Performance of Bayesian Priors in Validation of Correlate of Protection for High Efficacy Vaccine Trials

Igwebuike Enweonye and Edith Uzoma Umeh
Department of Statistics, Nnamdi Azikiwe University, PMB 5025, AWKA.

ABSTRACT

Although the use of intermediate clinical endpoint or correlate of protection (CoP) has increased over the years, the validation of CoP for high efficacy vaccine trials has remained a challenge due to sparse data and conventional statistical methods which are not adequate. Be it in the frequentist or the Bayesian world, the meta-analytic approach is a well-accepted method of CoP validation. However, the full joint bivariate models suffer computational issues. And there is a push for the use of individual level instead of aggregate data in validation process. In this quest, the Bayesian approach is emerging as the future as regards the validation of CoP but one recurring criticism about this method is its application of prior distributions. To elucidate which makes better sense, in the context of CoP validation, the non-informative (NIP) and weakly informative prior (WIP) distributions are compared in a meta-analytic approach using simulated data. It was found that, 1) there are no convergence issues when either of the models are used, 2) WIP models take about 20% longer time than NIP models to converge, and 3) the NIP models consistently perform better than the WIP models.

Keywords: Validation, Correlate of protection, Clinical endpoint, Bayesian Hierarchical Modelling, prior distribution

1. Introduction

Vaccines are mostly given as prophylactics - of which the true clinical endpoint is difficult to measure if there is no disease outbreak. It turns out that the development and regulatory approval of vaccines relies largely on the immunogenicity data. The *protective threshold* of a vaccine is desirable in identifying the level of an immune marker above which vaccinees have a defined probability of being protected, and to make a statement over the vaccine efficacy. Such quantity defines the *vaccine response threshold*, used to calculate the *response rate* (Voysey et al. 2018). Sadly, during clinical development, vaccine correlate of protection (CoP) is generally unknown (Callegaro and Tibaldi, 2019).

CoP are used in lieu (Buyse and Molenberghs, 1998) when clinical endpoints of primary interest are hard or unethical to measure. CoP is useful because it can be measured earlier, more conveniently, or more frequently than the true endpoint (Ellenberg and Hamilton, 1989). Its use in clinical studies has increased, necessitating the development of sound statistical methods in its validation process (Burzykowski et al., 2006).

Health authorities around the world are opening doors to CoP, for example, between 2010 and 2012, the United States Food and Drug Administration (US FDA) approved 45 percent of new drugs applications based on various surrogate endpoints (FDA, 2018). A beneficial correlate of protection generally allows for more efficient drug development programs (FDA, 2018).

The Bayesian statistics provide a flexible tool for complex applications including the validation of correlate of protection. The beauty of Bayesian inference lies in the prior distribution which is its backbone, although, it has caused controversies among the scientific community; with some arguing that prior distribution introduces external information from the data. This paper compares the performances of the non-informative (NIP) and weakly informative (WIP) prior distributions in a meta-analytic approach for validation of CoP using Gibbs sampler.

Gibbs Sampling is Markov Chain Monte Carlo (MCMC) method which involves successive sampling from the complete conditional densities. For the working of MCMC algorithm, we refer to Gelman and Shalizi (2013), Gelman et al. (2013), Gallager (2013) and Congdon (2014).

2. Study Design and Methods

True clinical and CoP endpoints were simulated using data from a previous efficacy clinical trials as starting values. Each data set consists 50 trials characterised by a 1:1 randomization of size of 100 subjects per trial leading to 5,000 subjects for each data set. Bayesian hierarchical model using Markov Chain Monte Carlo (MCMC) was applied to each of the simulated data sets. In turn the models combined non-informative prior and then weakly informative prior distributions with simulated data to obtain posterior information for inferences. Each MCMC model has 3 parallel chains with adaptation at every 1,000 simulation steps. The final inference used 10,000 draws while 1,000 draws were discarded as burn-in samples. Standard inference calls R to run the model through Just another Gibbs Sampler (JAGS) and extract predicted values for the monitored parameters: variance-covariance matrix D between random treatment effects of the true and CoP endpoints and coefficient of determination, R^2.

Let S_{ij} and T_{ij} represent the continuous and binary underlying values of the CoP and the true endpoints, respectively, for subject j in trial i and Z_{ij} an indicator for treatment effect. That a subject j in trial i has the disease is depicted by $T_{ij} = 1$. And further consider the meta-analytic framework in the single trial setting, in which the units are randomized subgroups of trials. At the first level of the hierarchical Bayesian meta-analytic approach, assuming full mediation, reduced bivariate models are specified as follow:

$$S_{ij} = \alpha_0 + (\alpha_1 + a_i)Z_{ij} + e_{ij} \quad (1)$$

$$logit\, T_{ij} = \beta_0 + (\beta_1 + b_i)Z_{ij} \quad (2)$$

Where,

$$\begin{pmatrix} S_{ij} \\ logit\, T_{ij} \end{pmatrix} \sim N\left[\begin{pmatrix} \alpha_0 + (\alpha_1)Z_{ij} \\ \beta_0 + (\beta_1)Z_{ij} \end{pmatrix}, \Sigma\right] \quad (3)$$

And,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D\right], D = \begin{pmatrix} d_{aa} & d_{ab} \\ d_{ab} & d_{bb} \end{pmatrix} \quad (4)$$

$\alpha's$ and $\beta's$ are fixed effects of treatment Z_{ij} on the endpoints in trial i, a_i and b_i are the trial specific random effects of treatment on the endpoints. The error structure e_{ij} are CoP associated normally distributed random error terms with zero mean and variance δ^2. Further $\sum$ is the joint error structure of the fixed effects.

Provided the variance-covariance matrix D of the random treatment effects of endpoints in eqn. (4) is positive definite, the validation is captured by means of the trial-level coefficient of determination,

$$R^2_{trial(r)} = R^2_{b_i|a_i} = \frac{d^2_{ab}}{d_{aa}\, d_{bb}} \quad (5)$$

At the second level of the hierarchical model, hyper priors for the fixed effects are specified. For NIP models:

$$\mu_s \sim N\left(0, \delta^2_{\mu s}\right),$$

$$\alpha_0 \sim N(0, \tau^2_{\alpha 0}),$$

$$\alpha_1 \sim N(0, \tau_{\alpha 1}^2),$$
$$\theta_T \sim Bern(p_T),$$
$$\beta_0 \sim N(0, \tau_{\beta 0}^2),$$
$$\beta_1 \sim N(0, \tau_{\beta 1}^2), \qquad (6)$$
$$\delta_{\mu s}^{-2} \sim U(0,100)$$
$$\tau_{\alpha 0}^{-2} \sim Gamma(10^{-4}, 10^{-4})$$
$$\tau_{\alpha 1}^{-2} \sim Gamma(10^{-4}, 10^{-4}),$$
$$\tau_{\beta 0}^{-2} \sim Gamma(10^{-4}, 10^{-4})$$
$$\tau_{\beta 1}^{-2} \sim Gamma(10^{-4}, 10^{-4}),$$

Next, specify a prior distribution for the association between the treatment effects of the two endpoints and the random effects. As the hyper-prior distribution for the variance-covariance matrices, a Wishart distribution is assumed:

$$D^{-1} \sim Wishart(R_D),$$
$$\sum^{-1} \sim Wishart(R_\sum) \qquad (7)$$

The NIP differs from the WIP model in the assignment of priors for the variance-covariance matrix D of the random effects. All other priors remain same as for both models. However, the difference is huge as the coefficient of determination depends on the matrix. Chung et al. (2015) proposed for WIP model the default prior as a function of the variances d_{aa} and d_{bb} and the correlation between the two varying random effects a_i and b_i given by,

$$p(D)\ \alpha\ |D|^{1/2} = d_{aa}\, d_{bb} \sqrt{1-\rho^2} \qquad (8)$$

The trial-level surrogacy is assessed using the posterior means for R^2 eqn. (5). A sufficiently large R^2 is an indicator of a good surrogate. Beside statistics, clinical and epidemiological judgments, as deemed fit by the experts, are taken into account before a surrogate can be finally adopted.

Modelling was performed in Just another Gibbs Sampler in R (RJAGS) as an interface to JAGS (JAGS 4.3.0 release July 18, 2017). In JAGS there is

no flexibility of specifying any one sampling method rather it runs as a black box and chooses the most efficient sampling method. For details on the working of MCMC algorithms refer to Congdon (2014) and Gelman et al (2013).

3. Results

With a range of vaccine efficacy (VE = 30%-100%), a total of 70 scenarios were simulated in R. The simulated data contain both true binary (protected / not protected) outcome and a continuous immunogenicity values as correlate of protection, using the models eqns. (1) and (2).

Each scenario sample size n=5,000 subjects. Randomisation was performed within 50 trials in a 1:1 ratio to treated or untreated groups of 100 subjects in each trial.

The following parameters were used in data simulation: α_0 = (4.609, 5.458); α_1 =5.458; β_0 = (-2.0, -3.5,-4.0, 4.5,-5.0,-5.6); β_1 = (-1.43, -1.45, -1.7591, -3); Var(a_i) = 10; Var(b_i) = 4. The correlation between the treatment random effects ρ = Corr(a_i ,b_i) = $\sqrt{0.9}$, with R^2= 0.9. Vaccine efficacy VE = {1- p(T=1|Z=1) / p(T=1|Z=0)}x100% where p(T=1|Z=1) and p(T=1|Z=0) are the probabilities of disease among vaccinated and unvaccinated subjects, respectively.

The simulated data were loaded and prior values specified for MCMC steps. The modelling steps were performed for both NIP and WIP alike. Each MCMC step used 1000 samples as burn in, while 10,000 iterations were used for inference. The sampler adapts its behaviour to maximize efficiency after every 1000 iterations. Trace plots reveal the stability and proper mixing of the monitored parameters R^2 and D matrix across the 3 parallel chains.

Table 1: Comparison of results for VE=30%

	Non informative prior (NIP) model				**Weakly informative prior (WIP) model**			
Param	**Mean**	**SD**	**naïve SE**	**Time series SE**	**Mean**	**SD**	**naïve SE**	**Time series SE**
Dmat[1,1]	4.0533	0.92149	0.00532	0.01425	3.9449	0.89939	0.00519	0.01394
Dmat[1,2]	6.5132	1.42206	0.00821	0.01474	6.1999	1.35424	0.00782	0.01435
Dmat[2,2]	11.5323	2.44915	0.01414	0.02106	11.2986	2.365	0.01365	0.02288
R^2	0.9087	0.03288	0.00019	0.00062	0.8639	0.04793	0.00028	0.00151
Dmat[1,1]	5.9584	1.33923	0.00773	0.01745	5.8872	1.29869	0.00750	0.01861
Dmat[1,2]	9.0683	1.97891	0.01143	0.01612	8.7076	1.87628	0.01083	0.01970
Dmat[2,2]	15.1804	3.22186	0.01860	0.02447	14.889	3.09118	0.01785	0.03218

R^2	0.9098	0.03115	0.00018	0.00052	0.8657	0.04572	0.00026	0.00141
Dmat[1,1]	4.7558	1.07016	0.00618	0.01543	4.6378	1.04233	0.00602	0.01675
Dmat[1,2]	6.8041	1.48787	0.00859	0.01626	6.4635	1.41823	0.00819	0.01803
Dmat[2,2]	10.5903	10.5903	0.01316	0.02588	10.327	2.19202	0.01266	0.02647
R^2	0.9203	0.03052	0.00018	0.00072	0.8733	0.04697	0.00027	0.00179
Dmat[1,1]	2.8118	0.64071	0.00370	0.00901	2.7921	0.62779	0.00363	0.00895
Dmat[1,2]	4.3419	0.96447	0.00557	0.00912	4.1523	0.92241	0.00533	0.00942
Dmat[2,2]	7.7676	1.67559	0.00967	0.01542	7.6992	1.63248	0.00943	0.01532
R^2	0.8643	0.04683	0.00027	0.00093	0.8036	0.06496	0.00038	0.00189
Dmat[1,1]	4.794	1.09171	0.00630	0.01424	4.6427	1.04932	0.00606	0.01602
Dmat[1,2]	6.8001	1.49962	0.00866	0.01297	6.4538	1.42252	0.00821	0.01638
Dmat[2,2]	10.7203	2.29062	0.01323	0.02058	10.6217	2.22064	0.01282	0.02160
R^2	0.9005	0.03531	0.00020	0.00069	0.8454	0.05433	0.00031	0.00169
Dmat[1,1]	3.7355	0.84472	0.00488	0.01136	3.6527	0.81543	0.00471	0.01128
Dmat[1,2]	5.8277	1.28343	0.00741	0.01231	5.5241	1.21619	0.00702	0.01332
Dmat[2,2]	10.1792	2.17406	0.01255	0.02019	9.9498	2.09989	0.01212	0.02141
R^2	0.8938	0.03753	0.00022	0.00070	0.8403	0.05427	0.00031	0.00164

Though the MCMC samples often have high autocorrelation, the naive MCMC error disregards the potential auto-correlation, and therefore the naive MCMC error is not realistic. The time-series MCMC error takes this auto-correlation into an account for the estimation of the error. As expected higher time-series MCMC than naïve MCMC errors are observed for both NIP and WIP models. Coefficient of determination is consistently higher for NIP compared to WIP models in low, moderate and high vaccine efficacies as shown in tables 1-3.

Table 2: Comparison of results for VE=70%

	Non informative prior (NIP) model				**Weakly informative prior (WIP) model**			
param	**Mean**	**SD**	**naïve SE**	**Time series SE**	**Mean**	**SD**	**naïve SE**	**Time series SE**
Dmat[1,1]	5.7900	1.60136	0.00925	0.08041	5.8065	1.67774	0.00969	0.09424
Dmat[1,2]	7.5704	1.78781	0.01032	0.06125	7.1525	1.71157	0.00988	0.05861
Dmat[2,2]	10.6336	2.31954	0.01339	0.03841	10.2679	2.18146	0.01260	0.03198
R^2	0.9365	0.02997	0.00017	0.00101	0.8658	0.06055	0.00035	0.00338
Dmat[1,1]	3.2587	0.81166	0.00469	0.02334	3.2397	0.80682	0.00466	0.02318
Dmat[1,2]	4.8893	1.10962	0.00641	0.02221	4.6432	1.04631	0.00604	0.01540
Dmat[2,2]	8.0961	1.75978	0.01016	0.02259	7.9282	1.67500	0.00967	0.01743
R^2	0.9096	0.03658	0.00021	0.00099	0.8437	0.06246	0.00036	0.00256
Dmat[1,1]	5.1164	1.26863	0.00733	0.04437	5.0483	1.25725	0.00723	0.04832
Dmat[1,2]	7.0250	1.59357	0.00920	0.04003	6.7147	1.51666	0.00876	0.03542
Dmat[2,2]	10.2632	2.23856	0.01292	0.04211	9.9694	2.11865	0.01223	0.02873
R^2	0.9427	0.02557	0.00015	0.00074	0.8991	0.04429	0.00026	0.00210
Dmat[1,1]	4.3508	1.09760	0.00634	0.03569	4.5347	1.14335	0.00660	0.04065

Dmat[1,2]	5.8134	1.34244	0.00775	0.02562	5.6567	1.31957	0.00762	0.03049
Dmat[2,2]	9.1136	1.97332	0.01139	0.02381	9.1855	1.96122	0.01132	0.02800
R^2	0.8553	0.05192	0.00030	0.00128	0.7718	0.07445	0.00043	0.00245
Dmat[1,1]	3.5487	0.88040	0.00508	0.02654	3.6140	0.89989	0.00520	0.02802
Dmat[1,2]	5.2356	1.21270	0.00700	0.02243	4.9960	1.16611	0.00673	0.02066
Dmat[2,2]	9.1817	1.98300	0.01145	0.02238	9.1420	1.93127	0.01115	0.01720
R^2	0.8318	0.05771	0.00033	0.00128	0.7590	0.07882	0.00046	0.00247
Dmat[1,1]	3.5222	0.88773	0.00513	0.02868	3.5165	0.90040	0.00520	0.02946
Dmat[1,2]	4.7017	1.07601	0.00621	0.01988	4.4540	1.02670	0.00593	0.01745
Dmat[2,2]	7.0281	1.54080	0.00890	0.01787	6.9352	1.48350	0.00856	0.01723
R^2	0.8969	0.04313	0.00025	0.00131	0.8185	0.07140	0.00041	0.00284

Table 3: Comparison of results for VE=95%

	Non informative prior (NIP) model				Weakly informative prior (WIP) model			
param	**Mean**	**SD**	**naïve SE**	**Time series SE**	**Mean**	**SD**	**naïve SE**	**Time series SE**
Dmat[1,1]	1.0901	0.6997	0.00404	0.06956	1.1261	0.8512	0.00492	0.08368
Dmat[1,2]	2.3933	1.1262	0.00650	0.09791	1.3831	0.9997	0.00577	0.06248
Dmat[2,2]	8.8972	1.8953	0.01094	0.01606	9.5796	2.1271	0.01228	0.02597
R^2	0.6344	0.2219	0.00128	0.01568	0.2529	0.1930	0.00111	0.01062
Dmat[1,1]	1.9628	1.2498	0.00722	0.15305	1.6457	1.1713	0.00676	0.11282
Dmat[1,2]	3.5079	1.4537	0.00839	0.14525	2.1113	1.1799	0.00681	0.08162
Dmat[2,2]	8.4924	1.8466	0.01066	0.02173	8.9194	1.9596	0.01131	0.02292
R^2	0.7868	0.1548	0.00089	0.01148	0.3749	0.2163	0.00125	0.01250
Dmat[1,1]	1.7748	1.0017	0.00578	0.10410	1.1600	0.8601	0.00497	0.07584
Dmat[1,2]	4.0094	1.5388	0.00888	0.14085	2.0780	1.3820	0.00798	0.12022
Dmat[2,2]	11.837	2.5109	0.01450	0.02125	12.557	2.7445	0.01585	0.03471
R^2	0.8022	0.1345	0.00078	0.00863	0.3570	0.2213	0.00128	0.01490
Dmat[1,1]	1.1874	0.7746	0.00447	0.07767	1.1094	1.2385	0.00715	0.16124
Dmat[1,2]	2.8639	1.3421	0.00775	0.12163	1.7543	1.3439	0.00776	0.13620
Dmat[2,2]	10.480	2.2441	0.01296	0.01847	11.0808	2.4535	0.01417	0.04168
R^2	0.6989	0.2030	0.00117	0.01621	0.3277	0.2165	0.00125	0.01338
Dmat[1,1]	2.0189	1.2892	0.00744	0.14674	1.8263	1.4999	0.00866	0.15756
Dmat[1,2]	3.6996	1.4835	0.00857	0.13498	2.4119	1.5062	0.00870	0.11473
Dmat[2,2]	9.0266	1.9749	0.01140	0.02515	9.5311	2.0957	0.01210	0.02728
R^2	0.7988	0.1285	0.00074	0.00733	0.3972	0.2152	0.00124	0.01243
Dmat[1,1]	2.5260	1.2680	0.00732	0.12323	2.5511	1.5449	0.00892	0.16747
Dmat[1,2]	4.8010	1.6178	0.00934	0.14030	3.7371	1.5216	0.00879	0.10025
Dmat[2,2]	11.164	2.4055	0.01389	0.02296	11.4913	2.4507	0.01415	0.03154
R^2	0.8510	0.0971	0.00056	0.00598	0.5298	0.1783	0.00103	0.00994

4. Conclusion

The comparison of coefficient of determination shows that R^2 is consistently better for NIP than for WIP model for all vaccine efficacy,

VE=30% -100%. None of both models experienced computational issues. However, WIP models take about 20% longer time than NIP model to converge. The NIP model outperforms the WIP model and suggests a better validation tool.

5. References

[1]. M. Voysey, M. Sadarangani, A. J. Pollard, and T. R. Fanshawe (2018). Computing threshold antibody levels of protection in vaccine clinical trials: An assessment of methodological bias," PloS one, vol. 13, no. 9.

[2]. A. Callegaro and F. Tibaldi (2019). Assessing correlates of protection in vaccine trials: statistical solutions in the context of high vaccine efficacy. BMC medical research methodology, vol. 19, no. 1, p. 47.

[3]. M. Buyse and G. Molenberghs (1998). Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics, pp. 1014-1029.

[4]. S. S. Ellenberg and J. M. Hamilton (1989). Surrogate endpoints in clinical trials: cancer. Statistics in medicine, vol. 8, no. 4, pp. 405-413.

[5]. T. Burzykowski, G. Molenberghs, and M. Buyse (2006). The evaluation of surrogate endpoints. Springer Science & Business Media.

[6]. U.S Food and Drug Administration (2018, July 24). Surrogate endpoint resources for drug and biologic development. https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development.

[7]. A. Gelman and C. R. Shalizi (2013). Philosophy and the practice of bayesian statistics. British Journal of Mathematical and Statistical Psychology, vol. 66, no. 1, pp. 8-38.

[8]. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin (2013). Bayesian data analysis. CRC press.

[9]. R. G. Gallager (2013). Stochastic processes: theory for applications. Cambridge University Press.

[10]. P. Congdon (2014). Applied bayesian modelling, vol. 595. John Wiley & Sons.

[11]. Y. Chung, A. Gelman, S. Rabe-Hesketh, J. Liu, and V. Dorie (2015). Weakly informative prior for point estimation of covariance matrices in hierarchical models. Journal of Educational and Behavioral Statistics, vol. 40, no. 2, pp. 136-157.