development*. *Chapman and Hall/CRC* (2020).
8. Goodman, Steven N. "A comment on replication, p-values and evidence." *Stat Med* 11 (1992): 875-879.
9. Shao, Jun and Shein-Chung Chow. "Reproducibility probability in clinical trials." *Stat Med* 21 (2002): 1727-1742.
10. Hozo, Stela Pudar, Benjamin Djulbegovic and Iztok Hozo. "Estimating the mean and variance from the median, range, and the size of a sample." *BMC Med Res Methodol* 5 (2005): 1-10.
11. Begg, Colin B. and Madhuchhanda Mazumdar. "Operating characteristics of a rank correlation test for publication bias." *Biometrics* (1994): 1088-1101.
12. Egger, Matthias, George Davey Smith, Martin Schneider and Christoph Minder. "Bias in meta-analysis detected by a simple, graphical test." *Bmj* 315 (1997): 629-634.
13. Levenson, Mark, Weili He, Jie Chen and Yixin Fang, et al. "Biostatistical considerations when using RWD and RWE in clinical studies for regulatory purposes: A landscape assessment." *Stat Biopharm Res* (2021): 1-11.
14. Sheikhalishahi, Seyedmostafa, Riccardo Miotto, Joel T. Dudley and Alberto Lavelli, et al. "Natural language processing of clinical notes on chronic diseases: Systematic review." *J Med Internet Res* 7 (2019): e12239.
15. Miguel-Alvarez, Marina, Alejandro Santos-Lozano, Fabian Sanchis-Gomar and Carmen Fiuza-Luces, et al. "Non-steroidal anti-inflammatory drugs as a treatment for Alzheimer's disease: A systematic review and meta-analysis of treatment effect." *Drugs Aging* 32 (2015): 139-147.

How to cite this article: Zhao, Yuankang and Shein-Chung Chow. "Statistical Evaluation of the Validity of Real World Data and Real World Evidence." *J Biom Biosta* 14 (2023): 150.