Covariate Adjusted Logit Model (CALM) for Generating Dose-Response Curves from Observational Data with Applications to Vaccine Effectiveness Trials

Nong Shang*, Stephanie Schrag, Rebecca Kahn, Julia Rhodes

Division of Bacterial Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

nms6@cdc.gov

Abstract

Establishing dose-response relationships with observational data is challenging due to confounding covariates and selection bias, particularly when the bias causes marginal distributions of confounding covariates to differ between the sample and the target population. In this scenario, standard causal analysis approaches like g-formula are not applicable without further modeling of the dose-response relationship. To address this limitation, we extend the log-linear structure mean model for binary treatments into a Covariate Adjusted Logit Model (CALM) for continuous treatments, introducing a new causal estimand that measures relative dose-response relationship. Based on one interpretation of vaccine effectiveness, the relative dose response curve can be described as a function that reduces susceptibility to disease, making the adjustment for confounding covariates equivalent to revealing latent susceptibility status. Accordingly, we leverage a Gibbs sampler called the All-or-Nothing (AoN) algorithm for Bayesian and non-parametric estimation of the relative response curve, providing new options beyond regular causal techniques, such as those based on generalized propensity scores. In the context of vaccine effectiveness (VE) clinical trials, CALM addresses the issue of confounding covariates that affect dose-response relationships differently in placebo and vaccine arms. While deriving a common absolute dose-response curve for both arms to evaluate VE is generally infeasible, our model provides a common relative dose-response curve for this purpose. Simulation studies confirm that CALM outperforms alternative methods in scenarios with confounding factors, making it a valuable tool for future immunological response-based vaccine trials.

*To whom correspondence should be addressed.

Key words:

observational studies, dose-response curve, casual risk ratios, correlate of protection, vaccine effectiveness

1. Introduction

Dose-response curves are essential tools for establishing causal relationships between exposure and outcome, typically derived from experimental data. In such experiments, dose levels are pre-selected, and environmental factors are meticulously controlled through systematic designs or randomization. Daniels *and others* [1] provided a comprehensive overview of current approaches for constructing dose-response curves using common parametric functions.

However, real-world observational studies present unique challenges, as environmental factors cannot be controlled. This leads to issues such as confounding variables and potential selection bias, meaning there might be unknown and significant differences between observational sample and the target population in terms of marginal distribution of the environmental factors. Examples of such studies include examining the effects of sun exposure on skin cancer risk (Kricker *and others*. [2]), air pollution on mortality (Di *and others* [3]), and vitamin D intake on coronary heart disease risk (Wang *and others* [4]; Crowe *and others* [5]).

1.1 Antibody disease risk curve for infant Group B Streptococcus (GBS) disease

Our research is motivated by the need to establish antibody-disease risk curves as part of a multi-national effort to license and evaluate maternal vaccines aimed at preventing infant Group B Streptococcus (GBS) disease (Gilbert *and others* [6]; Izu *and others* [7]; Madhi *and others* [8]). Vaccinated pregnant women transfer anti-GBS antibodies to their infants, providing protection against GBS disease. Given the low incidence of infant GBS disease, anticipated licensure trials are likely to use vaccine-induced antibody levels as endpoints rather than actual clinical outcomes. However, the precise relationship between antibody levels and the probability of developing infant GBS disease remains unknown. To establish an antibody-disease risk curve, a series of natural immunity studies are collecting data on infants' antibody levels (exposure), infant GBS disease status (response), and potential maternal and infant confounders using diverse epidemiological designs.

One of the major studies (Rhodes *and others* [9]), built on CDC's Active Bacterial Core surveillance (ABCs) platform, employs a case-control design with a roughly 1:3 case-control ratio. Remnant newborn screening blood spots from infant GBS cases identified through active surveillance are tested for antibodies against the GBS capsular polysaccharide. Controls consist of healthy infants born to GBS-colonized mothers enrolled at selected clinical centers affiliated with

the ABCs sites. The target population for the study is infants of GBS-colonized mothers in the United States (GBS-exposed infants), with the objective of establishing a general antibody protection curve for this population.

In addition to controlling for potential confounders, a primary challenge of this study design—common to many observational epidemiology studies—is the increased likelihood of samples being drawn from specific subpopulations (e.g., the participating clinical centers) rather than the general target population. This type of selection bias might cause the marginal distribution of the confounding covariates to differ between the observational sample and the target population. Consequently, when defining estimands for our research questions, we must ensure that the estimands do not depend on the marginal distribution. Additionally, the corresponding analytic approaches should not yield estimations that depend on the marginal distribution while adjusting for the confounding covariates.

1.2 Absolute and relative dose-response curve under a causal analysis framework

Suppose our dose level is in an interval $[t_0, t_1]$. At treatment level $t \in [t_0, t_1]$, let $Y(t)$ be the counterfactual outcome in the target population, then the dose response curve is traditionally defined in terms of "absolute" response $E(Y(t))$, averaging over the target population. If the dose levels are randomized, then $E(Y(t))$ is identifiable and can be estimated from observed random samples of $(t_i, y_i), i = 1, \dots, n$, with many standard modeling approaches (Daniels *and others* [1]). If the dose levels are not randomized, then $E\big(Y(t)\big)$ is generally not identifiable. However, if all confounding variables X are measured and satisfy a week confoundedness assumption $Y(t) \perp T \mid X$ for all $t \in [t_0, t_1]$ (Hirano and others [10]), then $E\big(Y(t)\big)$ become identifiable and can be estimated from observed data through a g-formula:

$$E\big(Y(t)\big) = E\big(E(Y(t)|X, T = t)\big) = \int g(x, t)\, f(x\,|t) dx \qquad (1.1)$$

if the dose level is randomly observed, or:

$$E\big(Y(t)\big) = E\big(E(Y(t)|X, t)\big) = \int g(x, t)\, f(x)\, dx \qquad (1.2)$$

if the dose level T is controlled. Here $g(x,t) = P(Y = 1 \mid X = x, T = t)$, as the conditional diseases probability, can be estimated from random sample (t_i, x_i, y_i), $i = 1, \dots, n$, assuming no other confounders. $f(x \mid t)$ is the conditional distribution of covariate given $T = t$, and $f(x)$ is the marginal distribution of the covariate. Condition $Y(t) \perp T \mid X$ for all $t \in [t_0, t_1]$ is called week confoundedness assumption because we don't require joint independence of potential outcomes at different treatment levels.

However, without additional modeling assumptions, the g-formula in either Equation 1.1 or Equation 1.2 will result in $E\big(Y(t)\big)$ that depends on the marginal distribution $f(x)$. Hence, they are not suitable for addressing our concerns on both the confounding factors and sample selection bias. In other words, if $f(x)$ is unknown and cannot be estimated from our observational sample, the g-formula cannot be applied to derive the causal absolute dose-response curve $E\big(Y(t)\big)$ defined for the target population.

One solution is to consider a relative dose response relationship $\frac{E(Y(t))}{E(Y(t_0))}$ rather than the absolute relationship $E\big(Y(t)\big)$. When there are only two levels of treatment, $\frac{E(Y(t))}{E(Y(t_0))}$ is just the causal risk ratio, where t_0 is the baseline treatment level, and t is the other level, In the presence of confounding covariate X and if X satisfying the conditional exchangeability assumption, a standard approach is to model the conditional causal risk ratio with a log-linear structure mean model (Robins [11], Dukes and Vansteelandt [12]):

$$\frac{E(Y(t)|X)}{E(Y(t_0) \mid X)} = \exp(\beta t) \qquad [1.3]$$

or $E(Y(t)|X) = g(X)\exp(\beta t)$, where $\exp(\beta)$ is the causal risk ratio conditional on the covariate X, and $g(X) = E(Y(t_0)|X)$ is an unspecified function and we refer it the null disease risk (NDR) in the target population. Because the model assumes the independence of X and T, $\exp(\beta)$ is also the marginal casual risk ratio, provided X satisfies the conditional exchangeability assumption. Various causal analysis techniques, such as the propensity score-based approaches or g-estimation, can be applied to estimate $\exp(\beta)$ (Robins [11,13]).

We extend the log-linear structure mean model in Equation 1.3 from two treatment levels to continuous treatment levels as in:

[1.4]

Here again, represents the null disease risk (NDR) at the baseline treatment level . If X satisfies the weak unconfoundedness assumption, is the marginal causal risk ratio function of which can be interpreted as a relative causal dose response curve. This curve is designed to reduce the disease risk from the NDR with a constrain that) = 1. Although not necessary, we generally assume , indicating that the dose is with protective effect. This curve can be modelled according to our understanding of the underlying dose-response relationship, such as using a monotone function, or the popular two parametric logit function $h(t) = \frac{1}{1+e^{\alpha+\beta t}}$, truncated and scaled to ensure) = 1. Acknowledging the significance of the logit function in the history of dose response curves, we refer to the model in Equation 1.3 as a Covariate Adjusted Logit Model (CALM), although the model is flexible, allowing for various parametric or non-parametric approaches to model

Our next objective is to develop analytic methods to estimate and make sure that the estimation does not depend on the marginal distribution . Recent advancements in generalizing propensity scores from binary to continuous exposures (Imbens [10], Hirano and Imbens [14], Brown et al. [15], Fong et al. [16]) offer promising avenues for developing causal analysis solutions. For instance, Hirano and Imbens propose a two-step procedure for estimating and introduce a technique known as "blocking" to assess covariate balance post-adjustment using propensity scores. Additionally, inverse probability weighting (IPW) with generalized propensity scores has been explored for similar problems, although more studies are needed in evaluating their performance on estimating

In this paper, we adopt a traditional regression approach to develop a practical and efficient estimation method to estimate h(t), leaving the development of causal analysis approaches for future research. The main difference between our approach and casual approaches is that we

model the NDR $g(x)$ directly, while the propensity score-based approaches focus on modeling the propensity scores. However, modeling $g(x)$ together with $h(t)$ leverages a unique interpretation of the CALM model, where $h(t)$ represents the proportion of subjects who are susceptible to diseases. Consequently, adjusting for confounding covariates is equivalent to revealing the latent status of susceptibility to diseases. To facilitate this, we have developed a new Gibbs Sampler-based algorithm, referred to as the AoN (All or Nothing) algorithm, which offers more flexibility in modeling $h(t)$, such as using a non-parametric isotonic function.

1.3 Traditional regression modeling approaches for dose response curve

Before we proceed with our modeling approach, we would like to address an often-overlooked aspect in the common use of logistic regression models for constructing covariate-adjusted dose-response curves.

Regular regression modeling approaches, such as linear logistic regression models, have been applied extensively in practice to derive covariate adjusted dose-response curves (Daniels and others [1] and Cox Jr. and others [17]). The idea originated from scenarios that there are no confounders, or if treatment (dose) levels are assigned through randomization. In that case, the marginal causal dose-response curve can be modeled and estimated using observed dose levels and outcomes alone. For example, employing a univariate logistic regression model allows for the derivation of a two-parameter logit function to describe the dose-response relationship.

Building on this idea and recognizing the extensive use of multivariate logistic regression models in epidemiological and medical research to adjust for covariate effects on odds ratio, it has been proposed to use the same logistic regression model to generate covariate-adjusted dose-response curves: the intercept and the coefficient for the dose variable derived from the model serve as the estimated parameters for the two-parameter logit function.

However, a critical limitation of this conventional practice is often overlooked: logistic regression models adjust only for the odds ratio or the slope term of the dose-response curve, not for the

intercept term. Consequently, the derived dose-response curve is not covariate-free but represents the curve at a specific covariate value, typically at X = 0. This limitation means that the estimated curve may not accurately reflect the underlying dose-response relationship unless the curve can be characterized solely by the slope parameter, such as in cases of exponentially decreasing function.

Our proposed method, the Covariate Adjusted Logit Model (CALM), offers a solution to mitigate the drawbacks from both the conventional regression approaches and the g-formula based causal approaches. It conceptualizes the impact of dose as a relative reduction in disease risk from the null disease risk (NDR). With the assumption of independence between treatment and confounding covariates added to the weak unconfoundedness assumption, a relative dose-response curve is defined as an extension of the causal risk ratio for binary treatment. The relative dose-response curve not only has a marginal causal interpretation but has potential to be estimated from biased observational samples. This paper presents CALM and its application in vaccine effectiveness studies.

In Section 2, we introduce CALM, outlining its structure, properties, computational methods, and interpretative framework. Section 3 applies the CALM model to establish an antibody disease risk curve for our Group B Streptococcus (GBS) research, using an interim dataset as an illustrative example. Section 4 explores the complexities of employing immunologic response curves as evaluation metrics for vaccine effectiveness (VE) in vaccine trials, focusing on the application of CALM for VE estimation. Section 5 conducts a comprehensive simulation study to assess CALM's performance and robustness against sample selection bias. We simulate vaccines with predetermined vaccine effectiveness (VE) and evaluate various approaches for deriving immunologic response curves to estimate the vaccine's effectiveness in clinical trials. Finally, Section 6 concludes by summarizing the key findings and implications of our research, highlighting the contributions and potential applications of the CALM model in epidemiological studies and vaccine trials.

2. Covariate Adjusted Logit Model (CALM)

2.1 Relative dose response curves with CALM:

As an extension of log-linear structure mean model (Robin [11], Dukes and Vansteelandt[12]), our CALM model as presented in Equation 1.4, provided a causal definition of relative dose response curve . As we are intended to use traditional regression model to study , and leaving developing of causal analysis procedure for future research, we will describe CALM in following regression model:

[2.1]

with constrains that 0 ≤ $h(t)$ ≤ 1, and $h(t_0)$ = 1, implying that reduces the disease risk from the NDR to , as t increases from the baseline dose level .

We first examine a few special situations of the CALM model [2.1 to see its connection to some previously applied models in dose-response research. If the dose levels are randomized, then is identifiable and no confounders need to be considered. In that case, if we let , then model in Equation 1.3 can be presented as which is just the causal expression of the Scaled Logit Model introduced in Dunning [18]:

Another special situation is when $h(t)$ takes an exponential form, such as , then CALM exhibits a formulation similar to a proportional hazard model:

$$g(x,t) = g(x)e^{\beta(t-t_0)}$$

In this formulation, the NDR $g(x)$ is considered a nuisance when estimating the dose-response relationship $h(t)$, akin to the baseline survival function in a Cox regression model when estimating proportional hazard ratios. Further, if we model with a logit function, then the parameter can be estimated easily through a regular linear logistic regression model.

More general parametric function forms for can be found in Daniels *and others* [1] and Dunning *and others* [19]. The CALM model can also be extended to consider non-parametric formulation of $h(t)$. For example, a semi-parametric approach with splines can be seamlessly integrated into CALM. A more versatile non-parametric model can be explored:

$$h(t) = \frac{1}{1 + e^{s(t)}}$$

where $s(t)$ represents a smoothing function of t, such as a smoothing spline. To ensure remains a monotone function, isotonic regression might also be applied as a non-parametric option.

2.2 Interpretations of CALM's dose response curve

Moving on to the interpretations of $h(t)$ using concepts and terminology from vaccine effectiveness research, it's important to note that while these interpretations may not universally apply to all dose-response relationships, they offer a fresh perspective for understanding the nature of a relative dose-response relationship. Additionally, they pave the way for a computational algorithm for constructing the dose-response curve non-parametrically.

In the context of vaccine effectiveness study, two modes of vaccine effects exist at the individual subject level (Halloran *and others* [20], WHO [21]):

- Leaky Mode: This suggests that a vaccine uniformly reduces the disease risk for all individuals by a constant proportion.
- All-or-Nothing Mode: This mode assumes that individuals can be classified into two groups: susceptible to disease and immune. Susceptible individuals share the same disease risk (as the NDR), while immune individuals have zero risk of developing the disease. Under this mode, the vaccine effect is interpreted as reducing the proportion of susceptible individuals.

Applying these two interpretations of vaccine effectiveness to our dose-response curve $h(t)$ is straightforward. Under the Leaky mode, the disease risk for each individual at dose level t is reduced by $1 - h(t)$ compared to their NDR at t_0. Alternatively, under the All-or-Nothing mode (Dunning [18]), all individuals with zero (or baseline) dose are considered susceptible. As the dose level increases $h(t)$ is interpreted as the remaining proportion of susceptible individuals at dose level t. This interpretation provides a new understanding of the relative dose-response curves: the protective effect of the dose lies in its ability to reduce the latent susceptibility to disease. In other words, under the CALM model, while the baseline risk of disease NDR may vary based on confounding covariates, the latent susceptible status depends on the antibody (dose) level only. If we were able to observe the latent susceptibility status of all individuals, the relative dose-response curve could be constructed directly, without the need to control for confounding covariates. In other words, the process of adjusting for confounders is equivalent to revealing the latent susceptible status of the individuals.

2.3 Computation methods

2.3.1 *Computation methods for parametric models*

Parametric modeling is frequently employed to characterize dose-response relationships, with various models proposed in the literature (Daniels *and others* [1]). The specific parametric form chosen depends on our understanding of the underlying biological process, although in practice, generic models are sometimes used without considering these processes.

If we also opt for a parametric model to describe the Null Disease Rate (NDR) $g(x)$, we can estimate the parameters for both $h(t)$ and $g(x)$ easily through maximizing the log likelihood function:

$$logL = \sum_{i=1}^{n} (y_i \log(g(x_i)h(t_i)) + (1 - y_i) \log(1 - g(x_i)h(t_i))) \tag{2.2}$$

with a non-linear maximum likelihood estimation (MLE) algorithm.

Standard Bayesian approaches can also be applied to CALM with a proper selection of priors for the model parameters of $h(t)$ and $g(x)$. However, we will introduce a new Gibbs algorithm based on the All-or-Nothing interpretation of the dose-response curve $h(t)$ in next subsection. The algorithm offers more flexible options to estimate $h(t)$.

2.3.2 All-or-Nothing (*AoN) algorithm*

For each observed data combination (t_i,x_i,y_i), we introduce a binary latent variable z_i to represent the disease susceptible status of subject *i*. Specifically, $z_i = 1$ indicates that the subject is susceptible, while $z_i = 0$ indicates that the subject is not. Since non-susceptible subjects have zero probability of developing the disease, and all susceptible subjects with the same covariate value *x* share the same NDR risk $g(x)$, we can estimate $g(x)$ using only the susceptible subjects. At the same time, the binary susceptible status will be used to estimate $h(t)$.

The conditional probability of y_i given *x*,*t*, and *z* can be expressed as follows:

$$\begin{aligned} y_i \mid z_i = 1, x_i, b_X &\sim g(x_i, b_X) \\ y_i \mid z_i = 0 &= 0 \end{aligned} \tag{2.3}$$

Here, b_X is the parameter vector of $g(x)$. The probability of the susceptible status z_i for a given antibody value t_i is determined by the function $h(t)$ with parameter b_T:

$$z_i \mid t_i, b_T \sim h((t_i, b_T) \tag{2.4}$$

Adding prior distributions to the two parameter b_X and b_T:

$$\begin{aligned} b_X &\sim p(s_X) \\ b_T &\sim p(s_T) \end{aligned} \tag{2.5}$$

with s_X and s_T as super parameters of the corresponding prior distributions. The full joint likelihood function is then:

$$\prod_{i=1}^{n} [g(x_i, b_X)^{y_i} (1 - g(x_i, b_X))^{1-y_i}]^{z_i} 0^{y_i(1-z_i)} h(t_i, b_T)^{z_i} (1 - h(t_i, b_T))^{1-z_i} p(b_X \mid s_X) p(b_T \mid s_T)$$

(2.6)

with a notation specification that $0^0 = 1$.

Based on the full joint likelihood function in Equation 2.6, we obtain the posterior distribution of the susceptible status z_i given d_i and the function values of $g(x_i, b_X)$ and $h(t_i, b_T)$:

$$\begin{cases} P(z_i = 1 \mid y_i = 1) &= 1 \\ P(z_i = 1 \mid y_i = 0) &= \frac{(1-g(x_i))h(t_i)}{(1-g(x_i))h(t_i)+(1-h(t_i))} \end{cases} \tag{2.7}$$

The first part of Equation 2.7 implies that a subject with disease must also be a susceptible subject.

The posterior distribution for parameter b_X as in $g(x)$ is:

$$b_X \mid D, Z \sim \prod_{i:z_i=1} g(x_i, b_X)^{y_i}(1 - g(x_i, b_X))^{1-y_i} p(b_X \mid s_X) \tag{2.8}$$

In other words, the susceptible subjects, i.e. those with $z_i = 1$, will be used to estimate model parameters of the NDR $g(x)$ through a Bayesian binary regression.

The posterior distribution for parameter is also straightforward:

$$b_T \mid D, Z \sim \prod_{i}^{n} h(t_i, b_T)^{z_i}(1 - h(t_i, b_T))^{1-z_i} p(b_T \mid s_T) \tag{2.9}$$

Accordingly, the susceptible status of all subjects is used to estimate the model parameter b_T of $h(t)$ as in another Bayesian binary regression.

Thus, by introducing the concept of latent susceptible status, we can simplify a complex regression problem into two simpler regression tasks, one for each component of CALM. This approach, rooted in the All-or-Nothing interpretation of the vaccine effect, is referred to as the All-or-Nothing (AoN) algorithm. Specifically, if separate logit functions are used to model $g(x)$ and $h(t)$, a Bayesian logistic regression procedure can be employed to draw posterior samples for the parameters b_X and b_T. The Polya-Gamma Gibbs sampler is a popular and efficient algorithm to generate these samples, and the UPG R package provides a convenient implementation of this sampler (Polson *and others* [22] Choi and Hobert [23]; Fruhwirth-Schnatter *and others* [24]).

Non-parametric modeling of in CALM not only address potential model misspecification when specific parametric forms lack biological support, but also leverages efficient methods for modeling one-dimensional relationships between a univariate continuous exposure and a binary response within the Bayes framework. For instance, incorporating Bayesian isotonic regression approaches, as discussed by Neelon and Dunson [25] and Li and Fu [26], into our AoN algorithm allows adjustment for the impact of confounding covariates.

3. Antibody-disease risk curve from GBS study

The US GBS study described in the introduction section has completed its data collection process, and the final data will be published and analyzed in a separate application paper. Here, we utilize the interim analysis dataset from this study to illustrate the issues addressed in this paper and to evaluate the utility of the CALM model in establishing the relative antibody-disease risk curve.

For simplicity, we focus on one GBS serotype, serotype IA, as our example. A total of 39 early onset cases and 202 controls were included in the interim analysis. During the study design stage, subject matter experts constructed a Directed Acyclic Graph (DAG) to identify potential confounders. In this example, three potential confounders were considered: gestational age, maternal chorioamnionitis, and exposure to intrapartum antibiotic prophylaxis (IAP). However, all collected cases from the interim data set has a single value of IAP (= 1) due to the high rate of IAP use in U.S.A; hence, the variable is dropped in this analysis. The other two variables were dichotomized for simplicity of interpretation Additionally, geographic region was included as another potential confounder. This inclusion was necessary because more than half of the controls were enrolled from one state, which accounts for only 5% of the US population. Given the significant regional variations in disease risk and antibody distribution (see Table 1), we added region (as a binary variable indicating if in the state or not) as the third confounder.

3.1 The null disease risk and antibody distributions by confounders

Among the 241 collected subjects (39 cases and 202 controls), 57 (19 cases and 38 controls) had antibody measurements below the lower limit of quantification, indicating no antibody protection. This sizable proportion of subjects enables us to obtain reasonable estimates of the NDR, stratified by each of the three covariates. As this is a case-control design, we compare the percentage of cases instead. The results are presented in Table 1.

As shown in Table 1, the percentages of cases to controls, and consequently the null disease risk, vary significantly by region but not substantially for the other two covariates. In Table 1, we also compare the mean of the logarithm of antibody distributions between the two strata of each covariate. The variable gestational age shows significant differences, while the other two variables are not, possibly due to small sample size in this interim data. Nonetheless, since the covariates

were identified as a potential confounder by our subject knowledge-based DAG diagram, we will include them in our final model to adjust for their effects.

Table 1. Null disease rate and antibody distributions by covariates

Covariates	strata	Percentage of cases among those with minimal antibody	p-values for percentage of cases between strata of covariates	Mean of logarithm of antibody values	p-values to compare antibody distributions
Chorioamnionitis	1	0.455	0.35	-2.59	0.14
	0	0.304		-1.54	
Geographic region	1	0.176	0.002	-1.49	0.23
	0	0.565		-2.09	
Gestational age	1	0.362	0.33	-1.48	0.02
	0	0.200		-3.22	

3.2 Examining CALM's model assumptions:

CALM assumes independent effects of covariates and antibody values. We examine the assumption through stratifying our covariates. Due to sample size limitations, we focused on examining each of the three covariates by comparing their corresponding two strata. Isotonic regression was applied to generate the non-parametric estimation of response curves within each stratum.

The left three panels of Figure 1 display the "raw" isotonic regression curves stratified by covariate strata. These curves show significant differences across strata, particularly for Chorioamnionitis and Region. However, this difference could be influenced by variation in NDR as indicated in Table 1. To remove this difference, we scale the isotonic curves to a value of 1.0 at the

minimal dose level, transforming them into relative curves. The right three panels of Figure 1 illustrate that the scaled relative curves across strata are much closer to each other, which supports CALM's independence assumption.

To illustrate the influence of marginal covariate distribution $f(x)$ on the g-formula (Equation 1.2) estimated counterfactual dose response curve E(Y(t)), we use the variable Region as example. We first fitted a joint logistic regression model with the main effect of antibody level and region, together with their interaction. Subsequently, we applied the g-formula to estimate marginalized dose-response curves using both the sample-based marginal distribution of the covariate and a population-based proportion of $(0.05,\ 0.95)$, reflecting that the state accounts for only 5% of the national population. Figure 2 displays the results, revealing substantial discrepancies between the estimated curves (solid curves). To examine if the g-formula works for relative dose response curve, we scaled each individual dose response curves first so that they achieve a value of 1.0 at the minimal dose level before applying g-formula. The dotted curves in Figure 2 illustrates reducing of the discrepancy. although some residual influence remains.

Figure 1 Isotonic regression curves stratified by covariates.

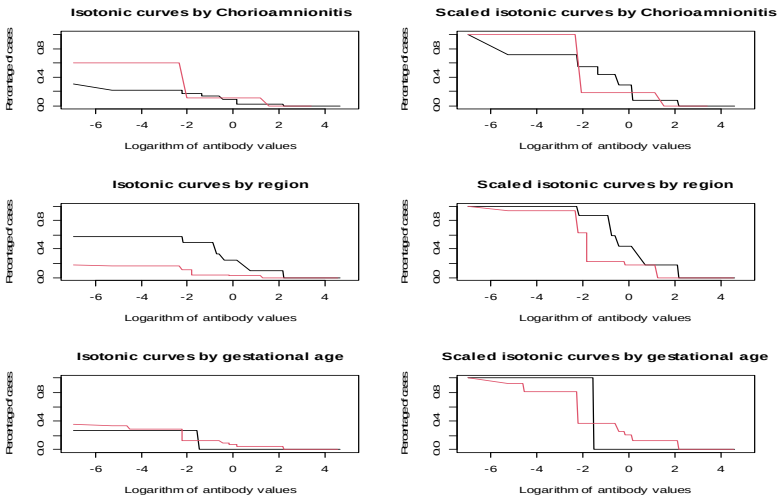
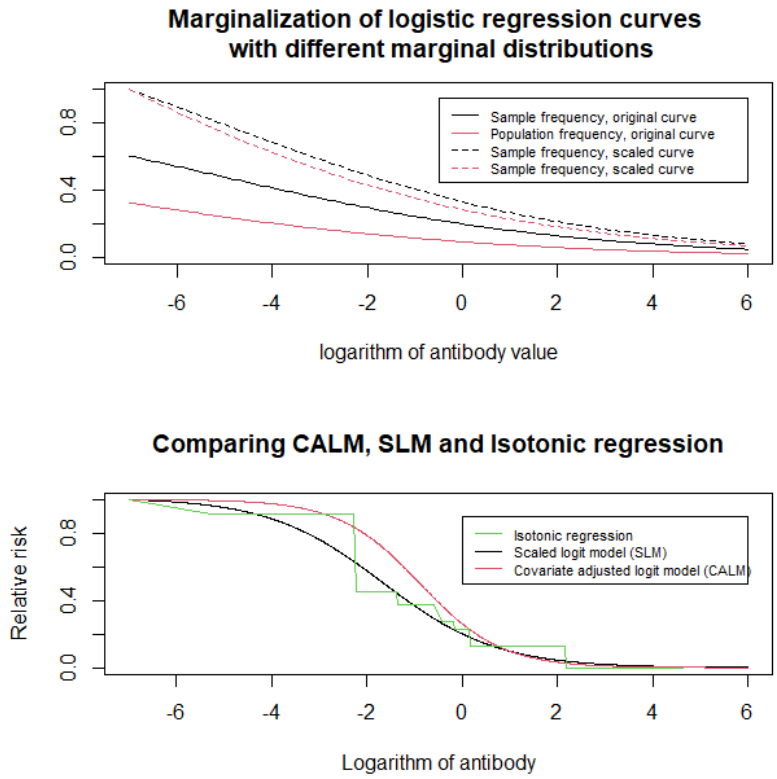


Figure 2. Dose response curves constructed by g-formula and by CALM



3.3 Covariate adjusted dose response curves with CALM

We applied CALM model to generate covariate-adjusted dose-response curves for our interim data. The fitted relative disease risk curve is depicted as the red curve in the bottom panel of Figure 2. We compared this curve with the one from the regular scaled logit model (black curve), which does not adjust for any confounders. Additionally, we included the overall isotonic regression curve (green curve) for comparison. The similarity between the non-parametric isotonic regression and the scaled logit model suggests that the two-parameter logit function is a reasonable choice for our curves. The figure also highlights the effect of adjusting for confounders, illustrated by the difference between the red and black curves.

We further explored interactions between individual covariates and antibody levels by incorporating interaction terms in the CALM model, thereby allowing the slopes of the logit function to vary across different covariates levels. Likelihood ratio tests were conducted to assess the significance of these interactions. The corresponding p-values for the three covariate interactions -- 0.82 (chorioamnionitis), 0.67 (region), and 0.14 (gestational age) -- support the independence assumption for the treatment and covariate effect in the CALM model.

4. Applications of CALM's relative dose-response curve in vaccine clinical trials

Traditionally, vaccines undergo licensure based on demonstrated efficacy in preventing clinical disease endpoints. However, certain scenarios -- such as evaluating next generation vaccines against existing standards, developing vaccines for low-incidence diseases, or responding to biological threats -- can pose challenges in demonstrating clinical disease protection. In such cases, immunologic response can serve as a surrogate for licensure evidence, a concept acknowledged by regulatory bodies (FDA and CBER [27], Rajam *and others* [28] ; Miller *and others* [29]; Cherry *and others* [30]).

This process of identifying immunologic response as a surrogate for clinical endpoints is termed establishing a "correlate of protection." This involves satisfying several causal criteria, often referred to as the Prentice criteria (Prentice [31], Qin *and others* [32]): demonstrating that the vaccine reduces disease risk, influences the immune response, links the immune response with disease risk, and ensures that the vaccine's total effect on disease risk reduction is captured by enhancing the immune response. Once these criteria are met, the next step is to identify an immune response curve $g(t) = P(y = 1 \mid t)$ and used it to evaluate VE in future clinical trials. Often, this curve is approximated by a dichotomous function with a threshold, where an immune response below the threshold indicates vaccine failure (See Siber *and others* [33], Chen *and others* [34]; Donovan *and others* [35]; Van der Laan *and others* [36]).

The central research question thus revolves around estimating a common antibody-disease risk curve applicable across both study arms. However, previous research on correlates of protection has primarily focused on antibody levels and disease status alone, often overlooking confounding covariates. The existence of such a common covariate-free function $g(t)$ remains an unanswered

question where there is a presence of confounding covariates. Section 4.1 elaborates on this issue, highlighting the impracticality of deriving such a function for unbiased VE estimation, employing an absolute dose-response curve. Nonetheless, a common relative dose-response curve from the CALM model offers a promising solution that outperforms other suggested alternatives. Section 4.2 will discuss these alternatives in detail.

4.1*Evaluation VE using immunologic response curves.*

In a typical two-arm vaccine trial comparing the vaccine arm to the placebo arm, clinical endpoints (disease status), immunologic response (antibody levels), and other covariates are considered as random samples from the joint distributions: $G_p(y,t,x)$ and $G_v(y,t,x)$ for the placebo and vaccine arms, respectively. The joint distribution can be expressed as:

$$G_a(y,t,x) = g(x,t)f_a(t \mid x)f(x) \tag{4.1}$$

In this equation, a denotes the study arms v and p, indicating the vaccine and placebo arms. $g(x,t) = P(y = 1 \mid x,t)$ represents the probability of disease given a specific antibody level t and covariate x, $f_a(t \mid x)$ is the conditional distribution of antibody levels at x for arm a, and $f(x)$ is the marginal distribution of the covariates in the clinical trial's target population. Notably, $f(x)$ remains the same for both arms due to the randomization design of the clinical trial, and $g(x,t)$ as the general function of disease probability is also the same across arms. However, the distributions of antibody values between the study arms, namely $f_v(t \mid x)$ and $f_p(t \mid x)$, are different and vary by the confounding variable X.

The VE is then defined as the proportions of diseases being prevented from vaccination:

$$VE = 1 - \frac{\int G_v(y,x,t)dxdt}{\int G_p(y,x,t)dxdt} \tag{4.2}$$

Where the expectation is taken over the target population where the clinical trial is conducted. If the clinical endpoints (denoted as y_i^a are observed in both arms of the clinical trial, VE can be simply evaluated

$$VE = 1 - \frac{\frac{1}{n_v}\sum I_{(y_i^v==1)}}{\frac{1}{n_p}\sum I_{(y_i^p==1)}}$$

If the observations can be considered as simple random sample from the joint distribution $G_a(y,t,x)$ in both study arms respectively.

The purpose of establishing immunological surrogate is to find an immunological response curve $g(t)$ and use it to estimate the VE as in following equation:

(4.3)

where $t^v, i = 1,...,n_v$ and $t^p, j = 1,...,n_p$ are the observed immune response, in the vaccine and placebo arms, assuming randomly generated from the marginal distribution and respectively.

One sufficient condition for Equation 4.3 to provide an unbiased estimation of VE, as defined in Equation 4.2, is that both denominator and numerator in Equation 4.3 are unbiased estimation of their counterparts in Equation 4.2. In order to do that, needs to satisfy following condition in both vaccine and placebo arm:

$$g(t) = \int g(x,t) f(x \mid t) dx$$

which is just the causal dose response curve as described in Equation 1.1. As and remains the same across the study, we would require following condition holds for having a common function :

(4.4)

Equation 4.4 is satisfied when x is not a confounder because then , for both arms a=v and p. However, if x is a confounder, Equation (4.4) generally does not hold. For instance, since the vaccine induced antibody level typically exceed naturally generated levels, in the vaccine arm will be closer to 1.0 compared to in the placebo arm, making Equation 4.4 unsatisfactory. In such cases, it may not be feasible to have such a function , let alone establish a binary threshold, to evaluate VE unbiasedly.

From a causal analysis perspective, when confounders are present, the two study arms effectively represent two distinctive target populations because the conditional antibody distributions vary between the study arms. While the randomization procedure in clinical trial design ensures consistency in the marginal distribution of covariate across the two study arms, it

does not align or . Consequently, the counterfactual dose response curves based on the g-formula in Equation 1.1, which provides unbiased estimations for the numerator and denominator in Equation 4.2 via Equation 4.3, are unlikely to be identical across the study arms.

One remedy to the problem is to utilize different marginalized functions in each study arm, as follows:

$$VE = 1 - \frac{\frac{1}{n_v}\sum_{i=1}^{n_v} g_v(t_i^v)}{\frac{1}{n_p}\sum_{j=1}^{n_p} g_p(t_j^p)} \tag{4.5}$$

However, deriving $g_v(t)$ in the vaccine arm necessitates knowledge of the conditional antibody distribution which is often challenging due to difficulties in obtaining enough representative vaccinated subjects before conducting large-scale clinical trials.

Another potential solution is to assume that the conditional immune response distribution does not depend on the covariate in the vaccine arm. The assumption may be approximately reasonable if the vaccine induces significantly higher antibody levels compared to those naturally induced. In essence, this assumes that the effects of confounder diminish in the vaccine arm due to the vaccine's high potency. Consequently, the marginalized function $g_v(t)$ can be assessed approximately by simple population average, without relying on antibody distributions among vaccinated subjects:

$$g_v(t) = \int g(x,t)f(x)dx$$

4.2 *Using the relative dose-response curve $h(t)$ from CALM to evaluate VE*

Following the "All-or-Nothing" interpretation of vaccine effectiveness, as proposed by Halloran *and others* [15], where vaccine effectiveness (VE) is defined as the vaccine's impact on reducing the proportion of susceptible individuals, CALM's relative dose-response curve $h(t)$ can be used to measure the ability of antibody values to reduce this proportion. With this interpretation of VE, we can estimate VE as follows:

$$VE = 1 - \frac{\frac{1}{n_v}\sum h(t_i^v)}{\frac{1}{n_p}\sum h(t_j^p)} \tag{4.6}$$

If the objective remains to measure VE in terms of reducing the disease risk rather than the proportion of susceptible individuals, Equation 4.6 can still provide an unbiased estimation of VE if there are no confounders. For instance, if the conditional antibody distribution $f_a(t|x)$ does not depend on the covariate x, then we have:

$$\int g(x)h(t)f_a(t)f(x)dxdt = \int g(x)f(x)dx \int h(t)f_a(t)dt$$

Here, the first integration on the right side of above equation is the same for both the vaccine and placebo arms, cancelling out in the VE formula of Equation 4.2, resulting in VE estimation using Equation 4.6.

However, if confounders exist, applying CALM's relative response curve $h(t)$ with Equation 4.6 may introduce bias too, as reducing percentage of susceptible subjects does not perfectly equate to reducing diseases risk. Nevertheless, as demonstrated in our simulations (Section 5), CALM's approach still provides a robust and less biased VE estimation compared to other common alternative approaches.

5. Simulation studies

5.1 *Simulation Design*

Our simulations have broad applicability to dose-response curves, but we specifically frame them within the context of antibody-disease type of dose-response curves derived from vaccine studies. In the initial set of simulations, we demonstrate how CALM can adjust for confounders and sample selection bias. Subsequently, we simulate a vaccine with a given vaccine effectiveness and compare CALM with other approaches in terms of their performance in estimating the true vaccine effectiveness (VE).

5.1.1 *Confounding covariate and conditional antibody distribution*

To simplify and underscore our main point, we introduce a single binary covariate denoted as X as a confounder in our simulations, implying that the study population is a mixture of two distinct

groups, where $X = 1$ represents a high-risk group and $X = 2$ represents a low-risk group with $P(X = 1) = 0.25$. The conditional distributions of antibody levels are $f(t \mid X = 1) \sim N(-4,1.5)$ and $f(t \mid X = 2) \sim N(1.5,2)$ respectively. The range of dose values chosen mirrors logarithmically transformed values, as commonly employed for antibody assays. We set $t_0 = -6.5$, indicating a limit of detection on the original scale of $exp(-6.5) = 0.0015$. Any values generated from the normal distributions that fall below -6.5 will be truncated at 6.5. The top left panel of Figure 3 illustrates the distributions of antibody assay values for the two risk groups, represented by solid lines.

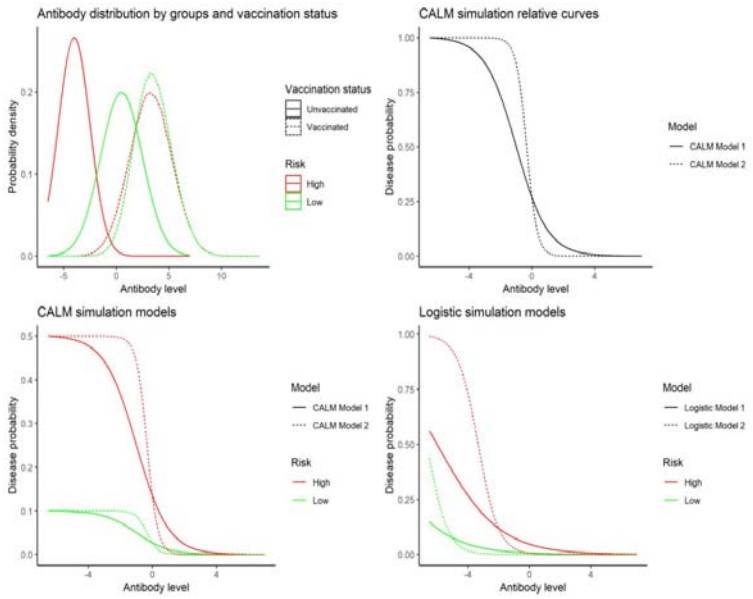
5.1.2 *The disease probability function*

We then use either CALM or logistic regression model to simulate the underlying disease response function $g(x,t)$. Here we assume there is not any other unmeasured confounders. Hence, $g(x, t)$ is identifiable and can be unbiasedly estimated using observed data. Under the CALM modeling structure, we assume a null disease risk (NDR) of 0.5 for the high-risk group and 0.1 for the low-risk group, namely, $g(x) = 0.5$ for $x = 1$ and 0.1 for $x = 2$.

The relative response curve $h(t)$ is modeled as $h(t) = 1/(1 + ex(a + b \cdot t))$ in the range of $t >= -6.5$ and scaled by $h(t)/h(-6.5)$. The function $h(t)$ is presented in the top right panel of Figure 3. for two different scenarios, Model1: (a=1,b=1) and Model 2 (a=1,b=3). The probability of disease, denoted as $g(x,t) = g(x)h(t)$ is depicted in the lower left panel of Figure 3.

In order to explore CALM's performance when CALM's model assumption does not hold, we also consider disease function $g(x,t)$ using a logistic regression model: $logit(g(x,t)) = \alpha + \beta t + \gamma x$ with the following combinations of parameters, Model 1: $\alpha = -1, \beta = -0.5, \gamma = -2$ and Model 2: $\alpha = 0, \beta = -1.5, \gamma = -5$ as in the lower right panel of Figure 3. Notice that, at this scenario, the covariate is an effect modifier under CALM model.

Fig. 3. Simulation designs



5.2 *General performance of CALM*

We now proceed to examine CALM's ability to generate relative dose-response curves under selection bias of observational sample. We first use a large sample size of 10,000 to investigate whether biased sampling affects the estimation of $h(t)$ in CALM. Additionally, we explore alternative ways to model $h(t)$ beyond the two-parameter logit function and evaluate Bayesian implementation of AoN algorithm.

We conducted experiments using two different samples with varying proportions of high risk group $P(X = 1) = 0.1$ and $P(X = 1) = 0.4$ to assess whether biased observational sample estimation of CALM's $h(t)$. The true relative dose response curve was specified as $h(t) = 1/(1 + exp(1 + t))$, scaled and truncated at $t0 = -0.65$.

For each sample, we applied four different methods for h(t). The first two are using a parametric CALM model with the correct function form, namely assuming $h(t) = 1/(1+exp(a+bt)$, with parameters estimated either with a MLE or a Bayesian approach. For the Bayesian approach, Normal priors were used for both parameters of $g(x)$ and $h(t)$, with standard errors set at 10,

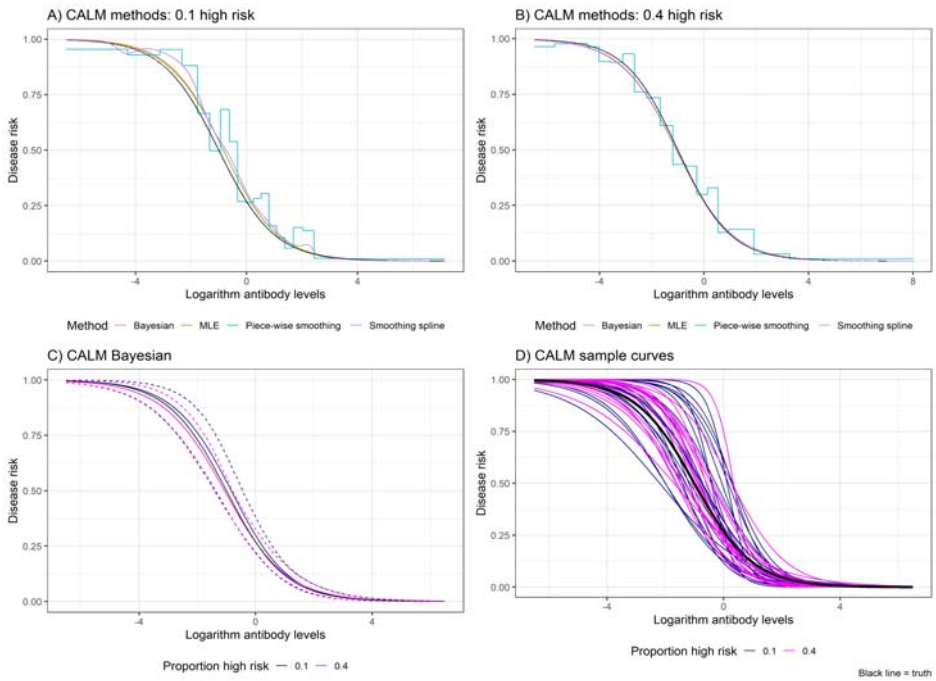
indicating non-informative prior. The next two methods are non-parametric without specifying the function form of $h(t)$. In the first non-parametric CALM model, a piece-wise step function was employed for $h(t)$, dividing the range of assay values into 20 bins, each containing 500 subjects. The AoN (All-or-Nothing) algorithm was applied with Beta priors B(1/2,1/2) for each of the 20 discrete values of $h(t)$. The second non-parametric CALM model utilized the modified AoN algorithm, incorporating a smoothing spline to model $h(t)$.

The top two panels of Figure 4 depict the estimated $h(t)$ curves using the two different proportions of the covariate. In both graphs, the relative dose-response curves constructed from the four CALM models are compared. Remarkably, all estimations closely fit the true dose-response curves, suggesting effective adjustment for confounding effects and selection bias in the sample, irrespective of whether the functional form of the true dose-response curve is specified in the model or not.

To provide further insight into the estimation accuracy of CALM, we present the Bayesian estimated dose-response curves along with their corresponding 95% credible intervals in the lower left panel of Figure 4, under two different marginal distributions (P(X = 1) = 0.1 or 0.4) of the covariate. It is evident that the bias introduced by observational samples has minimal impact on CALM estimations.

To assess the model's performance with smaller sample sizes, we generate 20 samples, each containing 1,000 observations, from each of the two biased scenarios. Subsequently, we obtain a total of 40 dose-response curves using the CALM parametric model. The lower right panel of Figure 4 displays these 40 estimated curves. Remarkably, there is no significant differentiation observed between the two scenarios with varying marginal covariate distributions, and the majority of the curves adeptly capture the true dose-response relationship.

Fig. 4 . Performance of CALM in estimating relative dose response curve



5.3 *Simulation to compare immunological response approaches in estimation of VE*

We will general observational samples of size 1,000 from the study population as described in the previous subsection, with the given conditional antibody distribution $f(t|\ x)$ and the four different underlying disease probability functions $g(x,t)$: two CALM based and two logistic model based. We will also consider the option that the samples are biased with marginal covariate distribution with $P(x = 1) = 0.1$ or unbiased, with $P(x = 1) = 0.25$.

Each simulated samples will be used to construct an antibody disease risk curves without incorporating any additional information beyond the observed sample. Subsequently, estimated curves will be employed to assess vaccine effectiveness (VE) in a simulated vaccine trial, with a designed true VE of 80%. We will compare the performance of 10 different methods for constructing the antibody-disease functions or thresholds in estimating the true VEs, based on 1,000 repeated samples. Further details regarding the vaccine trial and estimation methods are outlined below.

5.3.1 *Simulation of vaccine trial*

The placebo arm will be simulated by taking a simple random sample from the study population. The vaccine effect is defined as inducing an antibody level with a distribution of $N(\mu,2)$. Generally, a higher value of μ indicates a higher true VE. The observed antibody level for a vaccinated subject will be the higher of the subject's naturally induced and the vaccine induced antibody level (as in the placebo arm), represented as:

$$t_i^v = \max\,(t_i^p, t_i^{v_0})$$

where t_p^i is sampled from the mixture antibody distribution among unvaccinated subjects, and $t_i^{v_0}$ *is* sampled from the vaccine-induced antibody distribution. We choose μ = 3.5 as it results in a true VE of 80% under our simulation design.

The resulting antibody distributions are illustrated by the dotted lines in top left panel of Figure 3. As expected, the difference between the two risk groups becomes much smaller in the vaccine arm.

To minimize the sampling effect in evaluating the methods, we assume that a very large clinical trial is conducted with 100,000 subjects enrolled in both the placebo arm and the vaccine arm. Notice that, our simulation purpose is to use observational sample with sample size 1,000 to construct the antibody disease risk curve. The large vaccine trial sample size of 100,000 is used solely to obtain accurate estimation of VEs; it does not imply that we are simulating such a large-scale clinical trial.

5.3.2 *Approaches to be compared:*

We grouped the 10 approaches in four groups as presented in Figure 5, including some of the common approaches suggested in vaccine trial literature.

The first group contains only one method, named the Gold Standard. Our underlying disease probability function $g(x,t)$ is either based on a CALM or a logistic regression model of the variables t and x. Since we assume there are no additional measurable confounders, $g(x,t)$ can be estimated unbiasedly with the same function form to define $g(x,t)$. Correspondingly, the VE obtained with applying the estimated $g(x,t)$ should represent the best achievable estimation. Consequently, we consider this approach a gold standard and compare other approaches to it, despite that the approach is usually not proposed for actual clinical trials.

The second groups contain two methods that completely ignore the covariate *X*. The first method is a simple logistic regression model $g(t) = exp(\alpha+\beta t)/(1+exp(\alpha+\beta t))$ and the second method is the scaled logit model: $g(t) = \lambda h(t)$, where $h(t)$ is the logit function $1/(1+exp(a+bt))$, truncated at t_0 and scaled to $h(t_0) = 1$.

The four methods in the third group are based on using multivariate logistic regression model to estimate $g(x,t)$, then derive covariate free functions $g(t)$ from the estimation function $g(x,t)$. The first two methods are based on some common practice in using logistic regression to construct dose-response curves as in Daniels *and others* [1], either by removing the covariate term in $g(x,t)$, or by weighted averaging the intercepts of the individual dose response curve $g(t \mid x)$. The other two methods are based on applying a g-formula as in Equation 1.1 to derive $g(t)$ from $g(x,t)$ using conditional antibody distribution $f(t \mid x)$ and marginal distribution of covariate $f(x)$ estimated from the observed sample. Among them, the first method uses the same $g(t)$ in both the placebo and the vaccine arms, while the second method only use $g(t)$ in the placebo arm, but then assuming $f(t \mid x) = f(t)$ in the vaccine arm and use the formula in Equation 4.5 to estimate VE.

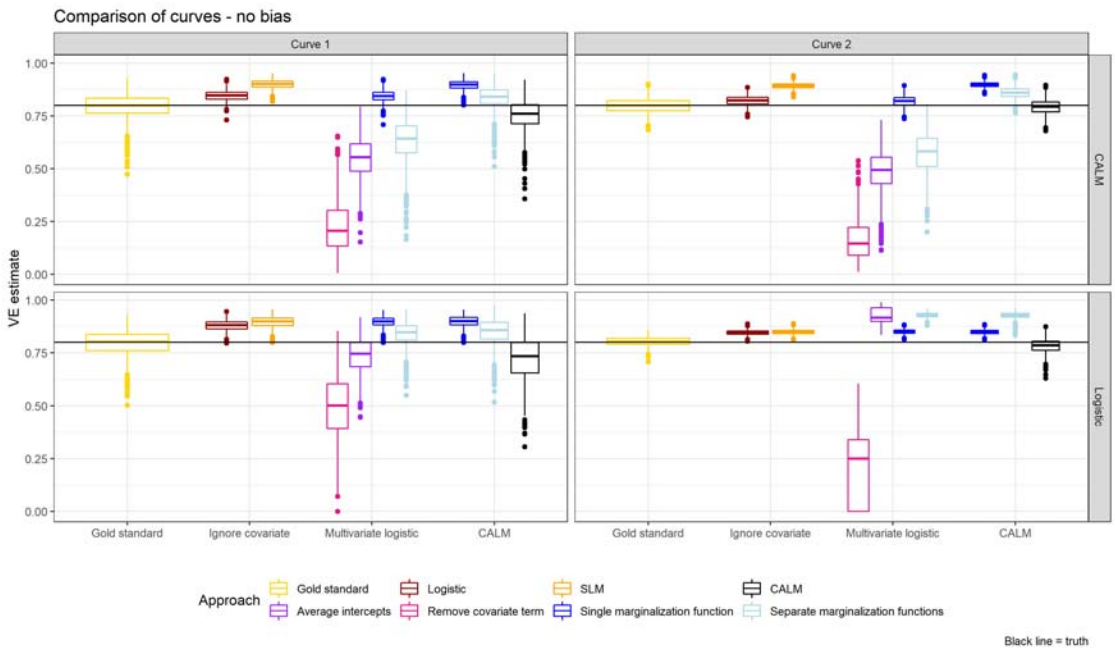
The three methods in the fourth group are based on applying CALM model to the observed sample. Although CALM has already generated a covariate free relative dose response curve $h(t)$, for comparison purpose, we also consider applying g-formula to CALM estimated function $g(x,t) = g(x)h(t)$, with the same two options: using the same marginalized $g(t)$ in both trial arms, or just use it in the placebo arm and assume $f(t \mid x) = f(t)$ in the vaccine arm.

5.3.3 *Simulation Results*

In Figure 5, we analyzed unbiased samples with $P(X = 1) = 0.25$. The figure comprises four panels corresponding to the four underlying disease probability models: two CALM models and two logistic regression model. Within the panels, each of the box plots represents the 1,000 estimated VEs based on the 1,000 simple random samples.

Figure 5 clearly demonstrates that excluding confounding covariates and relying solely on antibody levels is not advisable. Both methods in the " ignore covariates" group exhibit systematic bias in estimating VE for all four simulation scenarios.

Fig. 5. Comparing VE estimation methods with unbiased sample



The approach of deriving $g(t)$ from $\hat{g}(x,t)$ estimated by a multivariate logistic regression model is also not advisable, whether by dropping the covariate term, or averaging the intercepts of individual logistic response curve. Even when $\hat{g}(x,t)$ is the best estimation of the true $g(x,t)$ by study design under the simulation of logistic regression models, the methods exhibit drawbacks.

Similarly, marginalization approaches with g-formula are also suboptimal. Using the same marginalized $g(t)$ in both study arms should be avoided, as it systematically overestimated the true VE. The performance of using a different marginalization function in the vaccine arm is more nuanced, although theoretically justifiable if the assumption that the antibody distribution among vaccinated subjects does not depend on covariate might be reasonable. Similarly, applying g-formula to the CALM estimated function g(x,t) = g(x) h(t) is also not a good idea.

Nevertheless, in all scenarios examined, CALM's relative response curve $h(t)$ provides a more stable and less biased estimation compared to those derived from marginalized curves.

The impact of biased sample is illustrated in Figure 6, where we compare the performance of sample with biased $(P(X = 1) = 0.1)$ versus unbiased $(P(X = 1) = 0.25)$ covariate distributions. As

expected, the marginalization procedure shows susceptible to selection bias of sample. However, the use of CALM's relative response curve $h(t)$ is minimally affected by biased samples.

To further quantify the performance of different approaches, we calculated the mean square errors of each VE estimations compared to gold standard estimation. The results are summarized in Table 2, where rows represent the methods applied, and columns correspond to four simulation models, each with the sample being generated under either unbiased $P(X = 1) = 0.25$ or biased $P(X = 1) = 0.1$ covariate distributions.

Clearly, VE estimates from the CALM approach generally exhibit the smallest mean squared error from the gold standard estimation, across all simulation scenarios, regardless of bias introduced by observational samples. Therefore, overall, despite the inherent bias of using a single antibody disease curve to estimate VEs in both study arms, the CALM approach provides a straightforward yet more accurate estimation of true VE compared to other methods.

Fig. 6. Comparing VE estimation methods biased and unbiased sample

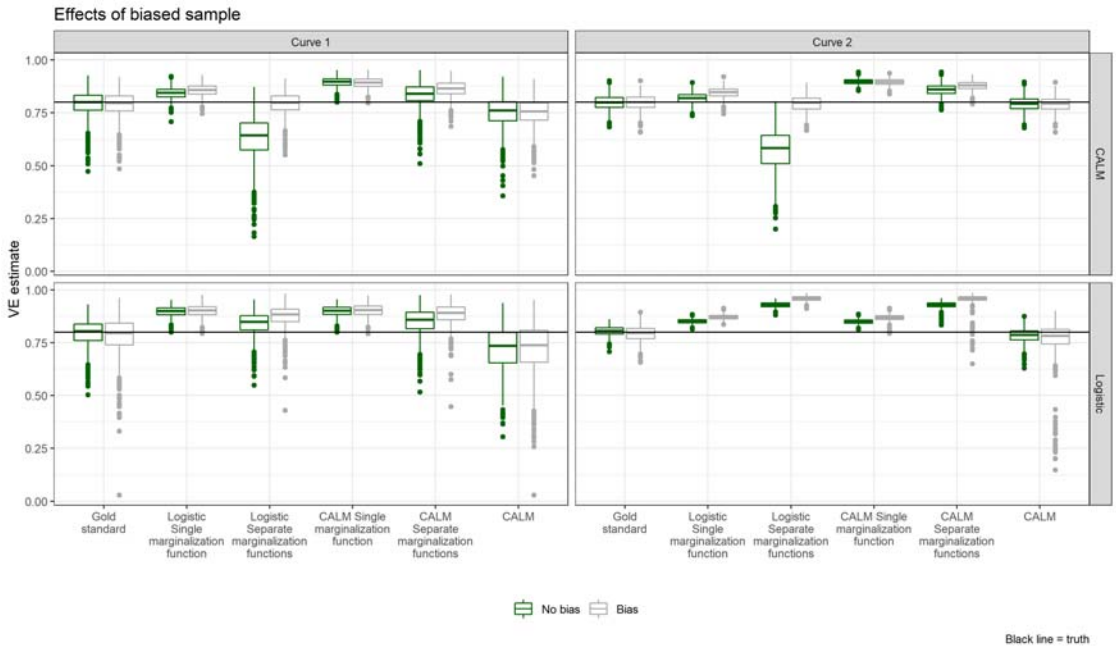


Table 2: Mean Square Errors of VE Estimations from gold standard

Methods	CALM				Logistic			
	Model 1		Model 2		Model 1		Model 2	
	unbiased	biased	unbiased	biased	unbiased	biased	unbiased	bias
Logistic, no covariate	0.005	0.01	0.002	0.005	0.009	0.019	0.002	0.0
Scaled logit model	0.014	0.013	0.01	0.009	0.013	0.02	0.002	0.0
Logistic, remove covariate term	0.331	0.199	0.409	0.192	0.103	0.099	0.394	0.3
Logistic, average intercept	0.073	0.031	0.107	0.025	0.013	0.024	0.016	0.0
Logistic, single marginalization function	0.005	0.007	0.002	0.003	0.013	0.019	0.002	0.0
Logistic, two marginalization function	0.032	0.001	0.06	0.001	0.002	0.011	0.015	0.0
CALM, single marginalization function	0.013	0.013	0.011	0.01	0.014	0.02	0.002	0.0
CALM, two marginalization functions	0.002	0.006	0.004	0.007	0.005	0.014	0.015	0.0
CALM	0.002	0.002	0	0	0.01	0.006	0.001	0.0

6. Summaries:

Establishing dose-response curves for causal relationships using observational data involves addressing two significant challenges: confounding covariates and sample selection bias. When selection bias leads to discrepancies in the marginal distribution of confounding covariates between the observational sample and the target population, traditional g-formula marginalization procedures cannot provide unbiased estimates of the dose-response curve.

A standard approach in causal analysis decomposes the absolute disease risk into relative disease risk and background disease risk through modeling with additional assumptions. For example, in binary treatment scenarios, log-linear structural mean models for causal risk ratios (Robins [10], Dukes and Vansteelandt [11]) have been extensively studied and applied in epidemiological and medical research. This paper extends this model to define a relative dose-response curve for continuous treatment. By separating the Null Disease Risk (NDR), defined as the counterfactual disease risk at the baseline treatment level, from a continuous function of causal risk ratios, we propose the Covariate Adjusted Logit Model (CALM) with a flexible definition of the relative dose-response curve.

Recent developments in generalized propensity scores for continuous treatments offer new methods for establishing causal dose-response curves (Imbens [13], Hirano and Imbens [14], Brown et al. [15], Fong et al. [16]). However, most studies focus on adjusting confounders without addressing selection bias issues. This paper adopts a standard regression modeling approach to estimate the relative dose-response curve in CALM, deferring causal analysis solutions for future research. Unlike propensity score-based approaches that treat NDR as an unspecified function, CALM models the NDR components directly. Given that observational studies often include a large proportion of subjects at the baseline dose level, our approach may be more efficient in utilizing observed data. Moreover, based on insights from vaccine effectiveness interpretation of the relative dose effect, modeling both the NDR and the dose-response curve facilitates a Gibbs sampler called the All-or-Nothing (AoN) algorithm. This algorithm provides a new perspective in adjusting for confounder effects by revealing the individual's latent susceptibility to disease rather than balancing continuous treatment levels with propensity score approaches.

Assessing vaccine effectiveness (VE) through immunologic response in randomized clinical trials is often preferred, especially for rare diseases. However, as discussed in Section 4, an absolute antibody-disease curve may not exist to provide an unbiased estimation of VE when confounding covariates influence the dose-response relationship. This challenge arises because the relationship between covariates and antibodies differs between placebo and vaccine arms. While randomization ensures similar marginal distributions of covariates, it does not align the conditional distribution of antibodies in the presence of confounding factors. As a result, the two study arms can be seen as representing different target populations when defining counterfactual dose-response relationships. This discrepancy generally precludes the establishment of a common causal dose-response curve applicable to both study arms.

While establishing a common "absolute" dose-response curve for both trial arms based on "absolute" disease risk may not be feasible, the CALM model offers the capability to derive a common relative dose-response curve applicable to both arms. CALM's modeling framework establishes a direct link between the relative response curve and a specific interpretation of vaccine effectiveness, namely in reducing disease susceptibility. Hence, this curve can be effectively utilized in vaccine trials to assess vaccine effectiveness. Our simulation studies demonstrate that CALM's response curve consistently outperforms other alternatives in scenarios

involving confounding factors and shows less susceptibility to biases introduced by selection bias of the observational samples.

Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

[1] Daniels, Robert D, Gilbert, Stephen J, Kuppusamy, Senthilkumar Perumal, Kuempel, Eileen D, ParkRobert M, Pandalai, Sudah P, Smith, Randall J, Wheeler, Matthew W, Whittaker, Christine and Schulte, Paul A. (2020). NIOSH Practices in Occupational Risk Assessment.

[2] Kricker, Anne, Armstrong, Bruce K, English, Dallas R and Heenan, Peter J. (1995). A dose-response curve for sun exposure and basal cell carcinoma. *International journal of cancer* **60**(4), 482–488.

[3] Di, Qian, Dai, Lingzhen, Wang, Yun, Zanobetti, Antonella, Choirat, Christine, Schwartz, Joel D and Dominici, Francesca. (2017). Association of short-term exposure to air pollution with mortality in older adults. *JAMA* **318**(24), 2446–2456.

[4] Wang, Thomas J, Pencina, Michael J, Booth, Sarah L, Jacques, Paul F, Ingelsson, Erik, Lanier, Katherine, Benjamin, Emelia J, D'Agostino, Ralph B, Wolf, Myles and Vasan, Ramachandran S. (2008). Vitamin d deficiency and risk of cardiovascular disease. *Circulation* **117**(4), 503–511

[5] Crowe, Francesca L, Thayakaran, Rasiah, Gittoes, Neil, Hewison, Martin, Thomas, G Neil, Scragg, Robert and Nirantharakumar, Krishnarajah. (2019). Non-linear associations of 25-hydroxyvitamin d concentrations with risk of cardiovascular disease and all-cause mortality: results from the health improvement network (thin) database. *The Journal of steroid biochemistry and molecular biology* **195**, 105480.

[6] Gilbert, Peter B, Isbrucker, Richard, Andrews, Nick, Goldblatt, David, Heath, Paul T, Izu, Alane, Madhi, Shabir A, Moulton, Lawrence, Schrag, Stephanie J, Shang, Nong *and others*. (2022). Methodology for a correlate of protection for group b streptococcus: Report from the Bill & Melinda Gates foundation workshop held on 10 and 11 February 2021. *Vaccine* **40**(32), 4283–4291.

[7] Izu, Alane, Kwatra, Gaurav, Madhi, Shabir A and Rigat, Fabio. (2022). Estimation of invasive group b streptococcus disease risk in young infants from case-control serological studies. *BMC Medical Research Methodology* **22**(1), 1–16.

[8] Madhi, Shabir A, Izu, Alane, Kwatra, Gaurav, Jones, Stephanie, Dangor, Ziyaad, Wadula, Jeanette, Moultrie, Andrew, Adam, Yasmin, Pu, Wenji, Henry, Ouzama *and others*. (2021). Association of group b streptococcus (gbs) serum serotype-specific anticapsular immunoglobulin g concentration and risk reduction for invasive gbs disease in south african infants: an observational birth-cohort, matched case-control study. *Clinical Infectious Diseases* **73**(5), e1170–e1180.

[9] Rhodes J, Bolcen S, Kahn R, McGee L, Moore A, Britton A, Nadle J, Barnes M, Hansen A, Lynfield R, Rowlands J, LeVasseur MK, Vishwanathan S, Shang N, Schrag S (2023) The U.S. CDC-led study to establish immunologic endpoints associated with protection against infant invasive group B streptococcal disease: updates and preliminary findings of interim analyses. *International Symposium for Streptococcal Agalactiae Disease (ISSAD).* Rio de Janeiro, Brazil.

[10] Robins, JM, (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Commun State Theory Methods* **23**(8), 2379–2412.

[11] Dukes and Vansteelandt (2018). A note on G-estimation of Casual Risk Ratios. *American Journal of Epidemiology* **187**(5), 1079-1084.

[12] Robins, JM, (1999). Association, causation, and marginal structure models. *Synthese* **121**, 151-179
.

[13] Imbens, Guido W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87:** 706-710.

[14] Hirano, Keisuke and Imbens, Guido W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* **226164**, 73–84.

[15] Brown, Derek, Greene, Thomas, Swartz, Michael, Wilkinson, Anna and DeSantis, Stacia. (2021). Propensity score stratification methods for continuous treatments. *Statistics in Medicine* **40**(5), 1189–1203.

[16] Fong, Christian, Hazleft, Chad and Imai, Kosuke. (2018). Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *The Annals of Applied Statistics* **12**(1), 156–177.

[17] Cox Jr, Louis Anthony. (2020). Implications of nonlinearity, confounding, and interactions for estimating exposure concentration-response functions in quantitative risk analysis. *Environmental Research* **187**, 109638.

[18] Dunning, Andrew J. (2006). A model for immunological correlates of protection. *Statistics in medicine* **25**(9), 1485–1497.

[19] Dunning, Andrew J, Kensler, Jennifer, Coudeville, Laurent and Bailleux, Fabrice. (2015). Some extensions in continuous models for immunological correlates of protection. *BMC medical research methodology* **15**(1), 1–11.

[20] Halloran, M Elizabeth, Longini, Ira M, Struchiner, Claudio J and Longini, Ira M. (2010). *Design and analysis of vaccine studies*, Volume 18. Springer.

[21] WHO. (2013). Correlates of vaccine-induced protection: methods and implications. *Technical Report*, World Health Organization.

[22] Polson, Nicholas G, Scott, James G and Windle, Jesse. (2013). Bayesian inference for logistic models using Polya–gamma latent variables. *Journal of the American statistical Association* **108**(504), 1339–1349.

[23] Choi, Hee Min and Hobert, James P. (2013). The Polya-gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics* **7**, 2054–2064.

[24] Fruhwirth-Schnatter, Sylvia, Zens, Gregor and Wagner, Helga (2020). Ultimate Polya gamma samplers–efficient mcmc for possibly imbalanced binary and categorical data. *arXiv preprint arXiv:2011.06898*.

[25] Neelon B and Dunson D (2004). Bayesian Isotonic Regression and Trend Analysis *Biometrics* **60** (2) 398-406

[26] Li W and Fu H (2017). Bayesian Isotonic Regression dose-response model *Biopharm Stat* **27** (5) 824-833

[27] FDA and CBER. (2018). Approaches for demonstrating effectiveness of group b streptococcus (gbs) vaccines intended for use in pregnant women to protect the newborn infant. 152nd meeting of the vaccines and related biological products advisory committee (vrbpac).

[28] Rajam, Gowrisankar, Carlone, George, Kim, Ellie, Choi, Jin, Paulos, Simon, Park, SoHee, Jeyachandran, Amilia, Gorantla, Yamini, Wong, Emily, Sabnis, Amit *and others*. (2019). Development and validation of a robust multiplex serological assay to quantify antibodies specific to pertussis antigens. *Biologicals* **57**, 9–20

[29] Miller, Elizbeth, Salisbury, D and Ramsay, Mary. (2001). Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup c disease in the uk: a success story. *Vaccine* **20**, S58–S67.

[30] Cherry, James D, Gornbein, Jeffrey, Heininger, Ulrich and Stehr, Klemens. (1998). A search for serologic correlates of immunity to bordetella pertussis cough illnesses. *Vaccine* **16**(20), 1901–1906.

[31] Prentice, Ross L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine* **8**(4), 431–440.

[32] Qin, Li, Gilbert, Peter B, Corey, Lawrence, McElrath, M Juliana and Self, Steven G. (2007). A framework for assessing immunological correlates of protection in vaccine trials. *The Journal of infectious diseases* **196**(9), 1304–1312

[33] Siber, George R, Chang, IH, Baker, Sherryl, Fernsten, Philip, O'Brien, Katherine L, Santosham, Mathuram, Klugman, Keith P, Madhi, Shabir A, Paradiso, Peter and Kohberger, Robert. (2007). Estimating the protective concentration of antipneumococcal capsular polysaccharide antibodies. *Vaccine* **25**(19), 3816–3826.

[34] Chen, Xuan, Bailleux, Fabrice, Desai, Kamal, Qin, Li and Dunning, Andrew J. (2013). A threshold method for immunological correlates of protection. *BMC medical research methodology* **13**(1), 1–10.

[35] Donovan, Kevin M, Hudgens, Michael G and Gilbert, Peter B. (2019). Nonparametric inference for immune response thresholds of risk in vaccine studies. *The annals of applied statistics* **13**(2), 1147.

[36] van der Laan, Lars, Zhang, Wenbo and Gilbert, Peter B. (2022) Nonparametric estimation of the causal effect of a stochastic threshold-based intervention. *Biometrics*.

Table 1. Null disease rate and antibody distributions by covariates

Covariates	strata	Percentage of cases among those with minimal antibody	p-values for percentage of cases between strata of covariates	Mean of logarithm of antibody values	p-values to compare antibody distributions
Chorioamnionitis	1	0.455	0.35	-2.59	0.14
	0	0.304		-1.54	
Geographic region	1	0.176	0.002	-1.49	0.23
	0	0.565		-2.09	
Gestational age	1	0.362	0.33	-1.48	0.02
	0	0.200		-3.22	

Figure 1 Isotonic regression curves stratified by covariates.

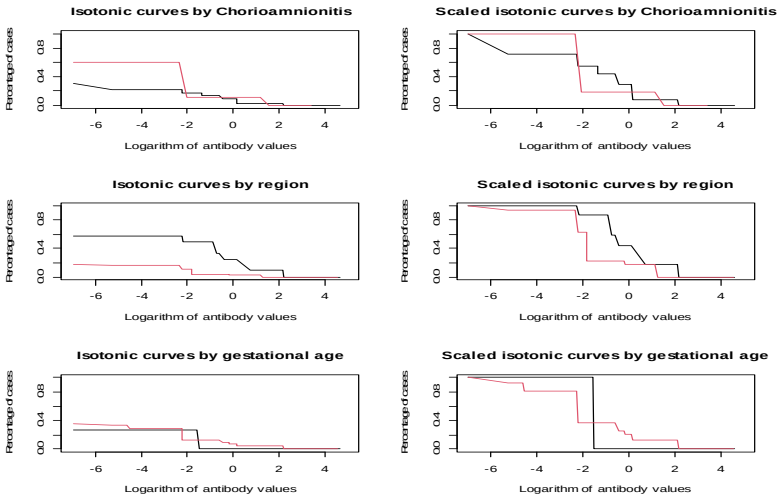


Figure 2. Dose response curves constructed by g-formula and by CALM

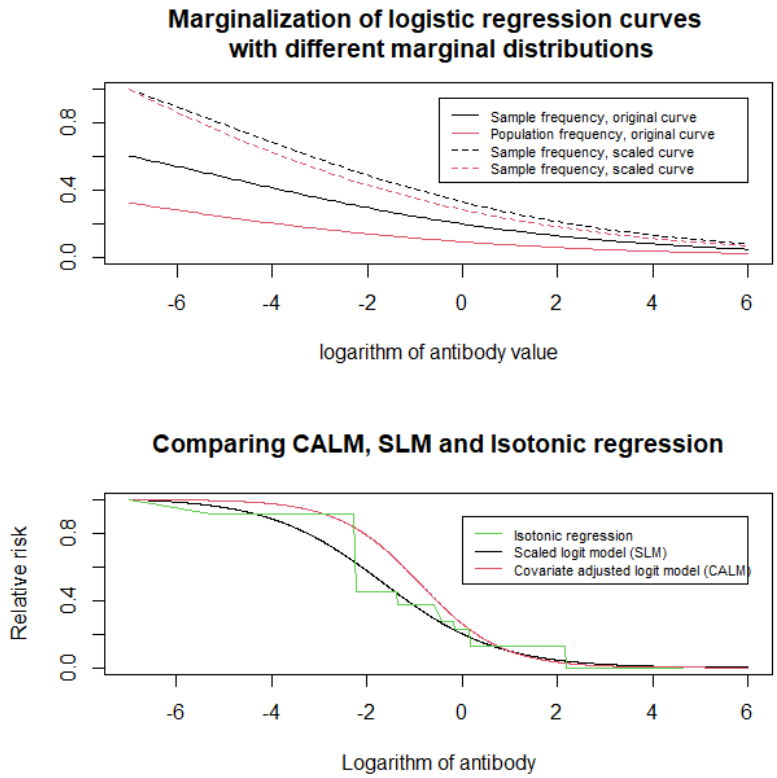


Fig. 3. Simulation designs

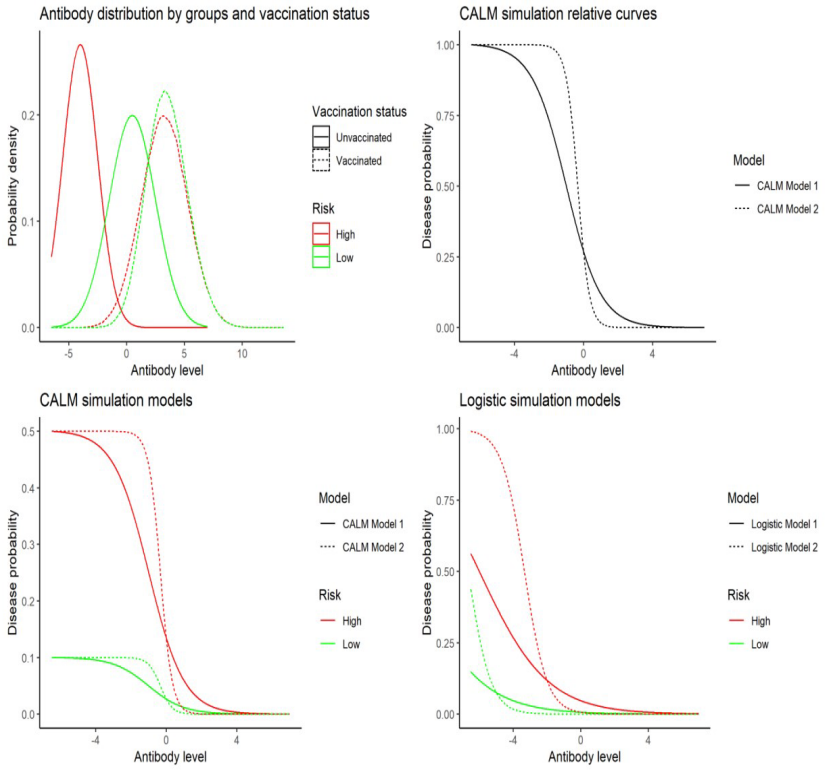


Fig. 4 . Performance of CALM in estimating relative dose response curve

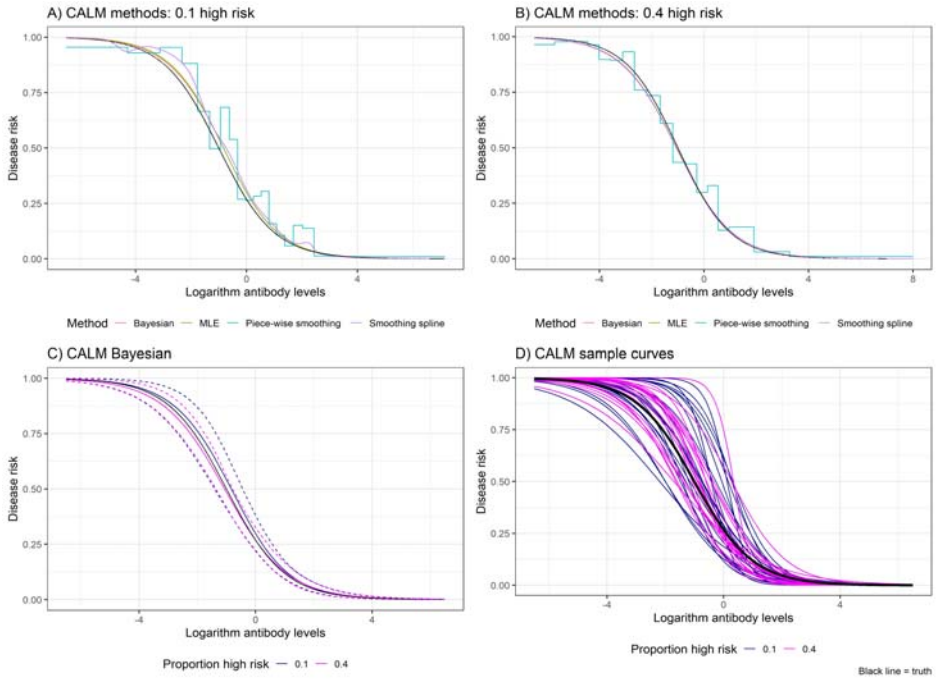


Fig. 5. Comparing VE estimation methods with unbiased sample

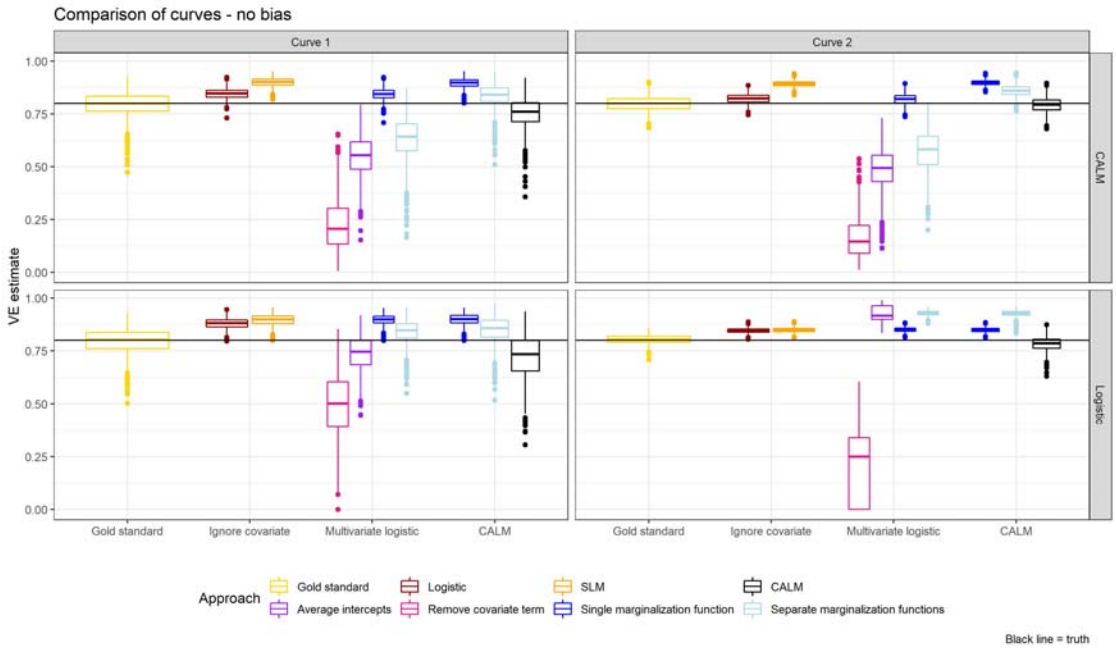


Fig. 6. Comparing VE estimation methods biased and unbiased sample

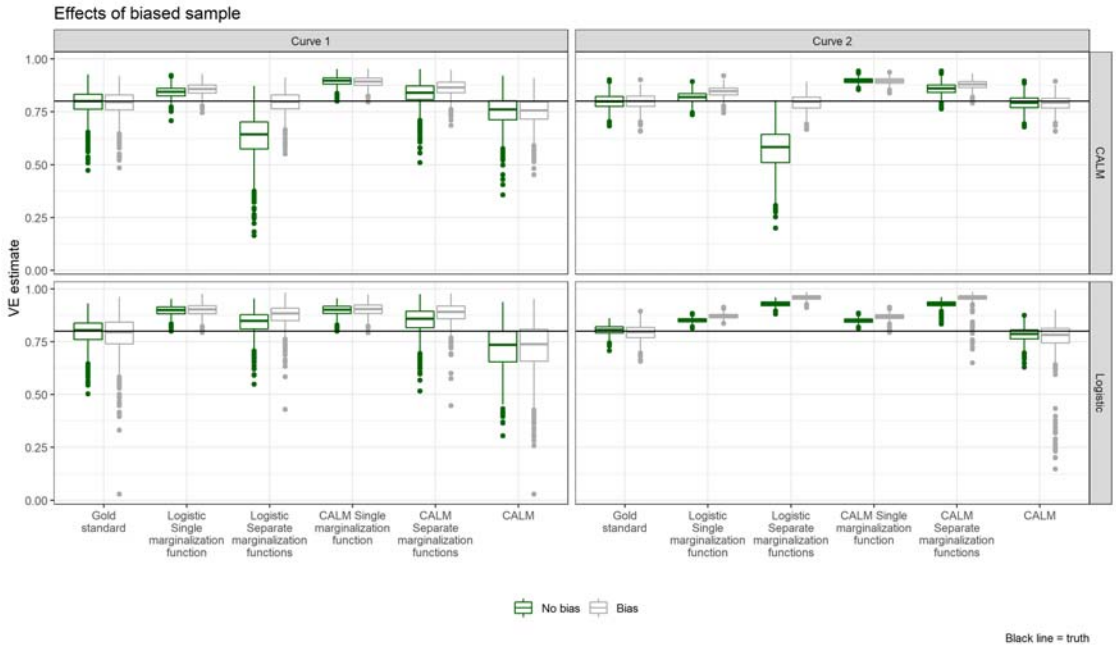


Table 2: Mean Square Errors of VE Estimations from gold standard

Methods	CALM				Logistic			
	Model 1		Model 2		Model 1		Model 2	
	unbiased	biased	unbiased	biased	unbiased	biased	unbiased	bias
Logistic, no covariate	0.005	0.01	0.002	0.005	0.009	0.019	0.002	0.0
Scaled logit model	0.014	0.013	0.01	0.009	0.013	0.02	0.002	0.0
Logistic, remove covariate term	0.331	0.199	0.409	0.192	0.103	0.099	0.394	0.3
Logistic, average intercept	0.073	0.031	0.107	0.025	0.013	0.024	0.016	0.0
Logistic, single marginalization function	0.005	0.007	0.002	0.003	0.013	0.019	0.002	0.0
Logistic, two marginalization function	0.032	0.001	0.06	0.001	0.002	0.011	0.015	0.0
CALM, single marginalization function	0.013	0.013	0.011	0.01	0.014	0.02	0.002	0.0
CALM, two marginalization functions	0.002	0.006	0.004	0.007	0.005	0.014	0.015	0.0
CALM	0.002	0.002	0	0	0.01	0.006	0.001	0.0