

Principal Surrogates in Context of High Vaccine Efficacy

Andrea Callegaro (✉ andrea.x.callegaro@gsk.com)

GSK Vaccine <https://orcid.org/0000-0003-4805-5118>

Tibaldi Fabian

GSK Vaccines

Dean Follmann

National Institutes of Health

Research article

Keywords: Causal inference, High Vaccine Efficacy, Principal Surrogate, Surrogate Endpoint, Vaccine Clinical trial

Posted Date: November 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.16785/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Statistical Communications in Infectious Diseases on January 1st, 2021. See the published version at <https://doi.org/10.1515/scid-2020-0003>.

RESEARCH

Principal Surrogates in Context of High Vaccine Efficacy

Andrea Callegaro^{1*}, Fabian Tibaldi¹ and Dean Follmann²

*Correspondence:

andrea.x.callegaro@gsk.com

¹GSK Vaccines, Rue de l'Institut
89, Rixensart, Belgium

Full list of author information is
available at the end of the article

Abstract

Background: The use of correlates of protection (CoPs) in vaccination trials offers significant advantages as useful clinical endpoint substitutes. Vaccines with very high vaccine efficacy (VE) are documented in the literature (95% or above). Callegaro and Tibaldi, (2019) showed that the rare infections observed in the vaccinated groups of these trials poses challenges when applying conventionally-used statistical methods for CoP assessment such as the Prentice criteria and meta-analysis.

Methods: In this paper, we extended Callegaro and Tibaldi, (2019) simulation study by evaluating the impact of high VE on the Principal stratification approach.

Results: Similarly to the Prentice framework, we showed that the power decreases when the VE grows. It follows that it can be challenging to validate a principal surrogate (and a statistical surrogate) when rare infections are observed in the vaccinated groups.

Keywords: Causal inference; High Vaccine Efficacy; Principal Surrogate; Surrogate Endpoint; Vaccine Clinical trial

Background

An important factor influencing the duration and complexity of clinical trials is the choice of the endpoint used to assess vaccine efficacy. It would be extremely convenient to replace a late, costly or rare true endpoint by an immunological surrogate, which is measured earlier, cheaper, or more frequently. However, from a regulatory perspective, a surrogate endpoint (called sometimes Surrogate of protection, or Correlate of Protection) is not considered acceptable for the determination of efficacy, unless it has been validated, i.e. shown to predict clinical benefit. Prentice (1989) [1] introduced a formal definition of surrogacy based on the concept of mediation in a single-trial setting. Although appealing, Prentice's definition and criteria received criticism [2, 3]. In subsequent decades, many statistical methods have been proposed for the evaluation of surrogate endpoints, most of them framed within the causal inference [2, 4, 5] and meta-analytic paradigms [6, 7, 8].

Although not common, vaccines with very high efficacy are documented in the literature [9, 10, 11, 12, 13, 14]. These include the salmonella typhi vi coniugate [9], or the combined measles-mumps-rubella-varicella immunisation [14]. Assessing CoPs in the context of high VE using classical statical methods is problematic. Indeed, a very small number of cases/infections (corresponding to the vaccinated groups) can trigger considerable issues for such statistical models. There is therefore

a need to evaluate the statistical methods for CoP assessment to the context of high efficacy vaccines. Callegaro and Tibaldi (2019) [15] showed that the validation of a surrogate endpoint using the Prentice criteria and meta-analytic frameworks (by randomized subgroups in single trial setting) can be problematic in case of high VE because of the rare events available in the vaccine group. The aim of this paper is to evaluate the performance of the causal framework, specifically the Principal Surrogate approach [4, 5] in case of high VE.

Methods

The Prentice Criteria

The following set of notation will be used throughout the manuscript: Y_j and S_j are random variables denoting the observed clinical (binary) and the surrogate endpoint for subject $j = 1, \dots, n$ and Z_j is a binary treatment indicator.

Key concepts, including the hypothesis-testing approach to the validation of substitute endpoints using randomised clinical trial data, were introduced by Prentice in 1989 [1]. Prentice four criteria for the validation of a surrogate endpoint can be evaluated using the following 4 models:

$$\begin{aligned} \text{logit}(P(Y_i = 1)) &= \mu_T + \beta Z_i \\ S_i &= \mu_S + \alpha Z_i + \epsilon_{S_i} \\ \text{logit}(P(Y_i = 1)) &= \mu + \gamma S_i \\ \text{logit}(P(Y_i = 1)) &= \tilde{\mu}_T + \beta_S Z_i + \gamma_Z S_i. \end{aligned}$$

In this paper we will mainly focus on criterion 4. This criterion is met if the null hypothesis $H_{01} : \gamma_Z = 0$ is rejected ($\text{p-value}(s) < \alpha$) and the null hypothesis $H_{02} : \beta_S = 0$ is not rejected ($\text{p-value}(z) \geq \alpha$).

Principal Surrogate Framework

Many causal inference approaches/methods have been published in the literature. In what follows, we describe the Vaccine Efficacy Framework of Follmann and Gilbert [4, 5]. Since S_i can be affected by treatment, there are 2 naturally occurring counterfactual values of S_i : $S_i(1)$ under treatment, and $S_i(0)$ under control. The observed clinical endpoint (binary) is denoted by Y_i and the counterfactual values are $Y_i(1)$ under treatment, and $Y_i(0)$ under control. Criteria for S to be a good surrogate are based on risk estimands that condition on the potential surrogate responses [5]

$$\begin{aligned} \text{risk}_{(1)}(s(1), s(0)) &= Pr(Y(1) = 1 | S(1) = s(1), S(0) = s(0)) \\ \text{risk}_{(0)}(s(1), s(0)) &= Pr(Y(0) = 1 | S(1) = s(1), S(0) = s(0)) \end{aligned}$$

A contrast in $\text{risk}_{(1)}(s(1), s(0))$ and $\text{risk}_{(0)}(s(1), s(0))$ is a causal effect on the clinical endpoint. A classical contrast used in vaccines is the Vaccine Efficacy (VE)

$$VE(s(1), s(0)) = 1 - \frac{Pr(Y(1) = 1 | S(1) = s(1), S(0) = s(0))}{Pr(Y(0) = 1 | S(1) = s(1), S(0) = s(0))}$$

A Principal Surrogate (PS) is a biomarker satisfying two conditions: causal necessity

$$VE(s(1), s(0)) = 0 \text{ for all } s(1) = s(0)$$

and Wide Effect Modification (WEM) which means that

$$VE(s(1), s(0)) \text{ increasing in } s(1) - s(0),$$

WEM is similar in spirit to the Individual Causal Association (ICA) [16], which is the correlation between the individual causal effect on the endpoint and on the surrogate.

In this paper we only focus on WEM. In fact, current works [5, 17, 18, 19, 20] suggest that WEM criterion is of primary importance for a biomarker to be a PS. Furthermore, [16] showed that the average causal necessity definition may be extremely restrictive.

Estimating VE

Assumptions A1-A3 (A1: Stable unit treatment value assumption; A2: Ignorable treatment assignments; A3: Equal individual clinical risk up to the time of surrogate measurements) imply that $risk(Z)(s(1), s(0))$ would be identified if we knew the potential outcomes $S_i(Z)$ of subjects assigned the opposite treatment $1 - Z$

$$\begin{aligned} risk_{(1)}(s(1), s(0)) &= Pr(Y = 1|Z = 1, S(1) = s(1), S(0) = s(0)) \\ risk_{(0)}(s(1), s(0)) &= Pr(Y = 1|Z = 0, S(1) = s(1), S(0) = s(0)). \end{aligned}$$

It follows that it is necessary to impute (or integrate out) the missing potential biomarkers. The risk can be modeled using the following logistic model

$$logit(P(Y_i = 1|z_i, s_{1i}, s_{0i})) = \beta_0 + \beta_z z_i + \beta_{s(1)} s_{1i} + \beta_{s(1)z} s_{1i} z_i + \beta_{s(0)} s_{0i} + \beta_{s(0)z} s_{0i} z_i.$$

The model can be simplified in case of a Constant Biomarker ($S_i(0) = c$)

$$logit(P(Y_i = 1|z_i, s_{1i})) = \beta_0 + \beta_z z_i + \beta_{s(1)} s_{1i} + \beta_{s(1)z} s_{1i} z_i, \quad (1)$$

where the VE curve is used

$$VE(s(1)) = 1 - \frac{Pr(Y = 1|Z = 1, S(1) = s(1))}{Pr(Y = 1|Z = 0, S(1) = s(1))}.$$

The constant biomarker assumption is reasonable when subjects have been selected to have no meaningful exposure to the pathogen, so that $S(0) = 0$. Examples include HIV or varicella vaccine trials. This assumption is also reasonable for populations exposed to the pathogen when the biomarker S_i is the log10 Fold-Increase from baseline (FI_i), which is the difference between the log10 post (A_i) and the log10 baseline (B_i) values ($FI_i = A_i - B_i$).

Missing values imputation/integration

The key challenge in estimating these risk estimands is solving the problem of conditioning on counterfactual values that are not observable. This involves integrating out (or imputing) missing values based on some models, and under some set of assumptions and/or trial augmentations. [5] and [4] proposed to use the estimated maximum likelihood followed by bootstrap. [21] suggested a pseudoscore estimation procedure that does have a closed form variance estimator. [22] used a multiple imputation approach. In this paper we fit model 1 using the method implemented in the R package `pseval` [23]: Baseline Immunogenicity Predictor (BIP); parameters estimated using estimated maximum likelihood (missing information is integrated out) and the variance is estimated by bootstrap. Rcode is provided in the Appendix. This approach is similar in spirit to the method used in Follmann [4].

Results

Simulations of Callegaro and Tibaldi (2019)

To evaluate the impact of high vaccine efficacy on the PS validation, we repeated the simulations of Callegaro and Tibaldi [15]. The Dunning regression model [24] was used to simulate the data in an ideal CoP setting, where the treatment effect is fully explained by the post values (A_i) as follows:

$$P(Y_i = 1|\pi, A_i) = \pi \frac{e^{\mu+\gamma A_i}}{1 + e^{\mu+\gamma A_i}}. \quad (2)$$

Here, π can be interpreted as the probability of being exposed to the disease.

Simulations were run using the following parameter assumptions: total sample size $n = 5000$, 1:1 randomization, $\pi=0.1$, $\mu = 8.3$, $\gamma = \log(1 - 0.95)$; the immune response post vaccination is normally distributed $A|Z = 0 \sim N(3, 0.2)$ in the placebo group and $A|Z = 1 \sim N(3 + \Delta, 0.2)$ in the vaccine group, where $\Delta = 0.33, 0.75, 1, 1.5$. The value of the immune response at baseline is generated as $B \sim N(3, 0.2)$ with correlation between A and B of 0.90 in the placebo group and 0.50 in the vaccine group. For each scenario, 1000 clinical trials were simulated.

We fit Prentice model 4 on the simulated data with Fold-Increase ($S_i = FI_i = A_i - B_i$) as surrogate adjusting for the baseline (B_i) using logit regression

$$\text{logit}(P(Y_i = 1)) = \tilde{\mu}_T + \beta_S Z_i + \gamma_Z FI_i + \gamma_B B_i$$

and the scaled logistic model [24]

$$P(Y_i = 1) = \pi \frac{e^{\tilde{\mu}_T + \beta_S Z_i + \gamma_Z FI_i + \gamma_B B_i}}{1 + e^{\tilde{\mu}_T + \beta_S Z_i + \gamma_Z FI_i + \gamma_B B_i}}.$$

Note that this model is consistent with the model used to generate the data (Equation 2), with a slightly different parametrization. The power to meet Prentice criterion 4 (PC4) was measured as the proportion of simulated trials with p-value(s) = $2\Phi(-|\hat{\gamma}_Z/\sqrt{\text{Var}(\hat{\gamma}_Z)}|)$ < α and p-value(z) = $2\Phi(-|\hat{\beta}_S/\sqrt{\text{Var}(\hat{\beta}_S)}|)$ $\geq \alpha$, $\alpha = 0.05$.

Furthermore, we applied the Principal surrogate approach on vaccine induced fold-increase ($S(1)_i = FI(1)_i$) where missing information is integrated out using

the baseline surrogate measurement (B_i). The power of the WEM approach was measured as the proportion of simulated trials with significant Wald statistics for the $s(1)z$ coefficient of model (1) ($\text{pvalue}(s(1)z) < \alpha$, $\alpha = 0.05$). Appendix contains the R code used to apply the PS approach is provided.

Table 1 shows that the power of both PC4 and WEM decreases when the VE increases. This is due to the fact that there is less information (number of events) as the VE increases. Note that the power of the Prentice approach is higher than in Callegaro and Tibaldi [15] because of the inclusion of the baseline surrogate as covariate. Simulation results suggest similar power for PC4 and WEM approaches.

Simulations with constant biomarker under placebo

In the previous simulations the Fold-Increase was not constant in placebo (it was normally distributed). To evaluate the performance of the Prentice and PS approach in case of constant biomarker under placebo, which mimics vaccine trials in a naive population, we simulated data using the model described above. However, in the inferential models, we replaced FI by FI^* which is constant in Placebo. FI^* is defined as

$$FI^* = \begin{cases} FI & \text{if } FI > c \\ c & \text{if } FI \leq c \end{cases}$$

where c is the 99% quantile of the distribution of FI in Placebo.

Table 2 shows some loss of power of the PS approach when the VE increases. Even if the use of the Prentice framework is not justified in this context, table 2 shows the results of the Prentice criteria 4 (PC4 logistic model). Results from PC4 scaled logistic are not shown because the model is not converging. We observe a dramatic loss of power of Prentice criterion 4 when the VE is high.

Note that table 2 shows simulation results where the inferential models do not agree with the data generating mechanism, so it represents a situation of model miss-specification.

To disentangle the problem of model miss-specification from the constant biomarker problem, we generate additional constant biomarker data using a model consistent with the ‘inferential’ model used to fit the data. We simulated data using the following Dunning regression model:

$$P(Y_i = 1 | \pi, FI_i^*, B_i) = \pi \frac{e^{\mu + \gamma FI_i^* + \gamma_B B_i}}{1 + e^{\mu + \gamma FI_i^* + \gamma_B B_i}}.$$

Here, $\pi = 0.1$ and the other parameters are chosen to mimic Table 1 data: $\Delta = 0.33, 0.75, 1, 1.47$, $\mu = (8.66, 9.45, 9.82, 9.41)$, $\gamma = (-5.39, -5.15, -4.8, -4.45)$ and $\gamma_B = (-2.31, -2.63, -2.79, -2.66)$.

Table 3 shows that the loss of power of Prentice approach shown in Table 2 was mainly due to model miss-specification.

Simulations with low/moderate VE

For comparison, we considered simulations with low VE. We simulated data as described above with $\mu_1 = E(A|Z = 1) = 3, 3.075, 3.15, 3.23$, corresponding to

estimated VE about 0%, 10%, 20% and 30%, respectively. Note that Prentice criteria 1 will not be met in this situation. For simplicity, we focused only on Prentice criterion 4. Table 4 shows that both approaches (PC4 and WEM) are powerful in the case of low/moderate VE. Prentice criterion 4 seems to be slightly more powerful than PS.

Simulations using random intercept logistic (correlated potential outcomes)

Finally, we generated data in a different way more aligned with the causal inference setting (potential outcomes). We generated correlated post-vaccination values $(A(0), A(1))$ using a bivariate normal distribution

$$\begin{pmatrix} A(0) \\ A(1) \end{pmatrix} \sim N \left[\begin{pmatrix} 3 \\ 3 + \Delta \end{pmatrix}, \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \right]$$

with $\Delta = (0.33, 0.75, 1.1, 1.6)$. The mean and the variance of the baseline are the same as the post-dose surrogate in Placebo. The correlation between baseline and post is 90% in Placebo and 50% in Vaccinated, respectively. We generated the correlated clinical outcomes using a logistic model with individual random intercept (b_i)

$$\text{logit}(P(Y(z)_i = 1|A(z)_i, b_i)) = \mu + A(z)_i\gamma + b_i.$$

The variables $Y(0), Y(1)$ are conditionally independent given b but unconditionally (averaged over b) correlated. The extent of correlation depends on the variance of the random effect ($\text{var}(b)$).

We generated bridge distributed random intercept (using R package `bridgedist` [25]) such that the resultant marginal distribution follows a logistic regression model [26]. In fact, the marginal logistic regression model is $\text{logit}(P(Y(z) = 1|A(z))) = \mu/c + A(z)\gamma/c$ for $z = 0, 1$ with $c = \sqrt{1 + 3\text{var}(b)/\pi^2}$. We simulated data with the following parameters: $\text{var}(b) = 10$ (scale = 0.5), $\mu = 3.6$ and $\gamma = -3.8$. In this way, $p_0 = P(Y = 1|Z = 0) = 0.05$ and the estimated VE is about 0.45, 0.75, 0.85 and 0.95.

Table 5 shows that Prentice criterion 4 is more powerful than WEM.

The Prentice framework is more powerful than PS for different reasons: i) PS tests for an interaction, which is less powerful than a test for the main effect; ii) the covariate S (observed surrogate in vaccinated and placebo) has greater range in the Prentice model 4 than the covariate $S(1)$ in the PS model. It is easier to estimate a slope for a covariate with a bigger range. Figure 1 illustrate these differences.

Case study: Analysis of a simulated data-set with large VE

In this section we analyze one simulated dataset from the scenario with largest VE of table 1. The sample size is $N=5000$, with 1:1 randomization. The number of events observed in the two groups are 3 and 90, with an estimated VE of 96% (95%CI, 0.89%-98%). Figure 2 shows that the vaccine and placebo groups had similar log10 titer distributions at baseline while there is a small overlap in distributions post vaccination. Antibody responses clearly increased from baseline to post-dose in vaccine recipients but not in placebo recipients.

Figure 3 shows the Spearman correlation between baseline and post (left panel) and between baseline and Fold-Increase (right panel). First we examine the interaction between surrogate and the treatment. Table 6 shows that there is no interaction (p-value=0.49).

Secondly, we assess the four Prentice criteria. Table 7 shows that all criteria are met. In particular, the last 4 rows shows the results related to criterion 4. We can see that the effect of the surrogate is significant (p-value(s)=0.019), while the treatment effect is not significant, but is close to 5% (p-value(z)=0.078).

Slightly better results are obtained if Dunning model is used (see Table 8).

In summary, there is suggestive though not strong evidence that the Fold-Increase is a Statistical Surrogate.

Principal Surrogate framework

Table 9 shows the results from R package `pseval` with 50 bootstrap (R codes are provided in the Appendix). We can see that the interaction between the treatment group and FI(1) (test for wide effect modification) is borderline (p-value=0.053).

Figure 4 shows the estimated VE curve for Fold-Increase. The estimated VE curve is an increasing function of FI(1), however we can see large variability for small values of FI(1) and negative VEs for vaccine recipients with no rise.

In summary, there is suggestive though not strong evidence that the Fold-Increase is a Principal Surrogate.

Conclusions

Although not common, vaccines with very high efficacy (95% or above) are documented in the literature [9, 10, 11, 12, 13, 14]. These trials raise the problem of assessing CoPs in the context where small number of cases/infections in vaccinated groups are available.

Callegaro and Tibaldi (2019) [15] showed that the validation of a surrogate endpoint using the Prentice criteria and meta-analytic frameworks (by randomized subgroups in single trial setting) can be problematic in case of high VE. In this paper, we evaluate the performance of the causal framework, specifically the Principal Surrogate (PS) approach [4, 5] in case of high VE.

First, we replicated the simulation study of Callegaro and Tibaldi [15] where the clinical outcome was simulated using Prentice model 3 (assuming full mediation) and using the Dunning model [24]. These simulation results show that i) adjustments for important covariates (such as baseline surrogate) considerably improves the power of the Prentice approach (even if the model is miss-specified) in case of high VE. Furthermore, these simulation results show similar power of Prentice and PS frameworks. The power of both approaches decreases when VE grows.

Second, we slightly changed the Callegaro and Tibaldi scenario to consider the case of constant biomarker under placebo and the case of small/moderate VE. Simulation results show that i) PS is more powerful than Prentice in case of constant biomarker when the inferential model is miss-specified, otherwise Prentice is more powerful; ii) Prentice criteria 4 and PS frameworks are powerful when the VE is small (see Table 3). However, in this case Prentice criteria 1 is not met, so the two approaches give different conclusions.

Finally, we simulated correlated potential outcome data using a bivariate (random intercept) logistic regression. In this case the Prentice framework is more powerful than the PS approach. This can be due to the following reasons: i) Prentice model 4 corresponds to the model used to generate the data and so there is no lack of fit in the Prentice framework; ii) PS tests for an interaction, which is less powerful than a test for the main effect; iii) the covariate S (observed surrogate in vaccinated and placebo) has greater range in the Prentice model 4 than the covariate $S(1)$ in the PS model. It is easier to estimate a slope for a covariate with a bigger range (see Figure 1); (iv) Principal stratification has to impute $S(1)$ for placebo participants which increases the variability of estimates relative to knowing $S(1)$. In contrast S is known in all for the Prentice criterion.

It is important to highlight that the power comparison between the two approaches should be interpreted with care. In fact, the two approaches measure two different things: Prentice framework evaluates if the surrogate is a “statistical surrogate” while the PS evaluates if the surrogate is a “principal surrogate” (see [27] for more details).

For illustration, we analyzed one data-set simulated with full mediation (Dunning model 3) and with high VE ($\hat{VE} = 96\%$). Results showed suggestive thought not strong evidence that the FI is a Statistical Surrogate (Prentice criteria) or a PS. These results are due to the lack of power of these approaches in case of high VE. An interesting topic for future research is the implementation of the two approaches in a Bayesian framework with weakly informative priors (WIP). In fact, [15] showed that WIP can considerably increase the power of the meta-analytical approach in case of high VE.

In conclusion, we evaluated by simulation the impact of high VE on the PS approach. Similarly to the Prentice framework, we showed that the power decreases when the VE grows. It follows that it can be challenging to validate a principal surrogate (and a statistical surrogate) when rare infections are observed in the vaccinated groups.

Abbreviations

BIP: Baseline Immunogenicity Predictor
 CI: Confidence Interval
 CoP: Correlate of Protection
 FI: Fold Increase
 ICA: Individual Causal Association
 PC4: Prentice criterion 4
 PS: Principal Surrogate
 UL: Upper Limit
 VE: Vaccine Efficacy
 WEM: Wide Effect Modification
 WIP: Weakly Informative Prior

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

AC and FT are employees of the GSK group of companies and hold shares in the GSK group of companies. DF declares no conflict of interest.

Funding

GlaxoSmithKline Biologicals SA was the funding source and was involved in all stages of the study conduct and analysis. GlaxoSmithKline Biologicals SA also took responsibility for all costs associated with the development and publishing of the present manuscript.

Authors contributions

AC, FT and DF equally contributed to all steps of the manuscripts development, and approved its final version.

Author details

¹GSK Vaccines, Rue de l'Institut 89, Rixensart, Belgium. ²National Institute of Allergy and Infectious Diseases, 5601 Fishers Lane Room 4C11, Bethesda, MD 20892, U.S.A..

References

- Prentice, R.L.: Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* **8(4)**, 431–440 (1989)
- Frangakis, C.E., Rubin, D.B.: Principal stratification in causal inference. *Biometrics* **58(1)**, 21–29 (2002)
- Burzykowski, T., Molenberghs, G., Buyse, M.: *The Evaluation of Surrogate Endpoints*. Springer, New York (2005)
- Follmann, D.: Augmented designs to assess immune response in vaccine trials. *Biometrics* **62(4)**, 1161–1169 (2006)
- Gilbert, P.B., Qin, L., Self, S.G.: Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Stat Med* **27(23)**, 4758–4778 (2008)
- Daniels, M.J., Hughes, M.D.: Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* **16**, 1965–1982 (1997)
- Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H.: The validation of surrogate endpoints in meta-analysis of randomized experiments. *Biostatistics* **1**, 49–67 (2000)
- Gail, M.H., Pfeiffer, R., Houwelingen, H.C.V., Carroll, R.: On meta-analytic assessment of surrogate outcomes. *Biostatistics* **1**, 231–246 (2000)
- Mitra, M., Shah, N., Ghosh, A., Chatterjee, S., Kaur, I., Bhattacharya, N., Basu, S.: Efficacy and safety of vi-tetanus toxoid conjugated typhoid vaccine (pedatyph) in indian children: school based cluster randomized study. *Human vaccines & immunotherapeutics* **12(4)**, 939–945 (2016)
- Lin, F.Y.C., Ho, V.A., Khiem, H.B., et al.: The efficacy of a salmonella typhi vi conjugate vaccine in two-to-five-year-old children. *New England Journal of Medicine* **344(17)**, 1263–1269 (2001)
- Wei, M., Meng, F., Wang, S., Li, J., et al.: Two-year efficacy, immunogenicity, and safety of vigoo enterovirus 71 vaccine in healthy chinese children: a randomised open-label study. *The Journal of infectious diseases* **jiw502** (2016)
- Phua, K.B., Lim, F.S., Lau, Y.L., Nelson, E.A.S., et al.: Rotavirus vaccine RIX4414 efficacy sustained during the third year of life: a randomized clinical trial in an asian population. *Vaccine* **30(30)**, 4552–4557 (2012)
- Black, S., Shinefield, H., Fireman, B., Lewis, E., et al.: Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. *The Pediatric infectious disease journal* **19(3)**, 187–195 (2000)
- Prymula, R., Bergsaker, M.R., Esposito, S., Gothefors, L., et al.: Protection against varicella with two doses of combined measles-mumps-rubella-varicella vaccine versus one dose of monovalent varicella vaccine: a multicentre, observer-blind, randomised, controlled trial. *Lancet* **383(9925)**, 1313–1324 (2014)
- Callegaro, A., Tibaldi, F.: Assessing correlates of protection in vaccine trials: statistical solutions in the context of high vaccine efficacy. *BMC medical research methodology* **19(1)**, 47 (2019)
- Alonso, A., Van der Elst, W., Molenberghs, G., Buyse, M., Burzykowski, T.: On the relationship between the causal-inference and meta-analytic paradigms for the validation of surrogate endpoints. *Biometrics* **71(1)**, 15–24 (2015)
- Wolfson, J., Gilbert, P.: Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. *Biometrics* **66(4)**, 1153–1161 (2010)
- Huang, Y., Gilbert, P.: Comparing biomarkers as principal surrogate endpoints. *Biometrics* **67(4)**, 1442–1451 (2011)
- Gabriel, E., Gilbert, P.: Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. *Biostatistics* **15(2)**, 251–265 (2014)
- Gabriel, E., Follmann, D.: Augmented trial designs for evaluation of principal surrogates. *Biostatistics* **17(3)**, 453–467 (2016)
- Huang, Y., Gilbert, P.B., Wolfson, J.: Design and estimation for evaluating principal surrogate markers in vaccine trials. *Biometrics* **69(2)**, 301–309 (2013)
- Miao, C., Li, X., Gilbert, P., Chan, I.: *A multiple imputation approach for surrogate marker evaluation in the principal stratification causal inference framework. risk assessment and evaluation of predictions*. New York: Springer (2013)
- Sachs, C. M., Gabriel, E, E.: An introduction to principal surrogate evaluation with the pseval package. *The R journal* **8(2)**, 277–292 (2016)
- Dunning, A.J.: A model for immunological correlates of protection. *Stat Med* **25(9)**, 1485–1497 (2006)
- Swihart, B.: *Bridgedist: An Implementation of the Bridge Distribution with Logit-Link as in Wang and Louis (2003)*. (2016). R package version 0.1.0. <https://CRAN.R-project.org/package=bridgedist>
- Wang, Z., Louis, T.A.: Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika* **90(4)**, 765–775 (2003)
- Gilbert, P.B., Gabriel, E.E., Huang, Y., Chan, I.S.: Surrogate endpoint evaluation: Principal stratification criteria and the Prentice definition. *Journal of causal inference* **3(2)**, 157–175 (2015)

Figures

Figure 1. Simulated trial with $n = 200$ per arm under the scenario with $\Delta = 1.6$. The true probability of infection is graphed as a function of observed A s and $A(1)$ s respectively. The top panel is the data used for the Prentice criteria while the bottom panel is used to test WEM. Red denotes the placebo group while blue denotes the vaccine group.

Figure 2. Distribution of the surrogate endpoint: baseline, post and Fold-Increase (post-baseline).

Figure 3. Correlation between baseline and post (left panel) and between baseline and Fold-Increase (right panel).

Figure 4. Estimated Vaccine Efficacy curve across levels of vaccine-induced Fold-Increase from baseline to post-vaccination, with 95% confidence intervals (dashed lines).

Tables

Table 1 Simulation results of data generated using scaled logit Prentice model 3. Power ($\alpha = 0.05$) to assess Prentice criterion 4 (PC4) using logistic and scaled logistic model and power to assess the Wide Effect Modification (WEM) of a Principal Surrogate using the logistic model (p-value of the interaction $s(1)z$).

| Δ | \hat{VE} | PC4 logistic | PC4 scaled logistic | WEM |
|----------|------------|--------------|---------------------|------|
| 0.33 | 0.41 | 0.94 | 0.96 | 0.92 |
| 0.75 | 0.75 | 0.93 | 0.96 | 0.92 |
| 1.00 | 0.87 | 0.89 | 0.95 | 0.90 |
| 1.50 | 0.96 | 0.80 | 0.86 | 0.73 |

Table 2 Simulation results with constant biomarker (inferential models do not agree with the data generating mechanism). Power ($\alpha = 0.05$) to assess Prentice criterion 4 (PC4) using logistic model and power to assess the Wide Effect Modification (WEM) of a Principal Surrogate using the logistic model (p-value of the interaction $s(1)z$).

| Δ | \hat{VE} | PC4 logistic | WEM |
|----------|------------|--------------|------|
| 0.33 | 0.41 | 0.80 | 0.67 |
| 0.75 | 0.75 | 0.45 | 0.85 |
| 1.00 | 0.87 | 0.49 | 0.84 |
| 1.50 | 0.96 | 0.37 | 0.69 |

Table 3 Simulation results with constant biomarker (inferential models agree with the data generating mechanism). Power ($\alpha = 0.05$) to assess Prentice criterion 4 (PC4) using logistic model and power to assess the Wide Effect Modification (WEM) of a Principal Surrogate using the logistic model (p-value of the interaction $s(1)z$).

| Δ | \hat{VE} | PC4 logistic | WEM |
|----------|------------|--------------|------|
| 0.33 | 0.29 | 0.96 | 0.73 |
| 0.75 | 0.62 | 0.94 | 0.85 |
| 1.00 | 0.78 | 0.96 | 0.84 |
| 1.47 | 0.94 | 0.79 | 0.80 |

Appendix: R code used to apply the PS approach

```
library("pseval")
binary.est <- psdesign(data, Z = group, Y = y, S = FI, BIP = B) +
  integrate_parametric(S.1 ~ BIP) +
  risk_binary(model = Y ~ S.1 * Z, D = 50, risk = risk.logit) +
  ps_estimate(method = "BFGS")
binary.boot <- binary.est + ps_bootstrap(n.boots = 50, progress.bar = FALSE,
  start = binary.est$estimates$par, method = "BFGS")
```

Table 4 Simulation results with small/moderate VE. Power ($\alpha = 0.05$) to assess Prentice criterion 4 (PC4) using logistic and scaled logistic model and power to assess the Wide Effect Modification (WEM) of a Principal Surrogate using the logistic model (p-value of the interaction $s(1)z$).

| Δ | \hat{VE} | PC4 logistic | PC4 scaled logistic | WEM |
|----------|------------|--------------|---------------------|------|
| 0.000 | -0.01 | 0.95 | 0.96 | 0.92 |
| 0.075 | 0.09 | 0.95 | 0.96 | 0.91 |
| 0.150 | 0.20 | 0.95 | 0.97 | 0.91 |
| 0.250 | 0.31 | 0.95 | 0.97 | 0.93 |

Table 5 Simulation results with random intercept (bridge) logistic regression. Power ($\alpha = 0.05$) to assess Prentice criterion 4 (PC4) using logistic model and power to assess the Wide Effect Modification (WEM) of a Principal Surrogate using the logistic model (p-value of the interaction $s(1)z$).

| Δ | \hat{VE} | PC4 logistic | WEM |
|----------|------------|--------------|------|
| 0.33 | 0.46 | 0.95 | 0.85 |
| 0.75 | 0.75 | 0.94 | 0.84 |
| 1.10 | 0.87 | 0.91 | 0.77 |
| 1.60 | 0.95 | 0.89 | 0.74 |

Table 6 Logistic model with interaction between treatment group and surrogate.

| | Estimate | Std. Error | z value | p-value |
|-------------|----------|------------|---------|---------|
| (Intercept) | 0.833 | 0.717 | 1.161 | 0.245 |
| Z | -0.576 | 1.696 | -0.339 | 0.734 |
| FI | -1.060 | 0.555 | -1.909 | 0.056 |
| B | -1.434 | 0.257 | -5.583 | 0.000 |
| group:FI | -0.966 | 1.401 | -0.690 | 0.490 |

Table 7 Prentice criteria: Logistic and linear models.

| Criterion | variable | Estimate | Std Error | z value | p-value |
|-----------|-------------|----------|-----------|---------|---------|
| 1 | (Intercept) | 0.433 | 0.698 | 0.620 | 0.535 |
| 1 | Z | -3.448 | 0.588 | -5.865 | 0.000 |
| 1 | B | -1.293 | 0.249 | -5.196 | 0.000 |
| 2 | (Intercept) | 0.956 | 0.031 | 30.734 | 0.000 |
| 2 | Z | 1.472 | 0.009 | 163.653 | 0.000 |
| 2 | B | -0.317 | 0.010 | -31.166 | 0.000 |
| 3 | (Intercept) | 1.092 | 0.702 | 1.556 | 0.120 |
| 3 | FI | -1.983 | 0.285 | -6.970 | 0.000 |
| 3 | B | -1.545 | 0.250 | -6.176 | 0.000 |
| 4 | (Intercept) | 0.825 | 0.717 | 1.150 | 0.250 |
| 4 | Z | -1.644 | 0.933 | -1.763 | 0.078 |
| 4 | FI | -1.205 | 0.514 | -2.345 | 0.019 |
| 4 | B | -1.432 | 0.257 | -5.574 | 0.000 |

Table 8 Prentice criterion 4 using Dunning model.

| variable | Estimate | Std Error | z value | p-value |
|-------------|----------|-----------|---------|---------|
| (Intercept) | 8.528 | 3.442 | 2.478 | 0.013 |
| FI | -2.662 | 1.131 | -2.353 | 0.019 |
| Z | -0.620 | 1.305 | -0.475 | 0.635 |
| B | -2.978 | 0.963 | -3.092 | 0.002 |
| logit(pi) | -2.386 | 0.370 | -6.440 | 0.000 |

Table 9 Principal Surrogate Evaluation.

| | Estimate | boot se | lower CL 2.5% | upper CL 97.5% | p-value |
|-------------|----------|---------|---------------|----------------|----------|
| (Intercept) | -7.81 | 1.146 | -10.35 | -6.147 | 9.13e-12 |
| FI(1) | 2.64 | 0.573 | 1.79 | 3.891 | 4.18e-06 |
| Z | 2.90 | 2.846 | -3.66 | 6.593 | 3.08e-01 |
| FI(1):Z | -3.98 | 2.053 | -8.11 | -0.157 | 5.28e-02 |

Figures

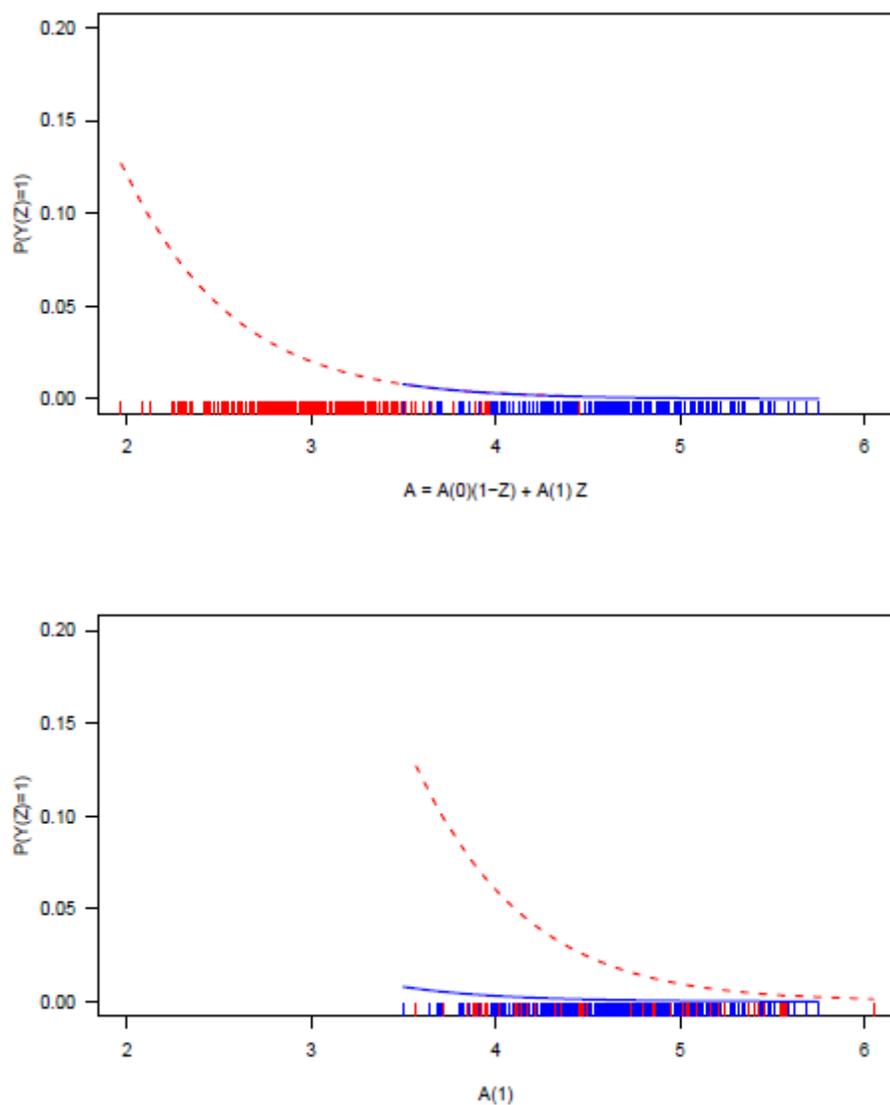


Figure 1

Simulated trial with $n = 200$ per arm under the scenario with $\lambda = 1:6$. The true probability of infection is graphed as a function of observed A s and $A(1)$ s respectively. The top panel is the data used for the Prentice criteria while the bottom panel is used to test WEM. Red denotes the placebo group while blue denotes the vaccine group.

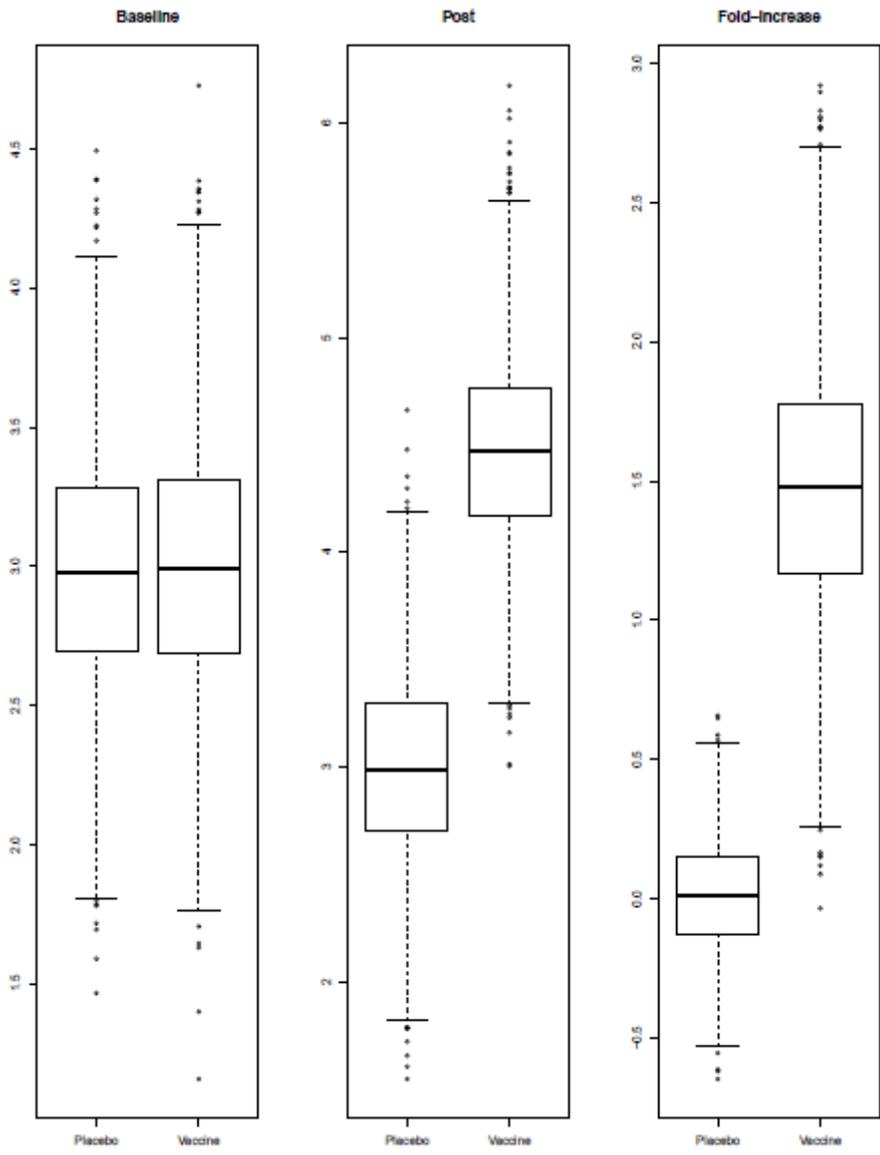


Figure 2

Distribution of the surrogate endpoint: baseline, post and Fold-Increase (post-baseline).

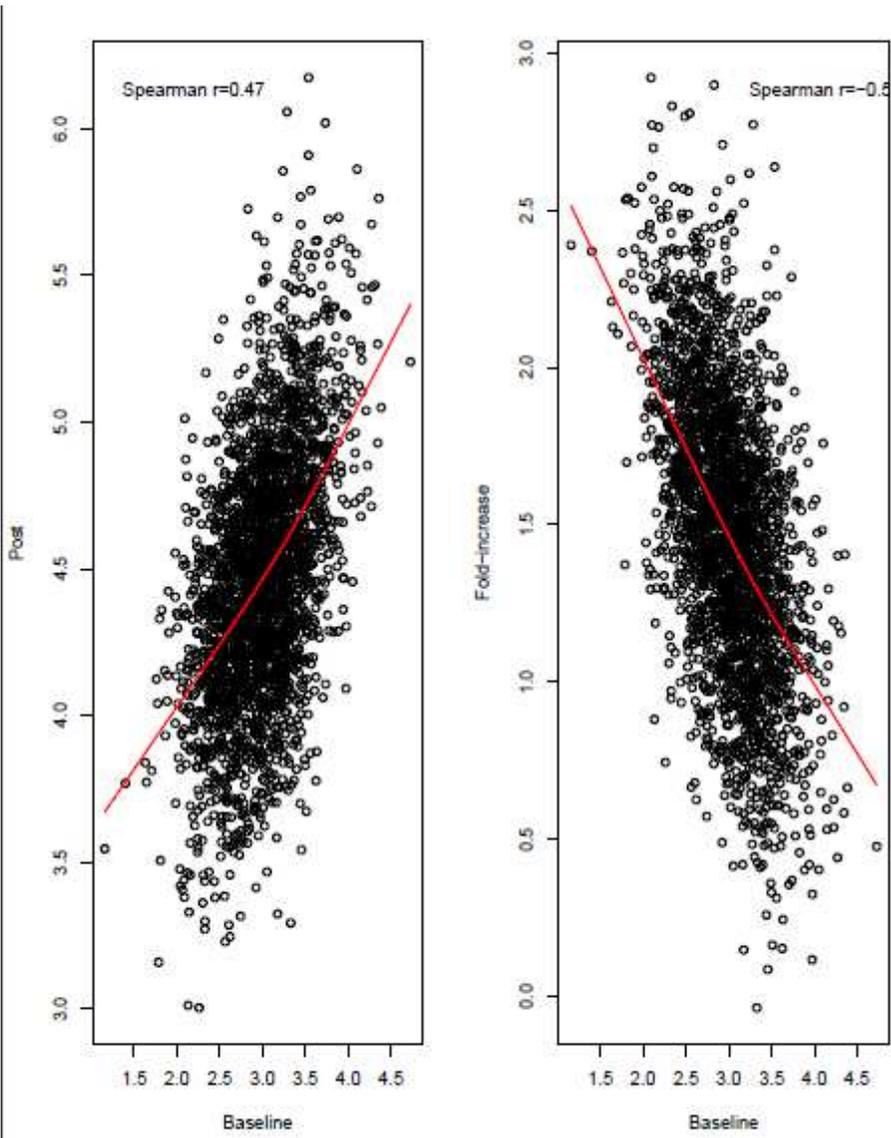


Figure 3

Correlation between baseline and post (left panel) and between baseline and Fold-Increase (right panel).

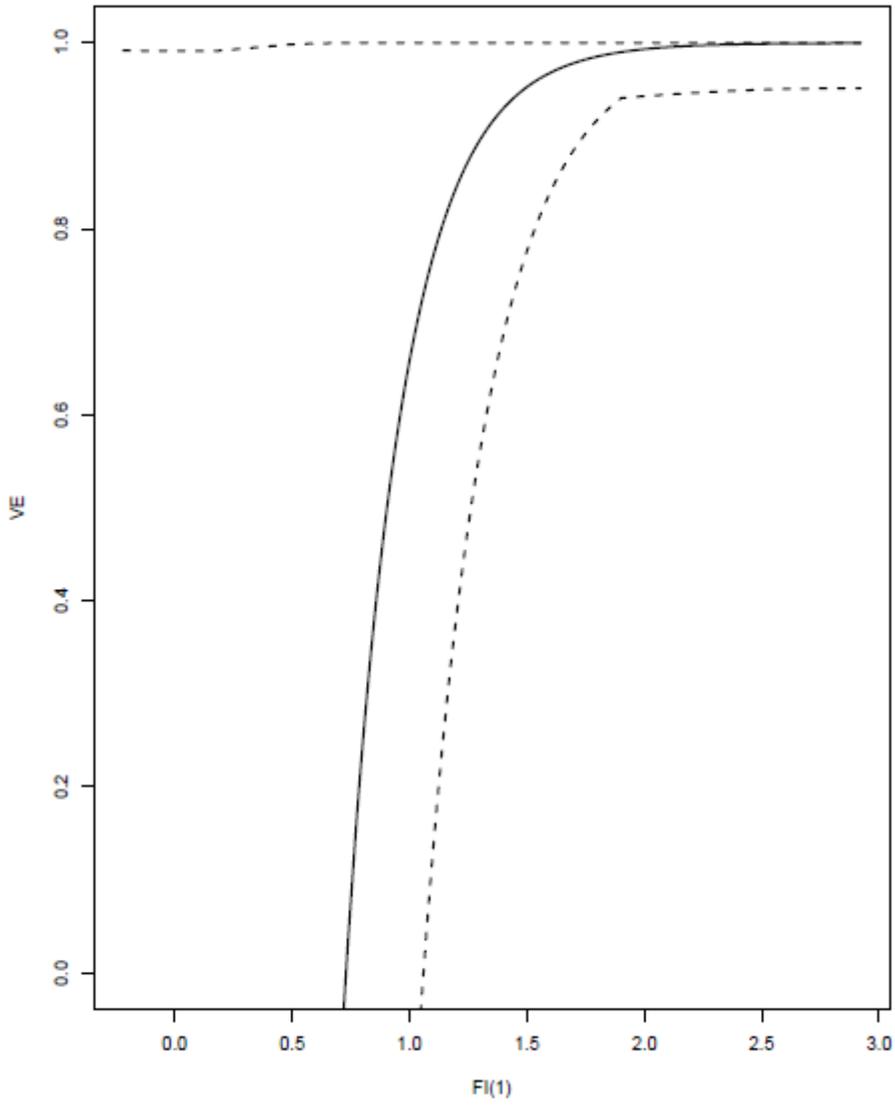


Figure 4

Estimated Vaccine Efficacy curve across levels of vaccine-induced Fold-Increase from baseline to post-vaccination, with 95% condence intervals (dashed lines).