TRIALS

A Bayesian prediction model between a biomarker and the clinical endpoint for dichotomous variables

Zhiwei Jiang[1,2], Yang Song[2], Qiong Shou[2], Jielai Xia[1*] and William Wang[2]

Abstract

Background: Early biomarkers are helpful for predicting clinical endpoints and for evaluating efficacy in clinical trials even if the biomarker cannot replace clinical outcome as a surrogate. The building and evaluation of an association model between biomarkers and clinical outcomes are two equally important concerns regarding the prediction of clinical outcome. This paper is to address both issues in a Bayesian framework.

Methods: A Bayesian meta-analytic approach is proposed to build a prediction model between the biomarker and clinical endpoint for dichotomous variables. Compared with other Bayesian methods, the proposed model only requires trial-level summary data of historical trials in model building. By using extensive simulations, we evaluate the link function and the application condition of the proposed Bayesian model under scenario (i) equal positive predictive value (PPV) and negative predictive value (NPV) and (ii) higher NPV and lower PPV. In the simulations, the patient-level data is generated to evaluate the meta-analytic model. PPV and NPV are employed to describe the patient-level relationship between the biomarker and the clinical outcome. The minimum number of historical trials to be included in building the model is also considered.

Results: It is seen from the simulations that the logit link function performs better than the odds and cloglog functions under both scenarios. PPV/NPV ≥0.5 for equal PPV and NPV, and PPV + NPV ≥1 for higher NPV and lower PPV are proposed in order to predict clinical outcome accurately and precisely when the proposed model is considered. Twenty historical trials are required to be included in model building when PPV and NPV are equal. For unequal PPV and NPV, the minimum number of historical trials for model building is proposed to be five. A hypothetical example shows an application of the proposed model in global drug development.

Conclusions: The proposed Bayesian model is able to predict well the clinical endpoint from the observed biomarker data for dichotomous variables as long as the conditions are satisfied. It could be applied in drug development. But the practical problems in applications have to be studied in further research.

Keywords: Biomarker, Clinical endpoint, Bayesian model, PPV, NPV

Background

The biomarker employed as a surrogate to evaluate efficacy is preferred in clinical trials. It leads to reduced trial durations and costs [1,2], improved compliance [3], better ethical satisfaction [4], *etcetera* For example, United States Food and Drug Administration (FDA) has enabled accelerated marketing approval based on a surrogate endpoint,

rather than the primary endpoint, in life-threatening diseases [5,6]. Controversy has surrounded the evaluation of a biomarker as a surrogate [7-20] since Prentice [7] formulated a statistical framework in the context of hypothesis testing. However, the evaluation of a biomarker covers various considerations [21-23]. The experience of the evaluation is limited and very few biomarkers are accepted as surrogate endpoints in practice [21]. On the other hand, as an intermediate endpoint, the biomarker is still helpful to predict the effect on clinical endpoint from the intervention even it cannot directly replace the true

* Correspondence: jielaixia@yahoo.com
[1]Department of Health Statistics, School of Preventive Medicine, Fourth
Military Medical University, No. 169 Changle West Road, Xi'an, Shaanxi, China
Full list of author information is available at the end of the article

clinical endpoint as a surrogate. In drug development, the prediction of a true clinical endpoint from an early biomarker makes sense in early drug efficacy evaluation and decision making. The key point here is how to build and evaluate the prediction model between biomarker and clinical endpoint.

A meta-analytic approach was first considered by Fleming [11] and Hughes *et al.* [12] to validate a surrogate endpoint from a collection of previous trials. Some approaches modeled the association between a biomarker and clinical endpoint based on trial-level summary data in a frequentist framework [16-19]. But few have made such evaluations within the Bayesian framework. The main paper involving a Bayesian meta-analytic method is that of Daniels and Hughes [13], who considered a mixed effect model for the trial-level association of treatment effect on biomarker and clinical outcome. However, it required patient-level data to estimate the correlation between biomarker and clinical endpoint in the model building [24]. Both patient-level data and trial-level summary data were necessary for the method of Daniels and Hughes. Furthermore, Van Walraven *et al.* [24] implemented a Monte Carlo model and Wang *et al.* [25] employed classical Cox model to predict clinical outcomes from biomarkers.

The evaluation of the model between the biomarker and the clinical endpoint is another key concern. It is vital that the model is able to describe the association between the biomarker and the clinical outcome well and to predict the clinical endpoint from the observed biomarker accurately and precisely in a new trial. Buyse *et al.* [14] proposed the coefficient of determination R^2_{trial} to evaluate the trial-level association and predictive ability. But the value of R^2_{trial} is difficult to interpret [20], and it is difficult to clearly define which model is appropriate for prediction. On the other hand, the accessibility of trial-level summary data from historical trials is one of the advantages of meta-analytic approaches. But this approach also leads to information loss compared with the use of patient-level data. Therefore, simulation studies [17], which evaluate the model based on the simulative data from a specific assumption, ignore the effect of information loss from trial-level data and may over-evaluate the predictive ability of meta-analytic model. It is more objective and reasonable to evaluate the meta-analytic model based on the assumption of the patient-level data of previous trials in the simulations.

In this paper, we consider a Bayesian meta-analytic method for building a prediction model between the biomarker and the clinical endpoint when both endpoints are dichotomous, which is completely independent of patient-level data. It is used to predict rate ratio (*RR*) of the clinical endpoint from an early biomarker. In addition, we evaluate the proposed prediction model by using extensive simulations from clinical practical considerations.

The patient-level data are simulated to evaluate the meta-analytic prediction model to avoid the assumption on trial-level data. Positive predictive value (PPV) and negative predictive value (NPV), which measure the patient-level association between biomarker and clinical endpoint for dichotomous variables, are employed in the simulations.

From the clinical point, a good biomarker is expected to have higher PPV and NPV to predict long-term clinical outcome (for example, in the early detection of a disease in clinical diagnosis, the early observed recurrence of a tumor in oncology study and so on). But a biomarker with higher NPV and lower PPV is also common in medical studies. For example, persistent infection of HPV (human papillomavirus) is considered to be a potential biomarker of high grade cervical disease and eventually cervical cancer. Because though HPV persistent infection case does not always progress to high grade cervical disease, the one who is not infected with HPV persistently has only a small probability of getting the disease from the viewpoint of medical mechanisms. In HPV vaccine clinical trials, a persistent infection of HPV, as a biomarker that has higher NPV and lower PPV, is able to help predict vaccine efficacy. Therefore, both scenarios, (i) equal PPV and NPV and (ii) higher NPV and lower PPV, are considered in the evaluation of the proposed model in the simulations.

The proposed prediction model between biomarker and clinical outcome in this article is built in a Bayesian framework. It is different from the frequentist meta-analytic approaches [14-19,26]. It intuitively describes the association between biomarker and clinical endpoint and is easy to be implemented by using Markov Chain Monte Carlo (MCMC) techniques. But compared with other Bayesian methods, the proposed model has its own features. The Bayesian mixed model proposed by Daniels and Hughes [13] is not a complete meta-analytic method and the advantage of the meta-analytic approach cannot be shown. But the proposed Bayesian model has no such restriction. Furthermore, the method of Daniels and Hughes ignores the variability of within-trial treatment effect [10], which is included in the proposed Bayesian model.

Methods
Bayesian model for prediction
Consider N randomized trials of size $n_i(i = 1, 2, ..., N)$. Equal sample size is assumed in the treatment and control group in each trial. In the i-th trial, B_{Ti} and B_{Ci} biomarker responses are observed in the treatment and control group, respectively. X_{Ti} and X_{Ci} subjects respond to the clinical outcome in the two groups. Let ϕ_{Bi} be the proportion of biomarker responses in the treatment group in the i-th trial. It is estimated by $\hat{\phi}_{Bi} = B_{Ti}/(B_{Ti} + B_{Ci})$. Correspondingly, $\hat{\phi}_{Xi} = X_{Ti}/(X_{Ti} + X_{Ci})$ is the estimate

of ϕ_{Xi}, which denotes the proportion of clinical outcome responses in the treatment group. Assuming the association between the biomarker and the clinical endpoint is equal across N trials irrespective of the intervention, a generalized linear model

$$g(\phi_{Xi}) = \beta g(\phi_{Bi}) + \beta_0 \tag{1}$$

is proposed to describe the relationship. In the model, $g(\cdot)$ is a link function. The biomarker and clinical endpoint are transformed by using the link function because they are both dichotomous. In this paper, three link functions, the odds function:

$$g(x) = \frac{x}{1-x}, \tag{2}$$

the logit function:

$$g(x) = \text{logit}(x) = \ln\left(\frac{x}{1-x}\right), \tag{3}$$

and the cloglog function:

$$g(x) = \text{cloglog}(x) = \ln(-\ln(1-x)), \tag{4}$$

are considered and will be further compared in the simulations.

Given B_i, B_{Ti}, X_i and X_{Ti}, ϕ_{Bi} and ϕ_{Xi} follow the distributions of

$$B_{Ti} \sim Binomial(B_i, \phi_{Bi}). \tag{5}$$

and

$$X_{Ti} \sim Binomial(X_i, \phi_{Xi}), \tag{6}$$

in which $B_i = B_{Ti} + B_{Ci}$ and $X_i = X_{Ti} + X_{Ci}$.

In the Bayesian model, we consider a uniform prior for the coefficient β and a normal prior for the intercept β_0 with mean zero and variance σ^2. The ϕ_{Bi} takes a prior distribution of $Beta(a_{Bi}, b_{Bi})$. The non-informative priors are considered for all three parameters β, β_0 and ϕ_{Bi}. The posterior distributions of the three parameters depend on the link function in the model. Given B_i, B_{Ti}, X_i, X_{Ti} and a specific link function, the estimates of posterior distributions for β, β_0 and ϕ_{Bi} are calculated by using the MCMC method based on the formulas (1), (5) and (6). The uncertainty on within-trial treatment effect is considered by incorporating formula (5) and (6) in the model.

For a new trial j, the rate ratio of the treatment and control group on the clinical endpoint

$$RR_j = \frac{\phi_{Xj}}{1-\phi_{Xj}} \tag{7}$$

is employed to evaluate the efficacy of the intervention. An equal association between the biomarker and the clinical endpoint across the new trial and historical trials is assumed here. It means that the biomarker in the new trial captures the same treatment effect as the one in the historical trials. Based on the Bayesian model built from N historical trials, the MCMC estimation of predictive distribution for ϕ_{Xj} is obtained when ϕ_{Bj} is given. Correspondingly, the predictive distribution of RR_j is derived from formula (7). We take the median of the predictive distribution as the point estimate of RR_j prediction and construct the 95% credible interval (CI) with 2.5% and 97.5% percentiles. The flow chart of Bayesian model building and prediction is depicted in Figure 1.

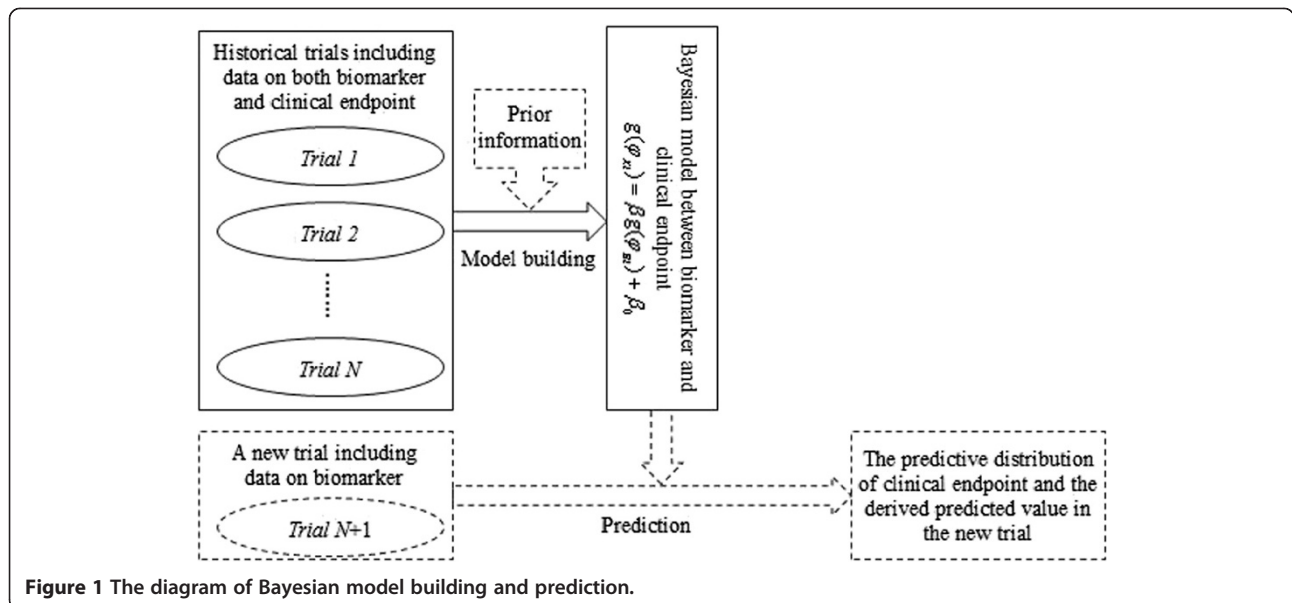The predictive ability of the proposed model is related to the association strength between the biomarker and

Figure 1 The diagram of Bayesian model building and prediction.

the clinical outcome. PPV and NPV are generally employed to measure the association from the patient level. These directly affect the predictive ability of the proposed Bayesian model. When one uses the proposed approach to build the Bayesian model and to predict the clinical outcome of the new trial, the first thing is to evaluate the strength of the association between the biomarker and the clinical endpoint, which is measured by PPV and NPV. Therefore, a simulation study is conducted to explore the application condition of the proposed method.

Simulation study

A simulation study is employed to (a) compare the predictive ability of different link functions (the odds, logit and cloglog functions) in the proposed Bayesian model when PPV and NPV vary; (b) explore the effect of the association strength between biomarker and clinical endpoint, which is measured by PPV and NPV, on the predictive ability of the model; and (c) discuss the number of historical trials to be included in model building for a good clinical prediction in the new trial. As we have mentioned above, the biomarkers with (i) equal PPV and NPV and (ii) higher NPV and lower PPV are both common in medical studies. From the practical perspective, both scenarios are considered in the simulations. All simulations are repeated for 5,000 times and performed by using R package for data generation and calculation and OpenBUGS for Bayesian model fitting.

Data generation process

In N historical randomized trials, it is assumed that the biomarker response rate of control group is 0.3, and that 400 patients complete the trial for simplicity. The sample size ratio of treatment and control group is 1:1.

Let π_{BTi} and π_{BCi} be the biomarker response rate of the treatment and control group in the i-th trial. Then $\phi_{Bi} = \pi_{BTi}/(\pi_{BTi} + \pi_{BCi})$. when equal sample size in the two groups. Because $0 < \pi_{BTi} < 1$ and $\pi_{BTi} = \frac{\phi_{Bi}}{1-\phi_{Bi}}\pi_{BCi}$, ϕ_{Bi} has to be $0 < \phi_{Bi} < 0.76$ when π_{BCi} is specified as 0.3. Therefore, we consider ϕ_{Bi} comes from the uniform distribution $U(0, 0.76)$ and the biomarker response rate of the treatment group π_{BTi} is derived when $\pi_{BCi} = 0.3$ in the simulations.

Let D_{pi} be the biomarker response identifier of the p-th patient in the i-th trial and $B_i = \sum_{p=1}^{n_i}(D_{pi} = 1)$. Here, $D_{pi} = 1$ when the biomarker responds; otherwise, $D_{pi} = 0$. It is randomly generated from $D_{pi} \sim Bernoulli(\pi_{BTi})$ for the treatment group and $D_{pi} \sim Bernoulli(\pi_{BCi})$ for the control group.

Let ρ be the concordance index between C_{pi} and D_{pi}, where C_{pi} denotes the clinical response identifier of the

p-th patient in the i-th trial and $X_i = \sum_{p=1}^{n_i}(C_{pi} = 1)$. ρ is from the distribution of $Bernoulli(\text{PPV})$ for $D_{pi} = 1$ and $Bernoulli(\text{NPV})$ for $D_{pi} = 0$. The patient-level data of clinical endpoint C_{pi} could be derived by using

$$\text{when } D_{pi} = 1, \; C_{pi} = \begin{cases} 1, & \text{if } \rho = 1 \\ 0, & \text{if } \rho = 0 \end{cases};$$

(8)

$$\text{when } D_{pi} = 0, \; C_{pi} = \begin{cases} 1, & \text{if } \rho = 0 \\ 0, & \text{if } \rho = 1 \end{cases}.$$

Finally, the summary statistics of each historical trial, B_i, B_{Ti}, X_i and X_{Ti}, are derived from the patient-level data and employed to build the Bayesian model by using formulas (1), (5) and (6).

The calculation of true values

For a new trial j with sample size of 400, given PPV, NPV and π_{BTj}, the true value of the clinical response rate of the treatment group is calculated by using

$$\pi_{XTj} = \pi_{BTj} \times \text{PPV} + (1-\pi_{BTj}) \times (1-\text{NPV}). \quad (9)$$

Equally, the true value of clinical response rate of control group π_{XCj} is calculated and the true value of RR_j is derived from $RR_j = \pi_{XTj}/\pi_{XCj}$.

Measures of predictive ability

The accuracy and robustness of RR prediction is vital to measure the predictive ability of the proposed Bayesian model. But the traditional measures, for example, bias, root mean square error (RMSE), *etcetera*, cannot be applied here because the endpoint is binary. The log transformation is considered to adjust the non-normality of the binary endpoint. The modified bias and modified RMSE are proposed to measure the accuracy of the RR prediction. For the new trial j, when K predicted values of RR_j are obtained, they are calculated by using

$$\text{Modified bias} = \left(\log^{-1}\left(\frac{\sum_{k=1}^{K}\left(\log RR_{kj}^{P} - \log RR_{j}^{True}\right)}{K} \right) - 1 \right) \times RR_{j}^{True}$$

(10)

and

$$\text{Modified RMSE} = \left(\log^{-1} \sqrt{ \frac{\sum_{k=1}^{K} \left(\log RR_{kj}^{P} - \log RR_{j}^{True} \right)^2}{K} } - 1 \right) \times RR_{j}^{True},$$

(11)

in which RR_{j}^{True} denotes the true value of RR_j and RR_{kj}^{P} is the k-th predicted value of RR_j. Similarly, modified RMSE includes the effect of bias direction to estimate prediction error compared with modified bias. The RR prediction is more accurate when the modified bias and the modified RMSE approach to 0.

To measure the precision of the RR prediction, the average width of the 95% CIs is calculated by using

$$\text{Average width of 95\% CIs} = \log^{-1} \left(\frac{\sum_{k=1}^{K} \left(\log RR_{kjL}^{P} - \log RR_{kjU}^{P} \right)}{K} \right)$$

(12)

where RR_{kjU}^{P} and RR_{kjL}^{P} are the upper and lower bound of 95% CI of the k-th RR_j prediction respectively. When the average width of 95% CIs approaches to 1, the better the precision of RR prediction is.

Results

Simulation I: comparison of Bayesian model with different link functions and positive predictive values/negative predictive values

Three link functions (the odds, logit and cloglog functions) are compared in this section. The ϕ_{Bj} is assumed to be 0.1, 0.3, and 0.5. Ten historical trials are generated randomly from patient level and their trial-level summary data are included for Bayesian model building. Modified bias and modified RMSE are calculated to measure the accuracy of RR prediction, and average width of 95% CIs is estimated to evaluate the precision. Both scenarios, (i) equal PPV and NPV and (ii) higher NPV and lower PPV, are considered here.

Scenario I: equal PPV and NPV

In the simulations, PPV/NPV is assumed to take the values of 0.1, 0.25, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 0.99. Figure 2 depicts the change of modified bias, modified RMSE and average width of 95% CIs for different ϕ_{Bj} and link functions. No matter which link function is employed, the modified bias and the modified RMSE are both worse when PPV/NPV <0.5. The modified RMSE fails to be estimated for the logit link function when PPV/NPV <0.5 and ϕ_{Bj} ≤0.3 because of the poor prediction. The other two link functions have the same problems when PPV/NPV = 0.1 and ϕ_{Bj} = 0.1. On the other hand, the
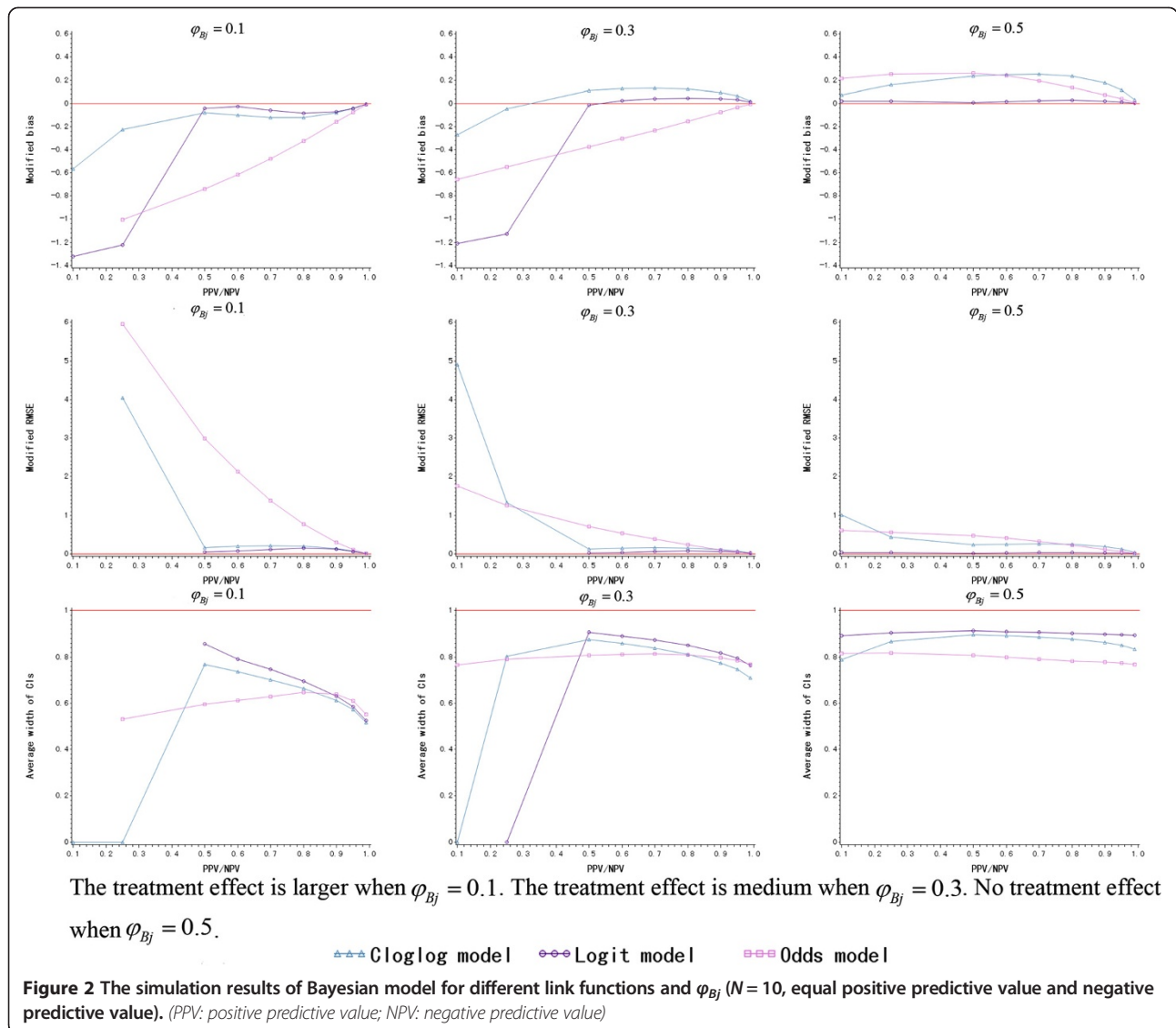
average width of the 95% CIs is not good when PPV/NPV <0.5, especially when ϕ_{Bj} <0.5. Therefore, PPV/NPV ≥0.5 is an indispensable condition of the proposed Bayesian model for RR prediction under the assumption of equal NPV and PPV.

Among the three link functions, both modified bias and modified RMSE of the logit function are the closest to zero when PPV/NPV ≥0.5, and the cloglog link function is better than the odds function. For the logit link function, modified bias and modified RMSE have a little inflation when PPV/NPV = 0.8 and approach to zero when PPV/NPV continues increasing. The accuracy of the prediction remains good, however, if ϕ_{Bj} varies when PPV/NPV ≥0.5 and logit link function is employed. Regarding the precision of the RR prediction, the logit link function has the narrowest average width of the 95% CIs, which increases and approaches to the other two link functions when PPV/NPV increases. As ϕ_{Bj} increases, the treatment effect becomes smaller and the average width of 95% CIs of the logit link function decreases. The logit link function is considered as the first choice for building the Bayesian model from the perspective of the accuracy and precision of RR prediction. The detailed simulation results are presented in Additional file 1: Table S1.

Scenario II: unequal positive predictive value and negative predictive value

In this scenario, NPV is considered to be 0.99, 0.95 and 0.9, and PPV is not larger than 0.5. The values of PPV and NPV are listed in Additional file 1: Table S2. As is depicted in Figure 3, the logit and cloglog link function have the better prediction accuracy on modified bias and modified RMSE when PPV + NPV ≥1. But the odds link function leads to an inaccurate prediction, especially when ϕ_{Bj} = 0.1. As ϕ_{Bj} increases, the modified bias of the cloglog link function is rising and deviates from zero when ϕ_{Bj} ≥0.3 though the modified RMSE of cloglog one still looks fine. However, the modified bias of the logit link function always fluctuates around zero regardless of how ϕ_{Bj} varies.

On the other hand, the average width of the 95% CIs of the logit link function is a little narrower than cloglog one when PPV/NPV remains the same and PPV + NPV ≥1. Considering both the accuracy and precision of the prediction, the logit link function is also the optimal one to build the Bayesian model when PPV + NPV ≥1 under the assumption of higher NPV and lower PPV. Furthermore, the average width of the 95% CIs of the logit link function is a little larger when ϕ_{Bj} = 0.1 and becomes smaller when ϕ_{Bj} increases. When PPV is fixed, the logit function has the narrower average width of 95% CIs, as NPV is smaller.

The treatment effect is larger when $\varphi_{Bj} = 0.1$. The treatment effect is medium when $\varphi_{Bj} = 0.3$. No treatment effect when $\varphi_{Bj} = 0.5$.

△△△ Cloglog model   ◇◇◇ Logit model   ⊞⊞⊞ Odds model

Figure 2 The simulation results of Bayesian model for different link functions and φ_{Bj} ($N = 10$, equal positive predictive value and negative predictive value). (PPV: positive predictive value; NPV: negative predictive value)
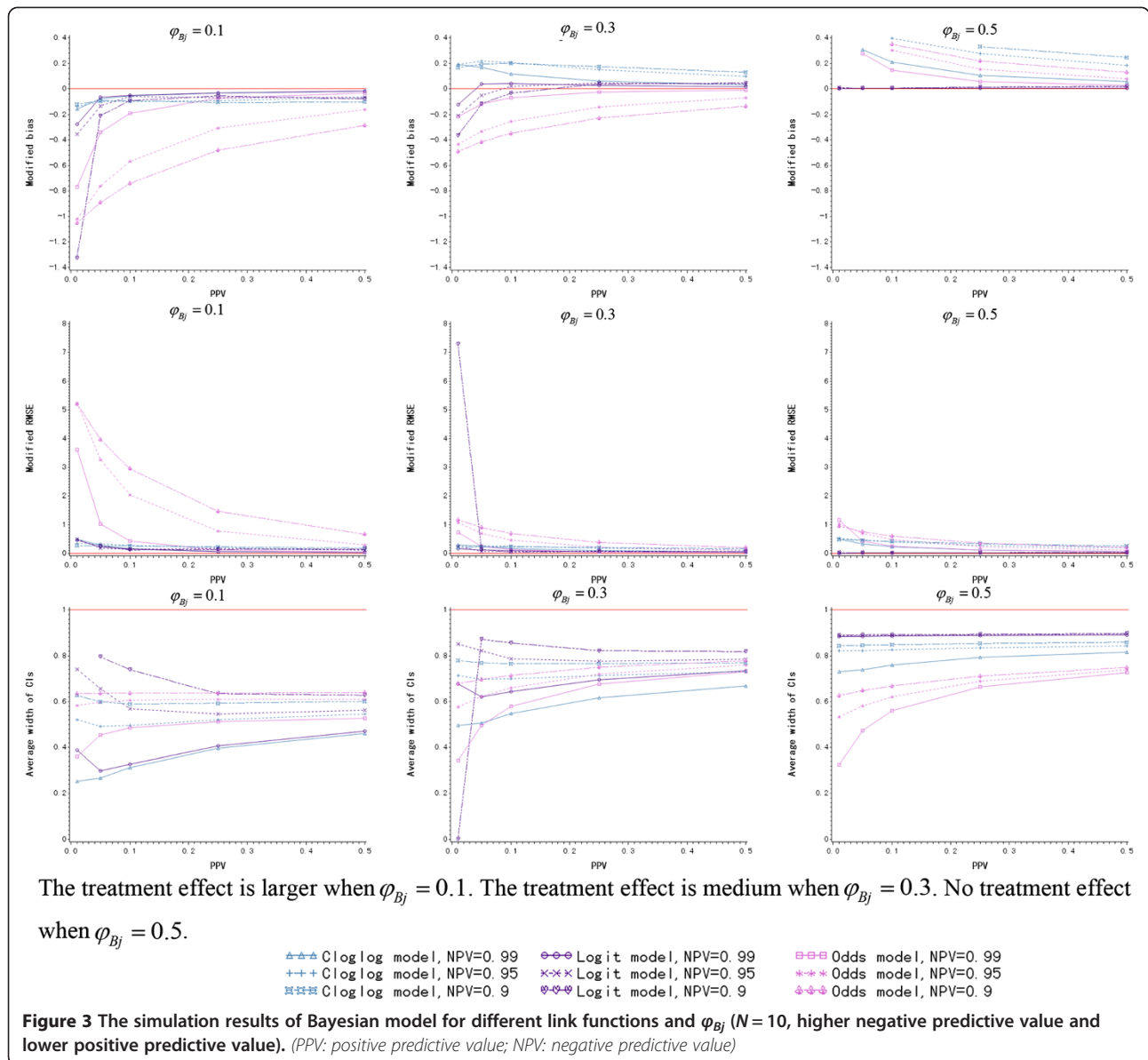
Simulation II: the number of historical trials to be included for model building

The number of historical trials to be included for model building is a key issue to meta-analytic approaches. In this section, 5, 10, 20, 30 and 50 historical trials are assumed to build the Bayesian prediction model, and their predictive abilities are compared under (i) equal PPV and NPV and (ii) higher NPV and lower PPV. Based on the simulation results of the last section, the logit link function is employed to build the Bayesian model here. The observed ϕ_{Bj} of the new trial is still assumed to be 0.1, 0.3, and 0.5.

Scenario I: equal positive predictive value and negative predictive value

Based on the results of the last section, PPV/NPV ≥0.5 is considered here. The values of PPV/NPV are listed in

Additional file 1: Table S3. As is seen in Figure 4, the prediction is a little underestimated when $\phi_{Bj} = 0.1$. The modified bias increases and even deviates from zero positively with the rise of ϕ_{Bj}. It varies within (-0.08,0.08) regardless of how ϕ_{Bj} changes. The modified RMSE is the smallest when 50 trials are included in model building for $\phi_{Bj} = 0.1$. But as ϕ_{Bj} increases, the larger N does not always bring an accurate prediction. For example, the model built from five trials has the smallest modified RMSE when $\phi_{Bj} = 0.5$. It is perhaps because there is no treatment effect when $\phi_{Bj} = 0.5$, and more historical trials include larger variations in the model, which reduces the accuracy of the prediction. Regarding the precision of *RR* prediction, the average width of the 95% CIs is smaller when N is larger and ϕ_{Bj} increases. The detailed simulation results are presented in Additional file 1: Table S3.
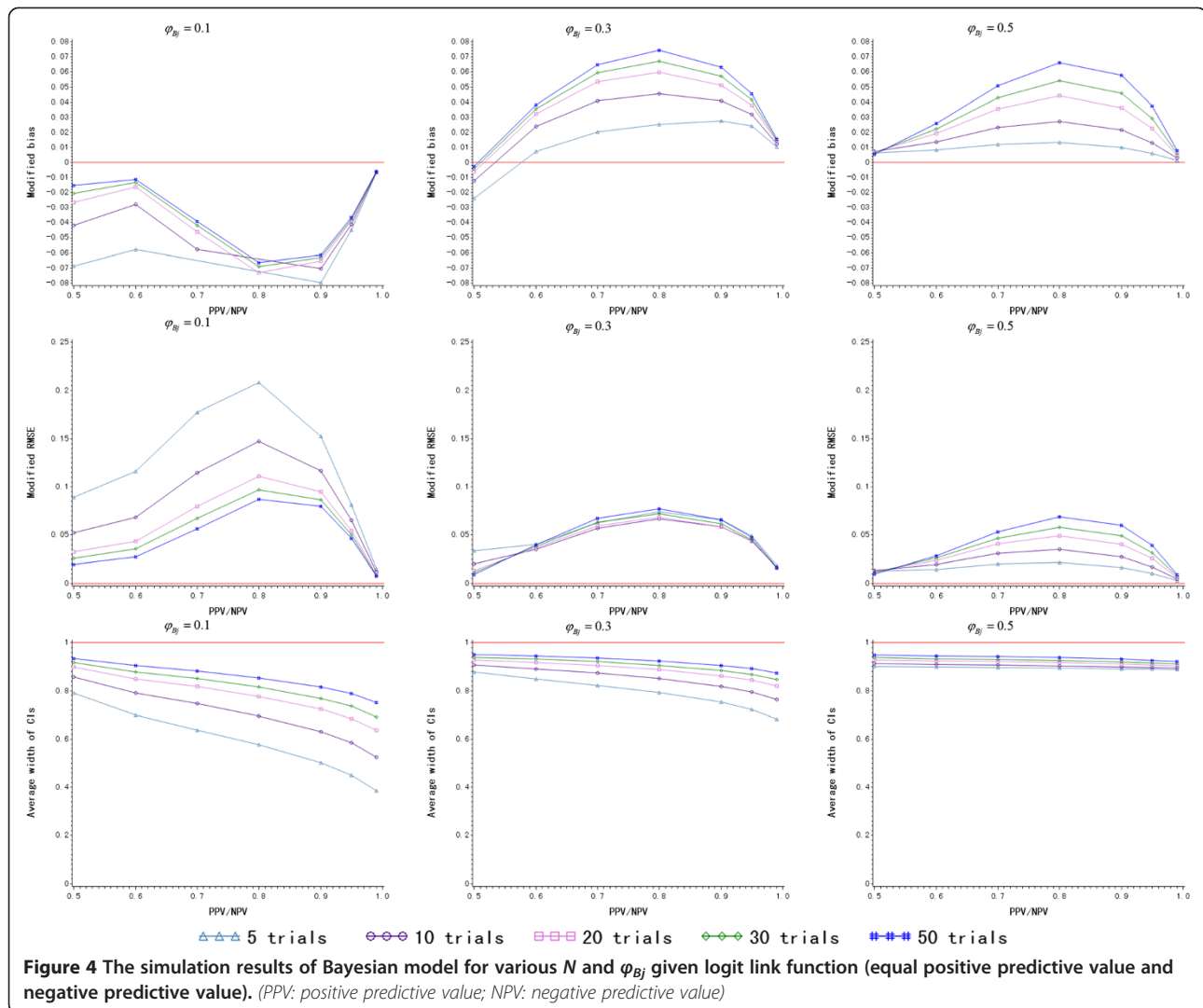
The treatment effect is larger when $\varphi_{Bj} = 0.1$. The treatment effect is medium when $\varphi_{Bj} = 0.3$. No treatment effect when $\varphi_{Bj} = 0.5$.

Figure 3 The simulation results of Bayesian model for different link functions and φ_{Bj} ($N = 10$, higher negative predictive value and lower positive predictive value). *(PPV: positive predictive value; NPV: negative predictive value)*

When $N = 20$, the maximum modified RMSE is only about 0.1, and the average width of the 95% CIs is also larger than 0.6 even when $\phi_{Bj} = 0.1$. Therefore, the minimum number of historical trials to be included for model building is proposed to be 20 for an accurate and precise prediction when PPV and NPV are equal.

Scenario II: unequal positive predictive value and negative predictive value

In the simulations, NPV is considered to be 0.99, 0.95 and 0.9 and PPV ≤0.5. As long as PPV + NPV ≥1 is satisfied, the accuracy of the prediction is good on both modified bias and modified RMSE. They both converge to zero when N and ϕ_{Bj} increase. The larger PPV + NPV is, the more accurate the prediction. Even when $N = 5$,

the modified RMSE is smaller than 0.4 for $\phi_{Bj} = 0.1$. Therefore, five historical trials are enough to build a model that has an accurate prediction.

On the other hand, the larger N leads to the smaller average width of 95% CIs, which is shown in Figure 5. When PPV and ϕ_{Bj} is fixed, the average width of 95% CIs rises with the increase of NPV. When $N = 5$ and $\phi_{Bj} = 0.1$, the minimum average width of 95% CIs is only about 0.2. The 95% CIs of the *RR* prediction is wide and the Bayesian model brings a conservative interval estimate. In conclusion, when a higher NPV and a lower PPV are considered, the Bayesian model built from five historical trials is enough for an accurate point estimate of the prediction, but leads to a conservative interval estimate. If a precise interval estimate of *RR* prediction is expected, $N = 30$ is

Figure 4 The simulation results of Bayesian model for various N and φ_{Bj} given logit link function (equal positive predictive value and negative predictive value). *(PPV: positive predictive value; NPV: negative predictive value)*

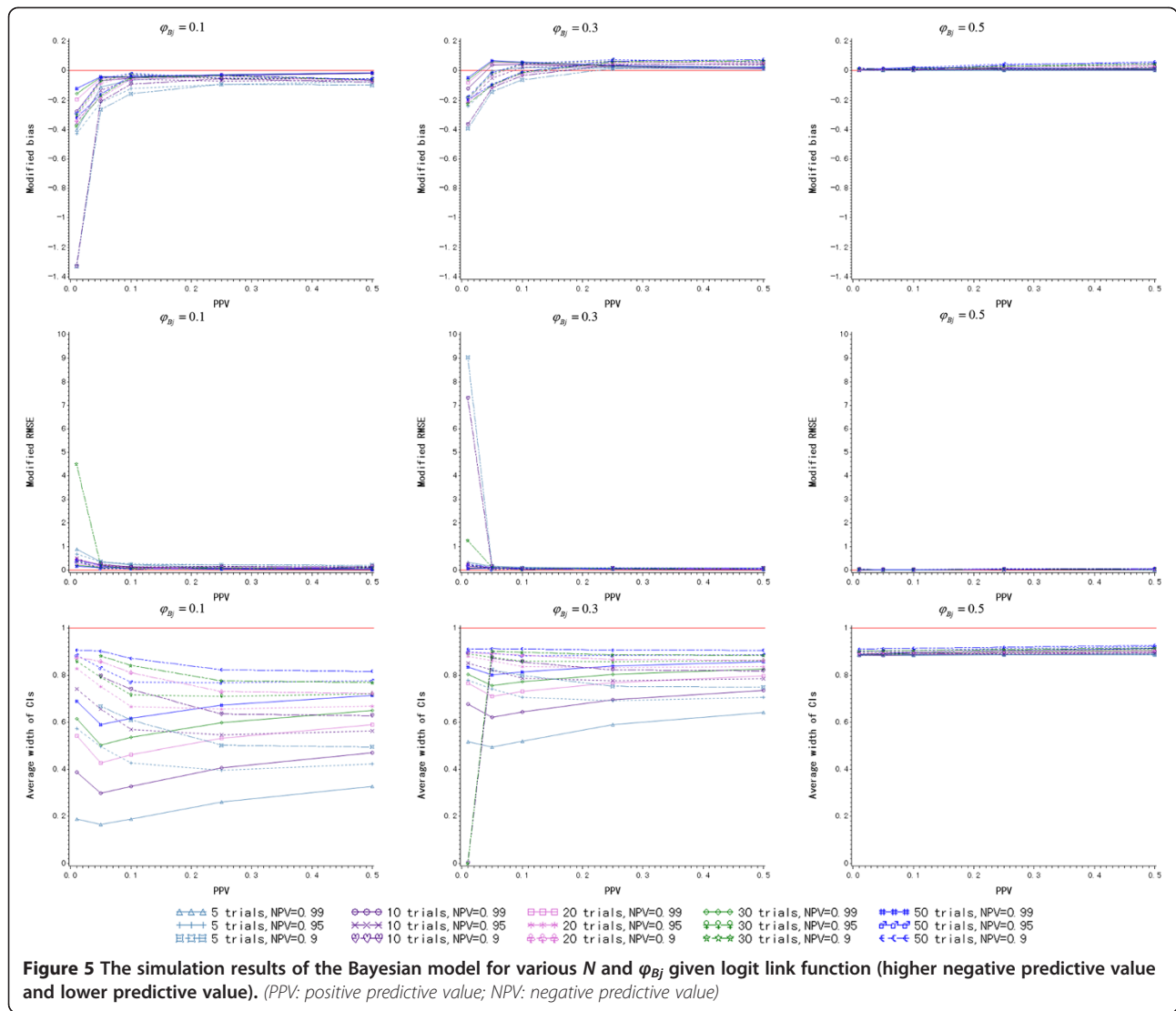proposed where the average width of 95% CIs is larger than 0.5 even when $\phi_{Bj} = 0.1$.

Hypothetical example

To evaluate the efficacy of a vaccine in clinical trial, the incidence rate of the disease of interest is usually observed. Vaccine efficacy (VE) is estimated by using VE = $1 - \pi_T/\pi_C = 1 - RR$, in which π_T and π_C denote the incidence rate of the disease in the vaccine and control group. However, the occurrence of disease requires too long a follow-up time. Virus infection is considered as an early biomarker of the disease. It has an NPV of 99.99% and PPV of 15% for predicting the occurrence of the disease. VE is just a transformation of RR. It is possible to predict vaccine efficacy earlier when the infection cases are observed by employing our proposed approach.

It was hypothesized that five global placebo-controlled vaccine trials with sample size ratio of 1:1 completed in the United States, Europe and Japan. The number of

disease cases and infection cases in the vaccine and placebo group in each trial are listed in Table 1. Recently, a new placebo-controlled trial was conducted in China to evaluate VE in the Chinese population. A total of 3606 subjects in two groups of equal size were recruited. A total of 36 infection cases were observed, with 7 cases from the vaccine group. No disease was observed, and we wanted to evaluate vaccine efficacy earlier based on the infection cases.

It is presumed that there is no ethnic difference in the association between infection and disease occurrence even if the vaccine efficacy is different between Chinese and other ethnic groups. The proposed approach can be applied to build the Bayesian model from historical vaccine trials and to predict VE from the observed infection cases in the Chinese population. A Bayesian model built from five historical vaccine trials in the United States, Europe and Japan could bring an accurate point estimate and a conservative interval estimate of VE prediction according

Figure 5 The simulation results of the Bayesian model for various N and φ_{Bj} given logit link function (higher negative predictive value and lower predictive value). *(PPV: positive predictive value; NPV: negative predictive value)*

to the simulation results in last section. Logit link function is employed in the Bayesian model. The VE on disease in the Chinese population is predicted as 81.84% with 95% CI of (53.98%,97.54%). Consequently, the vaccine is considered to have significant efficacy in the Chinese population because the lower boundary of the conservative 95% CI is still larger than zero.

Discussion

The proposed Bayesian prediction model between a bio-marker and the clinical outcome in this article is one of meta-analytic approaches. It focuses on the prediction of clinical outcome based on biomarker, not the replacement as a surrogate. According to Prentice criterion [7], the proposed Bayesian model requires that the association

Table 1 Five hypothetical historical vaccine trials

	N	Number of infection cases		Number of disease cases		Vaccine efficiency (95% CI) on disease
		Vaccine	Placebo	Vaccine	Placebo	
1	2954	2	213	0	23	100.0%(82.5%,100.0%)
2	3200	10	86	1	6	83.3%(-37.6%,99.6%)
3	529	2	45	0	2	100.0%(-431.4%,100%)
4	3674	27	130	3	8	62.7%(-55.5%,93.6%)
5	3452	24	296	0	30	100.0%(87.1%,100.0%)

CI: credible interval.

between the biomarker and the clinical endpoint be equal across the two groups. Compared with the Daniels and Hughes method, the proposed Bayesian model is a complete meta-analytic approach, which only requires the accessible trial-level summary data of previous trials, and no patient-level data is necessary. Nikolakopoulos *et al.* [27] also employed a Bayesian method to predict clinical endpoint based on observed biomarker for phase II trial decision making. But there are important differences between the two approaches. First, the model between a biomarker and the clinical endpoint must be known for clinical outcome prediction, and how to build the model from previous studies was not considered in [27]. But it is part of our work. Both model building from historical trials and model prediction for a new trial are considered here. Second, as a meta-analytic approach, our proposed model only involves trial-level summary data of historical studies for model building and predicts clinical endpoint with trial-level summary biomarker data for the new trial. The patient-level data of the new trial is not necessary for the prediction. PPV and NPV are not involved in model building and model prediction, but are needed for model evaluation. Third, the proposed prediction model is also different. Here, we propose a generalized linear model to describe the association between the biomarker and the clinical endpoint. We only consider the prior information on the model parameter β, β_0 and ϕ_{Bi} for model building, but not place prior information on the clinical endpoint in the prediction. Finally, the scenario of unequal PPV and NPV between the biomarker and the clinical endpoint, which is common in vaccine clinical trials, is discussed in this article. On the other hand, the proposed Bayesian model describes the relationship between biomarker and clinical endpoint regardless of treatment effect. The historical trials for model building do not have to contain the same treatment and control group. The observational studies are also available for model building as long as both biomarker and clinical endpoint are collected in the studies. It further increases the availability of historical studies and makes it easy to be applied in practical.

In the proposed Bayesian model, ϕ_{Bi} and ϕ_{Xi} are employed to measure the treatment difference of biomarker and clinical outcome between the two groups and connected with formula (1). That is because this paper is to predict the rate ratio of clinical outcome for binary endpoint. ϕ_{Bi} and ϕ_{Xi}, which are usually considered in vaccine clinical trials, have a direct connection with RR that is shown in formula (7) and does not depend on trial size. The association between the biomarker and clinical endpoint could be well described in the model in virtue of ϕ_{Bi} and ϕ_{Xi}. When unequal sample size in the two groups is considered, ϕ_{Bi} is still calculated by $\phi_{Bi} = \pi_{BTi}/(\pi_{BTi} + \pi_{BCi})$, but estimated by $\hat{\phi}_{Bi} = B_{Ti}/(B_{Ti} + B_{Ci} \times R)$

where R is the sample size ratio of treatment and control group. Equally, ϕ_{Xi} is estimated by $\hat{\phi}_{Xi} = X_{Ti}/(X_{Ti} + X_{Ci} \times R)$. Therefore, the proposed Bayesian model is still applicable for unequal sample size in two groups.

Though multiple meta-analytic models between biomarker and clinical outcome were proposed, few of them evaluated the model in a broad sense and explicitly describe the application guidance. The direct assumption on trial-level data in Baker's simulation study [17] possibly influences the evaluation. The association model may be over-evaluated because the information loss from the trial-level data was ignored. R^2_{trial}, introduced by Buyse *et al.* [14], is difficult to interpret, and it is difficult to define the application condition clearly. To overcome both problems, we make assumption on the patient-level data of historical trials and perform the simulations to evaluate the meta-analytic model.

Both scenarios, (i) equal PPV and NPV and (ii) higher NPV and lower PPV, are considered in the simulations from the practical perspective. In clinical diagnosis, higher PPV and NPV are usually required for a biomarker to detect the potential disease earlier [27,28]. But the biomarker with a higher NPV and lower PPV is also common in vaccine studies. According to the simulation results, the proposed Bayesian model leads to a good prediction of clinical RR based on a biomarker when both PPV and NPV are larger than 0.5. For higher NPV and lower PPV, the model makes sense of the prediction when PPV + NPV ≥1. In an actual trial, the exact values of NPV and PPV between biomarker and clinical outcome are generally unknown. But it is possible to estimate PPV and NPV within a range by clinicians from the medical mechanism and clinical experiences and evaluate if the model is applicable. Furthermore, the logit link function is better than other functions in model building from the point of accuracy and precision of RR prediction. But for a specific trial, an extensive simulation for the trial is recommended to choose the optimal link function for model building to satisfy the demand of the trial.

Regarding the minimum number of historical trials to be included in model building, it is advised that 20 historical trials be enough to build a model that predicts clinical RR accurately and precisely for equal PPV and NPV. When higher NPV and lower PPV is considered, the model built from five historical trials is able to lead to an accurate point estimate of the prediction, but a conservative interval estimate. If a more precise prediction is demanded, a larger N is required. However, if too low biomarker and clinical response rate is expected in all historical trials, more historical trials are proposed to be included in model building in order to describe the association between biomarker and clinical outcome accurately and well predict the clinical outcome from biomarker in the new trial. A simulation is also proposed

to evaluate the minimum number of historical trials for a specific trial.

The proposed Bayesian model for clinical outcome prediction based on biomarker in this paper is not only for binary endpoint, but also continuous variables as long as a suitable link function is given. The proposed method has extensive potential applications in drug development. The example shows an application in global drug development. As long as an equal association between the biomarker and the clinical endpoint across different ethnic groups is considered from the medical point, it is possible to 'bridge' the relationship from historical global studies to a new regional trial in virtue of the proposed Bayesian model and then predict clinical endpoint in the new region. It is different from the traditional bridging study, which bridges the treatment effect from historical trials, and equally enhances the efficiency of regional trial in another way. On the other hand, the early prediction of clinical outcome with the help of the Bayesian model built from the historical studies is able to help with making a go/no-go decision in new drug development. However, a few practical problems, for example, sample size estimation of regional trial involving biomarker, go/no-go decision rule based on biomarker, *etcetera*, will be encountered in the application of the proposed Bayesian model. They are not in the scope of this article and will be discussed in further studies.

Conclusions

A Bayesian prediction model between a biomarker and the clinical outcome is proposed in this paper. It is a complete meta-analytic approach and only requires trial-level data in model building. It is able to predict well the clinical outcome from an observed biomarker when PPV/NPV ≥0.5 for equal PPV and NPV and when PPV + NPV ≥1 for higher NPV and lower PPV. The Logit link function is preferred in both scenarios. The minimum number of historical trials to be included in model building is proposed to be 20 when PPV and NPV are considered to be equal. For higher NPV and lower PPV, the Bayesian model from five historical trials could lead to an accurate point estimate, but conservative interval estimate of the prediction. The proposed model has potential applications in decision making of new drug development and global-regional drug development program. But the practical problems have to be discussed in further studies.

Additional file

Additional file 1: Table S1. The simulation results of the Bayesian model for different link functions and φ_{Bj} under the assumption of equal PPV and NPV ($N = 10$). **Table S2.** The simulation results of the Bayesian model for the different link functions and φ_{Bj} under the assumption of higher negative predictive value (NPV) and lower positive predictive

value (PPV) ($N = 10$). **Table S3.** The simulation results of the Bayesian model for the various N and φ_{Bj} given logit link function and equal PPV and NPV. **Table S4.** The simulation results of the Bayesian model for the various N and φ_{Bj} given logit link function and equal higher NPV and lower PPV. **Table S5.** Table of notations in the article.

Abbreviations
NPV: negative predictive value; PPV: positive predictive value; RMSE: root mean square error; RR: rate ratio; VE: vaccine efficacy; HPV: human papillomavirus; MCMC: Markov Chain Monte Carlo; CI: credible interval; RMSE: root mean square error.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
ZJ, YS and WW conceived the study concept and designed the simulation methods. ZJ and JX performed the simulations. All authors contributed to the interpretation of the simulation study. ZJ and QS performed the analysis of the example. All authors revised and approved the submitted manuscript.

Author details
[1]Department of Health Statistics, School of Preventive Medicine, Fourth Military Medical University, No. 169 Changle West Road, Xi'an, Shaanxi, China. [2]Biostatistics and Research Decision Science, Merck Research Laboratory, Merck & Co., Inc., 2/F Building 22, Universal Business Park, No. 10 Jiuxianqiao Road, Beijing, China.

References
1. Boissel JP, Collet JP, Moleur P, Haugh M: **Surrogate endpoints: a basis for a rational approach.** *Eur J Clin Pharmacol* 1992, **43**:235–244.
2. De Gruttola V, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, Gail MH, Prentice R, Wittes J, Zeger SL: **Considerations in the evaluation of surrogate endpoint in clinical trials: summary of a National Institutes of Health workgroup.** *Control Clin Trials* 2001, **22**:485–502.
3. Chuang-Stein C, DeMasi R: **Surrogate endpoints in AIDS drug development: current status.** *Drug Inform J* 1998, **32**:439–448.
4. Biomarkers Definitions Working Group: **Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.** *Clin Pharmacol Ther* 2001, **69**:89–95.
5. US Department of Health and Human Services: *FDA: The Nation's Premier Consumer Health Protection Agency.* Washington DC: US Food and Drug Administration; 2004.
6. US Department of Health and Human Services: *Guidance For Industry: Fast Track Drug Development Programs-Designation, Development, And Application Review.* Washington DC: US Food and Drug Administration; 2004.
7. Prentice RL: **Surrogate endpoints in clinical trials: definitions and operational bias.** *Stat Med* 1989, **8**:431–440.
8. Freedman LS, Graubard BI, Schatzkin A: **Statistical validation of intermediate endpoints for chronic diseases.** *Stat Med* 1992, **11**:167–178.
9. Weir CJ, Walley RJ: **Statistical evaluation of biomarker as surrogate endpoints: a literature review.** *Stat Med* 2006, **25**:183–203.
10. Shi Q, Sargent DJ: **Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials.** *Int J Clin Oncol* 2009, **14**:102–111.
11. Fleming TR: **Surrogate endpoints in clinical trials: definition and operational criteria.** *Stat Med* 1994, **8**:431–440.
12. Hughes MD, DeGruttola V, Welles SL: **Evaluating surrogate markers.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, **10**:S1–S8.
13. Daniels MJ, Hughes MD: **Meta-analysis for the evaluation of potential surrogate markers.** *Stat Med* 1997, **16**:1865–1982.

14. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H: **The validation of surrogate endpoints in meta-analyses of randomized experiments.** *Biostatistics* 2000, **1**:49–67.
15. Gail MH, Pfeiffer R, Van Houwelingen HC, Carroll RJ: **On meta-analytic assessment of surrogate outcomes.** *Biostatistics* 2000, **1**:231–246.
16. Korn EL, Albert PS, McShane LM: **Assessing surrogates as trial endpoints using mixed models.** *Stat Med* 2005, **24**:163–182.
17. Baker SG: **A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint.** *Biostatistics* 2006, **7**:58–70.
18. Baker SG: **Two simple approaches for validating a binary surrogate endpoint using data from multiple trials.** *Stat Methods Med Res* 2008, **17**:505–514.
19. Baker SG, Sargent DJ, Buyse M, Burzykowski T: **Predicting treatment effect from surrogate endpoints and historical trials: an extrapolation involving probabilities of a binary outcome or survival to a specific time.** *Biometrics* 2012, **68**:248–257.
20. Burzykowski T, Buyse M: **Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation.** *Pharm Stat* 2006, **5**:173–186.
21. Rolan P: **The contribution of clinical pharmacology surrogates and models to drug development – a critical appraisal.** *Br J Clin Pharmacol* 1997, **44**:219–225.
22. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use: **Statistical principles for clinical trials.** *Stat Med* 1999, **18**:1905–1942.
23. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, Strand V, Conaghan PG, Ostergaard M, Maksymowych WP, Landewe R, Bresnihan B, Tak PP, Wakefield R, Mease P, Bingham CO 3rd, Hughes M, Altman D, Buyse M, Galbraith S, Wells G: **Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema.** *J Rheumatol* 2007, **34**:607–615.
24. Van Walraven C, Oake N, Coyle D, Taljaard M, Forster AJ: **Changes in surrogate outcomes can be translated into clinical outcomes using a Monte Carlo model.** *J Clin Epidemiol* 2009, **62**:1306–1315.
25. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, Gobburu J: **Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development.** *Clin Pharmacol Ther* 2009, **86**:167–174.
26. Van Houwelingen HC, Arends LR, Stijnen T: **Advanced methods in meta-analysis: multivariate approach and meta-regression.** *Stat Med* 2002, **21**:589–624.
27. Nikolakopoulos S, van der Wal WM, Roes K: **An analytical approach to assess the predictive value of biomarkers in phase II decision making.** *J Biopharm Stat* 2013, **23**:1106–1123.
28. Thoresen M, Hellstrom-Westas L, Liu X, de Vries LS: **Effect of hypothermia on amplitude-integrated electroencephalogram in infants with asphyxia.** *Pediatrics* 2010, **126**:131–139.