Statistical Methods for Standard Membrane-Feeding Assays to Measure Transmission Blocking or Reducing Activity in Malaria

Bruce J. Swihart *
and
Michael P. Fay
and
Kazutoyo Miura

February 5, 2016

Abstract

Transmission blocking vaccines for malaria are not designed to directly protect vaccinated people from malaria disease, but to reduce the probability of infecting other people by interfering with the growth of the malaria parasite in mosquitoes. Standard membrane-feeding assays compare the growth of parasites in mosquitoes from a test sample (using antibodies from a vaccinated person) compared to a control sample. There is debate about whether to estimate the transmission reducing activity (TRA) which compares the mean number of parasites between test and control samples, or transmission blocking activity (TBA) which compares the proportion of infected mosquitoes. TBA appears biologically more important since each mosquito with any parasites is potentially infective; however, TBA is less reproducible and may be an overly strict criterion for screening vaccine candidates. Through a statistical model, we show that the expected value of the sample TBA depends on μ_c, the mean number of parasites in the control mosquitoes, a parameter not easily experimentally controlled. We develop a standardized TBA estimator based on the model and a given target

*The authors gratefully acknowledge the *Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH and also the PATH Malaria Vaccine Initiative*

value for μ_c which has better mean squared error than alternative methods. We discuss types of statistical inference needed for using these assays for vaccine development[1].

Keywords: assay normalization, standardization, negative binomial, zero inflation; *Plasmodium falciparum*.

[1]Please regard everything after the references (Sections 9 and 10) as a separate online supplement excluded from manuscript page counts; included here for the purposes of reviewing ease.

1 Introduction

Malaria remains a daunting global public health problem, with the World Health Organization estimating that in 2015 there were about 214 million cases and 0.4 million deaths, with most deaths in children under 5 (WHO, 2015). Malaria is caused by infection with a *Plasmodium* parasite, and deaths are mainly due to *Plasmodium falciparum* (WHO, 2015). No highly effective malaria vaccine has been developed, although the new RTS,S/AS01 malaria vaccine given to infants and young children has shown vaccine efficacy during the 20 months after vaccination of 45% (95% CI: 41% to 49%) (RTS,S Clinical Trials Partnership, 2014). However, in 5-17 month old children, using a bi-phasic exponential model, the vaccine efficacy to prevent new infections is estimated to begin at 74%, but reduces to 28% at 1 year after vaccination (White et al., 2015). Therefore, next generation vaccines or other types of tools are needed to accelerate elimination and eradication of malaria.

Transmission blocking vaccines (TBVs) offer one such tool. Transmission blocking vaccines for malaria are not designed to confer direct and immediate protection of vaccinated people from malaria disease, but rather to reduce the probability of infecting other people by interfering with the growth of the malaria parasite in mosquitoes. Under conditions of sufficient vaccine efficacy and coverage, this approach is expected to accelerate local malaria parasite elimination. TBVs are designed to produce human antibodies that interfere with malaria parasite development inside the mosquito (WHO, 2015). Several TBVs are being developed and an efficacious TBV may be an important addition to malaria vaccines in controlling malaria epidemics and eradicating malaria (malERA Consultative Group on Vaccines et al., 2011; Nunes et al., 2014).

For a final Phase III test of the efficacy of TBVs, a cluster randomized trial would probably be ideal (Delrieu et al., 2015), but other approaches for licensure have been discussed (Nunes et al., 2014). For the early stages of TBV development it is useful to measure the transmission blocking efficacy for each vaccinated individual. Additionally, for models of malaria transmission, having estimates for the probability that a mosquito becomes infected during a single feed-day for both a

vaccinated individual and an unvaccinated individual is useful (Smith et al., 2011). Such estimates are possible via assays. There are many types of assays that have been developed to measure mosquito transmission of *Plasmodium* parasites (Sinden et al., 2012).

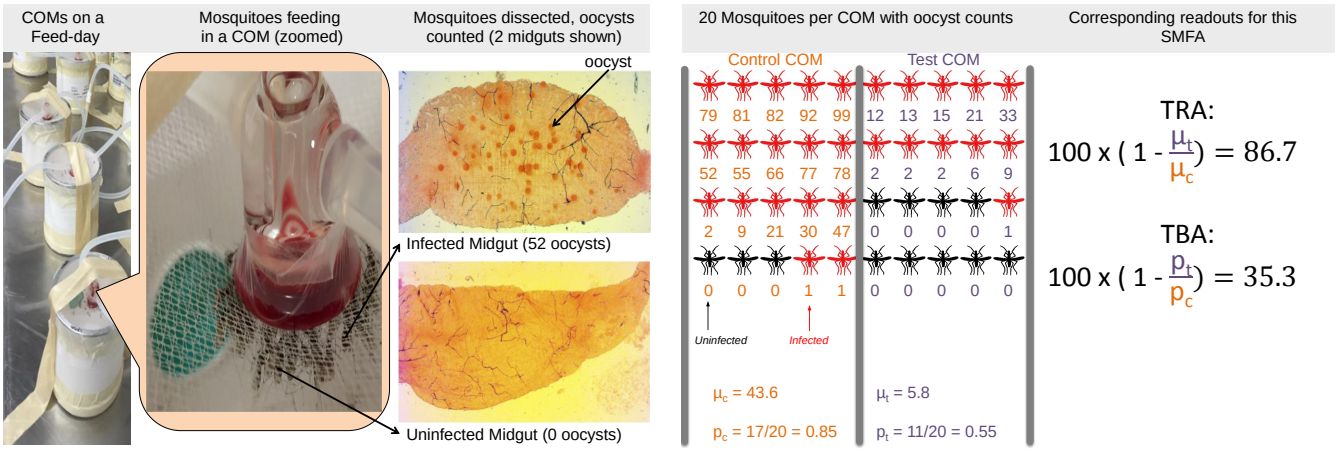

Figure 1: SMFA and the two readouts: TRA and TBA. From left to right: Several containers of mosquitoes (COMs) are fed on the same feed-day and are maintained for a week. Then each mosquito is dissected and oocysts in the midgut counted. Mosquitoes with 0 oocysts are uninfected, those with 1 or greater are infected. Within each test and control COM, 20 mosquitoes are summarized by their oocyst counts mean (μ) and prevalence (p) (e.g., proportion infected). Those summary measures then inform the calculations of TRA and TBA.

The direct skin-feeding assay (DSFA, mosquitoes fed malaria infected blood through skins of malaria infected subjects) and direct *membrane*-feeding assay (DMFA, mosquitoes fed mixture of malaria infected blood collected from human and control/test antibody through a feeding apparatus) need to be conducted in a field site where malaria is endemic (transporting gametocytes – parasites in the transmissible stage – destroys their viability). Unfortunately these procedures have high failure rates in producing infective mosquitoes for study (Bousema et al., 2012). Comparatively, the standard membrane-feeding assay (SMFA) has advantages over DSFA and DMFA.

SMFA is conducted in a lab and mixes a large enough number of lab raised parasites into the blood to make it infective to mosquitoes with very high probability. Another advantage of SMFA over DSFA is that SMFA contains a control sample which is used to adjust for some of the day-to-day biological variability of the assay. In SMFAs mosquitoes feed on one of two samples: the test sample or the control sample. The test sample contains blood mixed with the lab-raised gametocytes and with test antibodies (e.g., antibodies produced by a vaccinated individual), while the control sample is similar except using control antibodies (e.g., antibodies from individuals that have not been exposed to malaria or the vaccine). The two samples are fed to mosquitoes and the resultant yield of parasites (in the *oocyst* stage) growing in each mosquito is enumerated (see Section 2.2 for more details). Bousema et al. (2013) argue that "the SMFA is rightfully the current gold standard for ranking TRI [transmission-reducing intervention] candidates and prioritizing the most promising TRI candidates for further development". Due to SMFA's importance to the development of TBVs, this paper focuses solely on the SMFA.

Despite its status as the gold standard assay, the preferred readout used from the SMFA is not clear. Two common assay results are the transmission reducing activity (TRA, also called the percent inhibition of mean oocyst intensity) and the transmission blocking activity (TBA, also called percent inhibition of prevalence). We write TRA for a single individual as

$$r = 100\left(1 - \frac{\hat{\mu}_t}{\hat{\mu}_c}\right), \tag{1}$$

and TBA as

$$b = 100\left(1 - \frac{\hat{p}_t}{\hat{p}_c}\right) \tag{2}$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the mean number of oocysts per mosquito from the test sample and from the control sample, respectively, and $\hat{p}_t$ and $\hat{p}_c$ are the proportion of infected mosquitoes (that is, the proportion of mosquitoes with at least 1 oocyst) from the two samples (Figure 1). Some argue that TBA is biologically the more important measure (because in the field if a mosquito has a single oocyst, it has the potential to become infective).

We propose that TRA should be the preferred readout. While TBA might be the biologically preferred measure, its dependence on the control sample is problematic. Empirically for replicate SMFAs for a given test, the ratio of means in TRA stays roughly constant for different $\hat{\mu}_c$, whereas the ratio of proportions in TBA does not, indicating that TBA is more dependent on the control sample than TRA (e.g., see Figure 4 in Churcher et al., 2012). This implies that if a control mean in the field is different than that of the lab, the TBA of the same test will differ between the two settings, whereas the TRA will be consistent. Typically $\hat{\mu}_c$ is larger in SMFAs than in the field. For TBA to be a useful readout from SMFA, there is a need to be able to bridge the readout from these lab-based SMFAs to expected results from field-based assays such as the DSFA or DMFA (Nunes et al., 2014). When TBA is desired, we suggest a standardized TBA calculated from the TRA and a specified target control mean (Section 5.2).

The pipeline of TBV development starts from preclinical animal studies and then involves multi-stage human clinical trials (phase I, II, and III). Early on, SMFAs play a crucial role. An SMFA can be used on a single subject at one time to determine how antibodies in her blood would affect malaria transmission. Either TRA or TBA could be the readout, although for the latter some standardization is generally needed. For the single subject, confidence intervals to convey the uncertainty are needed. For studies with multiple samples per arm, effective ways to combine information are needed. For example, we may want to test individual efficacy in a randomized trial of a control vaccine vs. a TBV using some assay such as the SMFA. In Phase III development, a cluster randomized vaccine trial could be used in which case no SMFA would be needed for the primary endpoint; however, other approaches for this phase have been discussed which use individual assays (Nunes et al., 2014). The ultimate question of "How will vaccine roll-out affect future malaria rates" may require TBA estimates that are standardized for the field, for which our measures developed for SMFAs will be useful.

In this paper, we analyze an extensive data set for the SMFA. We use this data set to show that a reasonable model for the number of oocysts in each mosquito in the SMFA is a zero inflated,

negative binomial model with random effects specified for each container of mosquitoes and each feed-day that the assay is performed. Others have used this or similar models (Churcher et al., 2012; Miura et al., 2013), and have discussed the relationship between TRA and TBA estimates, r and b. The contribution of this paper is to more formally describe the estimands related to the TRA and TBA, develop a standardized TBA estimator that allows for standardizing the lab-based results to a TBA estimand that is interpretable in a malaria endemic field, and then carry out traditional statistical inferences using the estimators.

Using this model, the estimand for TRA, Ω_R, does not depend on μ_c, the true mean oocyst count for a mosquito in the control group. In contrast, the general estimand associated with TBA, $\Omega_B(\mu_c)$ depends on μ_c. These features of the estimands reflect empirical results and clarify that if we report b for TBA without reporting $\hat{\mu}_c$, we do not know which estimand within the class $\{\Omega_B(\mu_c)\}$ we are estimating. The important TBA estimand within the class is $\Omega_B(\mu_c^*)$, where μ_c^* is a fixed target mean oocyst count in the field, as opposed to $\Omega_B(\mu_c^\dagger)$, where $\mu_c^\dagger$ the mean oocyst count in the lab (which may vary substantially from feed-day to feed-day). Thus, when evaluating the effect of a vaccine candidate versus placebo, the size of the vaccine effect could be very different if each subject's TBA response is estimating the $\Omega_B(\mu_c^\dagger)$ for that subject rather than $\Omega_B(\mu_c^*)$.

We discuss two estimators for $\Omega_B(\mu_c^*)$. A restricted TBA estimator, $\hat{\Omega}_{B1}$, measures many control samples and only uses those with $\hat{\mu}_c \approx \mu_c^*$ in the calculation of b. A standardized TBA estimator, $\hat{\Omega}_{B2}(\mu_c^*)$, uses the statistical model so we can use all control samples, then input the target control mean value, μ_c^*, into that estimator. Through simulations, we show that the standardized estimator is much more efficient to use than the restricted estimator. The standardized estimator provides one method to predict the vaccine efficacy in a field from the results of a SMFA.

Here is an outline of the paper, written to address the needs of SMFA-use in the TBV development pipeline. In Section 2 we give more details on the biological issues of the problem. In Section 3, we describe the assay development data set which had many replicated assays for polyclonal control samples and motivate a representative count model (RCM) of the data generating

process on these controls. We then examine the fit of the RCM on SMFA test sample data of monoclonal and polyclonal antibodies. In Section 4, we define the estimand and estimator for TRA and the confidence intervals for a single subject. Analogously for TBA, in Section 5 we define the estimand, and propose two estimators, each with a different way to handle the dependence on the control mean. We compare the two estimators as well as provide confidence intervals for single subject TBA. In Section 6 we show how results from single subject TRAs can be combined. For example, in a randomized trial of an experimental transmission blocking vaccine versus a placebo vaccine, we could compare the two vaccine groups. We end with a short section on designing assays for vaccine studies (Section 7) and a discussion (Section 8).

2 Background

2.1 *Plasmodium falciparum*

Plasmodium falciparum has a complicated life cycle, with part of its development occurring within humans and part within mosquitoes. We start the description of the cycle when a mosquito takes a blood meal that includes red blood cells that contain the sexual stage of the parasite (the gametocytes). The parasites mature in the mosquito and develop into oocyst stage parasites in approximately 1 week's time. Each oocyst internally creates many sporozoites and eventually releases them, which travel to the salivary gland of the mosquito. If the mosquito bites a human after that, the sporozoites invade the liver cells. Within the liver cell the sporozoites change and multiply into many merozoites, which are released when the liver cell ruptures and invade red blood cells. In this blood-stage of the parasite, there can be many cycles where the red blood cells are invaded, the parasites multiply, and the red blood cells rupture to release more merozoites to repeat the cycle. This blood-stage parasite is responsible for all malaria symptoms, such as fevers, headaches, nausea, vomiting, fatigue, and other severe outcomes – including death. Some blood-stage parasites differentiate into gametocytes and may be taken up by a mosquito and complete

the life cycle.

2.2 Standard Membrane-Feeding Assay

There is no globally accepted standard reagents or protocols for the SMFA. Here we give a brief review of the way the SMFA was done for the data in this paper. To start, the NF54 strain of *Plasmodium falciparum* is maintained in the lab for 14-16 days to induce mature, infectious, gametocytes. For each day the feeding assay is performed, the cultures are mixed with normal human serum (the liquid part of blood without blood cells or clotting factors) and normal human red blood cells in predefined proportions (0.15-0.2% gametocytes and 50% red blood cells by volume). The two types of samples are created by combining that gametocyte mixture with the test antibodies or control antibodies in predefined amounts. The concentration of antibodies in the samples will differ depending on the test sample. The test samples are classified into two types: polyclonal antibody or monoclonal antibody. The polyclonal antibodies are produced *in vivo*, from either animals or humans, depending on the study. The monoclonal antibodies are produced from a hybrid of an antibody producing white blood cell (a B cell) that produces the antibody of interest and a cancerous white blood cell (myeloma). Then either the test sample or a control sample is put in a feeding apparatus that is covered with a membrane which is exposed to about 50 mosquitoes (*Anopheles stephensi*) of approximately the same age (3-7 days old). After the mosquitoes feed on a blood meal (a mixture of gametocyte culture and test (or control) antibody) through the membrane, the mosquitoes are maintained for 8 days. We call a group of mosquitoes which feed on the same blood meal a container of mosquitoes (COM). Then a subset of the mosquitoes that have taken blood meals (typically $n = 20$ per sample) are dissected *manually* to count the number of oocysts in each mosquito (see Figure 1).

As is clear from the description, many aspects of the SMFA have been standardized. Despite this standardization, there remains considerable variability in the assay. Some of this is due to the biological variability and difficulties in raising highly similar gametocytes and mosquitoes. Al-

though they are raised in a temperature-controlled environment, with highly regimented protocols, some biological variability cannot be easily eliminated. Moreover, this assay is very labor intensive and takes a long time to complete (about 3 weeks). Thus, repeating the assay on two different aliquots from the same sample would almost double the already large labor burden of the assay. Furthermore, in some cases there is not enough antibody to repeat the assay on several aliquots. Since the within-sample variability cannot be reduced to a negligible part of the between sample variability, it is important to have a good model of assay variability, and to report that variability (e.g., through confidence intervals for each assay readout).

3 Fitting a Model to Data

3.1 Description of Data

The ultimate goal is to develop methodology and inference for the TRA and TBA, but to do so we first need a reasonable model of the data-generating process, which we term the representative count model (RCM). We develop the RCM on data from mosquitoes in the control COMs using control antibodies. Later we will show that the model and the specific parameter estimates from it will fit test COMs.

One oocyst count is from the observation-level mosquito; mosquitoes are clustered into containers of mosquitoes (COMs); and COMs are clustered into feed-days. For this model-building data we have 9804 mosquitoes, in 492 COMs on 105 feed-days, translating to about 20 mosquitoes per COM and about 5 COMs per feed-day. For 9804 mosquitoes, the oocyst counts have a mean of 18.66 and a variance of 522.3, which is 28 times larger than the mean, so the data are clearly overdispersed compared to a simple independent Poisson model. The distribution of oocyst counts shows a pronounced proportion of zeroes and a skewing to the right, both extremes contributing to the overdispersion (Figure 2).

The data exhibit a hierarchical structure, which could explain some of that overdispersion. To

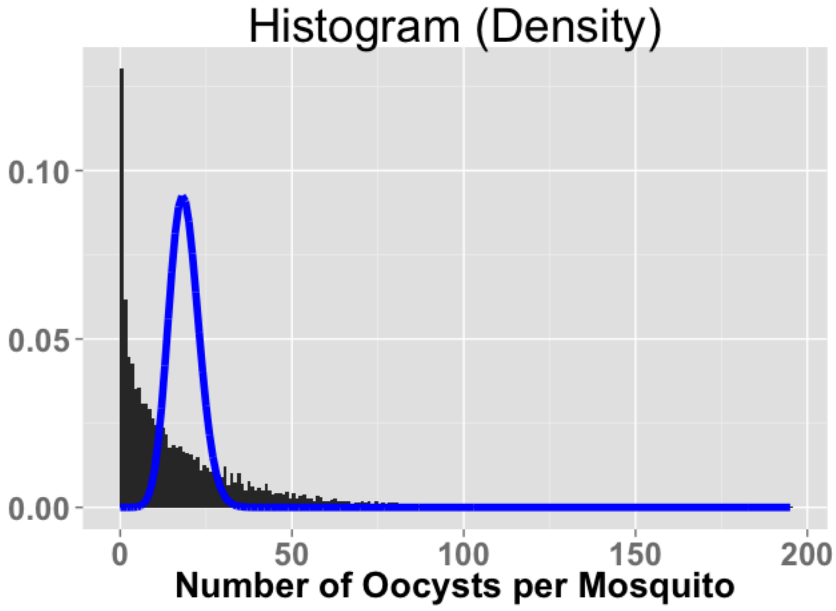

Figure 2: Histogram of 9804 oocyst counts from polyclonal control COMs with Poisson($\lambda = 18.66$) density overlay. Note the excessive proportion of zeroes and skewing to the right, both extremes contributing to overdispersion.

get a sense of the variation across feed-days, we calculate each COM's mean oocyst count over the 20 mosquitoes. Then within feed-day, we calculate the average of the COM-wise mean oocysts along with t-distribution confidence intervals on that average. (Figure 3; note: some days have only 1 COM, others 16). For this case study, the feed-day effect is substantial.

The relation between the sample COM-wise means and standard deviations (or variances) are of interest to inform model choice. We plot the within-COM standard deviation versus the within-COM mean (both on the log scale) against the equality line for the variance and mean (Figure 4).

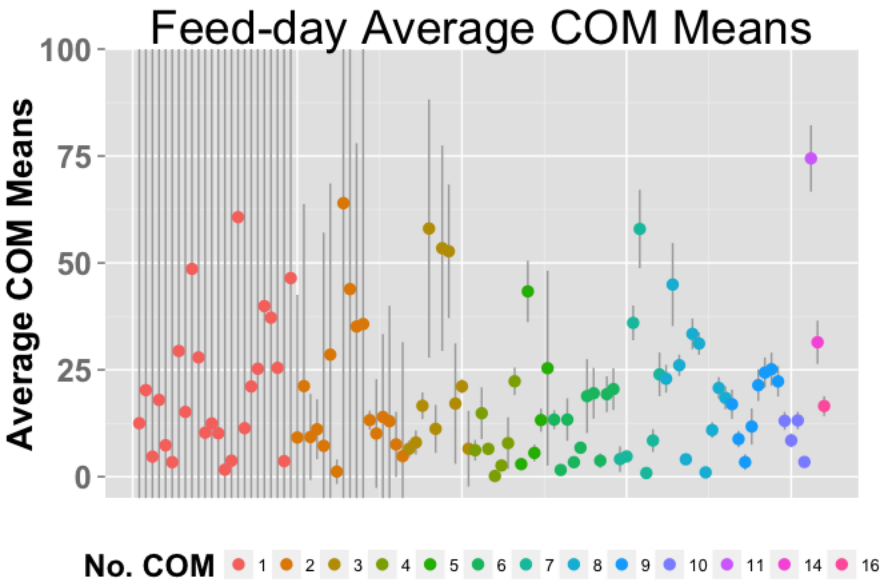

Figure 3: The oocysts found in 20 mosquitoes of each COM are averaged (we call this the "COM-wise mean"), and then the average of the COM-wise means is taken within each feed-day. 25 feed-days had 1 COM (left-most dots), 1 feed-day had 16 COMs (right-most dot). One feed-day had 11 COMs, and its average COM-wise mean was about 75. Each average COM-wise mean oocyst is presented with t-distribution 95% confidence intervals (feed-days with 1 COM set to ∞).

3.2 A Representative Count Model

To develop methods and inference, a statistical model of the data generating process will be justified. We propose a zero-inflated negative binomial random effects model, denoted as ZINB2-RE. Let Y_{ijk} be the oocyst count of the kth mosquito in the jth COM of the ith feed-day experiment,

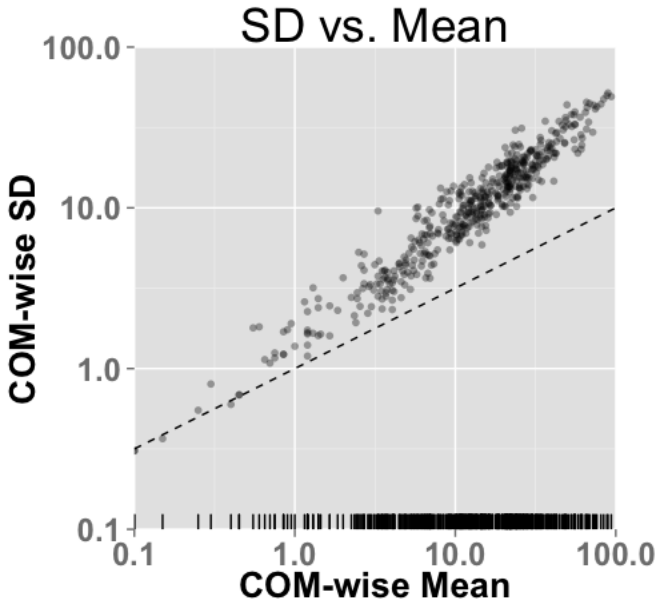

Figure 4: The COM-wise standard deviation vs. the COM-wise mean on the log scale for polyclonal controls and the mean-variance equality line (dashed). If these data followed the Poisson model, we would expect the dots to fall on the dashed line. From the rug plot and alpha blending, we see a majority of the data have a COM-wise mean greater than 10.

where each mosquito's likelihood contribution is:

$$L_{ijk}\left(Y_{ijk}=y_{ijk}|\psi\right)=\begin{cases}\pi+(1-\pi)\left(1+\frac{\lambda_{ij}}{\theta}\right)^{-\theta}, & y_{ijk}=0\\ (1-\pi)\,\frac{\Gamma\left(y_{ijk}+\theta\right)}{\Gamma\left(y_{ijk}+1\right)\Gamma(\theta)}\,\frac{\left(\frac{\lambda_{ij}}{\theta}\right)^{y_{ijk}}}{\left(1+\frac{\lambda_{ij}}{\theta}\right)^{y_{ijk}+\theta}}, & y_{ijk}=1,2,\ldots\end{cases}$$

The mean oocysts of the COM is modeled as $(1-\pi)\lambda_{ij}$, where π is the zero inflation parameter, and λ_{ij} is the random mean effect from the negative binomial portion of the model. Let $\log(\lambda_{ij})=\gamma+f_i+c_{ij}$, where γ is the overall (log) control mean oocysts, f_i is the random effect for the ith feed-day, and c_{ij} is the random effect for the ijth COM. We assume the random effects are all independent and normally distributed with means 0 and standard deviations either σ_f (for feed-day random effects) or σ_c (for COM random effects). Further there is a negative binomial dispersion parameter,

θ. The model is completely defined by the five parameters, $\psi = [\gamma, \pi, \sigma_c, \sigma_f, \theta]$. Such a model on the 9804 counts rendered $\hat{\psi}^{RCM} = [\hat{\gamma} = 2.57, \hat{\pi} = 0.056, \hat{\sigma}_c = 0.2306, \hat{\sigma}_f = 1.042, \hat{\theta} = 1.93]$. The necessity of zero-inflation, negative binomial (as opposed to Poisson), or random-effects can be tested via likelihood ratio tests (LRTs), due to the ZINB2-RE nesting simpler models. Each of the LRTs for zero-inflation ($\pi = 0$), negative binomial ($\theta = \infty$), and/or random-effects ($\sigma_f = \sigma_c = 0$) involve a parameter on the boundary of its domain, so the likelihood ratio test will be conservative (Self and Liang, 1987). Even so, the choice ZINB2-RE is resoundingly supported by the data (results in Supplementary Materials). Its fit to the 9804 control mosquitoes is deemed the RCM.

3.3 Validating the Representative Count Model

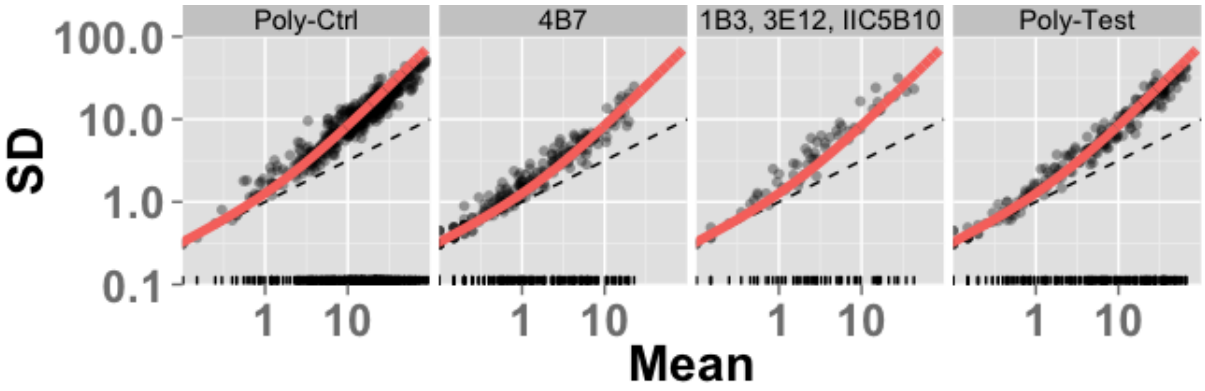

Figure 5: Standard deviation of oocyst counts (log10-scale) by mean oocyst counts (log10-scale). For each panel, the solid line is the predicted (log10-scale) standard deviation from the ZINB2-RE RCM as a function of the mean. The dashed line is the relation for a Poisson distributed oocyst count, where mean=variance. The points $(\bar{Y}, s)$, vary across panels, and are the sample mean and standard deviation of the oocyst counts from 20 mosquitoes in a COM with R^2 values of (0.91, 0.94, 0.92, 0.98) from left to right. Poly-Ctrl are the polyclonal controls upon which the RCM is built. The remaining three panels are the test COMs of monoclonal 4B7 antibodies; test COMs of monoclonal 1B3, 3E12, or IIC5B10 antibodies; and test COMs of polyclonal antibodies (Poly-Test).

The RCM model was developed on control data and we assess its generalizability to test COMs data. Specifically, we check how the mean-variance relationship of the RCM holds for the test COMs of different antibodies. For many of the feed-days for which we have polyclonal IgG controls, we also have test COMs. We group these test COMs into three distinct datasets: COMs of monoclonal 4B7 antibodies, COMs of monoclonal 1B3, 3E12, or IIC5B10 antibodies, and test COMs of polyclonal antibodies. Test COMs of monoclonal 4B7 antibodies consist of 4295 mosquitoes among 215 COMs across 104 feed-days tested at various doses. Test COMs of monoclonal 1B3, 3E12, or IIC5B10 antibodies are 1240 mosquitoes among 62 COMs across 18 feed-days tested at various doses. Test COMs of polyclonal antibodies are 5255 mosquitoes among 264 COMs across 33 feed-days.

Within each of the three test datasets, the observed log transformed standard deviation of the oocyst counts in a COM is compared to the RCM-predicted log transformed standard deviation at each observed log transformed mean oocyst counts in a COM to calculate a R^2 value. That is, the predicted RCM variance is a function of the COM-wise mean:

$$V_{ij}^{RCM}(m; \hat{\psi}^{RCM}) = m + (\frac{\hat{\pi}}{1-\hat{\pi}} + \frac{1}{\hat{\theta}}\frac{1}{1-\hat{\pi}})m^2.$$

For an observed COM-wise mean, $\bar{Y}_{ij} = \sum_{k=1}^{K_{ij}} Y_{ijk}/K_{ij}$, we calculate $\sigma_{ij}^{RCM} = \sqrt{V_{ij}^{RCM}(\bar{Y}_{ij}; \hat{\psi}^{RCM})}$ and compare it to the observed COM-wise standard deviation $s_{ij} = \sum_{k=1}^{K_{ij}}(Y_{ijk} - \bar{Y}_{ij})^2/(K_{ij}-1)$ on the log10 scale,

$$R^2 = 1 - \frac{\sum(\log_{10}\sigma_{ij}^{RCM} - \log_{10} s_{ij})^2}{\sum(\log_{10}\sigma_{ij}^{RCM} - \frac{1}{T}\sum_{i,j}\log_{10}\sigma_{ij}^{RCM})^2},$$

where T is the total number of COMs in the dataset.

These calculations lend themselves nicely to visual checks (Figure 5, with R^2 values listed in caption). The high R^2 and mean-variance line tracing centered on each cloud of data indicates a trustworthy representation of the mean-variance relation and the data generating process, for both control and test data of different antibody types.

3.4 The RCM with covariates

In practice there are test samples and control samples on the same feed-day. A covariate can be included in the mean of the RCM so that an effect for each test sample can be estimated. To accommodate test samples, let $\log(\lambda_{ij}) = \gamma + \beta_h I(X_{ij} = 1) + f_i + c_{ij}$, where β_h is the effect of the hth test sample (which is measured on the ijth COM), $X_{ij} = 1$ if the ijth COM (i.e., the jth COM of the ith feed-day) is a test sample and $X_{ij} = 0$ if the ijth COM is a control sample. With this update, the model is completely defined by the six parameters, $\psi^X = [\gamma, \pi, \sigma_c, \sigma_f, \theta, \beta_h]$ and the likelihood contribution, $L_{ijk}\left(Y_{ijk} = y_{ijk} | X_{ijk}, \psi^X\right)$, now depends on X_{ij}.

For the development of TRA and TBA, we assume any and all variation follows that of the controls as estimated by the RCM and that all of the information of the test sample enters the model through β_h via the λ_{ij} parameter (this assumption seems reasonable given the generalizability of the mean-variance relationship of the RCM to the test COMs in Section 3.3). We assume that the test sample does not affect the zero inflation parameter, π, so that the zero inflation represents some process independent of the vaccine effects that somehow stops all oocyst development for both test and control samples. That is, the RCM is analogous to fitting generalized linear models: once the family is selected, shifts around the linear predictor (in this case $\log(\lambda_{ij})$) are analyzed.

As we will show, nonparametric estimates of β_h are straightforward.

4 TRA

4.1 Defining the TRA Estimand

For all of Section 4, we consider only one test sample so we drop the h subscript in β_h. From the RCM, the TRA Ω_R directly corresponds to the fixed effect $\rho = \exp(\beta)$:

$$\begin{aligned}
\Omega_R &= 100\left(1 - \frac{E(Y_{ijk}|X_{ij}=1)}{E(Y_{ij'k'}|X_{ij'}=0)}\right) \\
&= 100\left(1 - \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(1-\pi)\exp(\gamma+\beta+s+t)\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)\frac{1}{\sigma_f}\phi\left(\frac{s}{\sigma_f}\right)ds\ dt}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(1-\pi)\exp(\gamma+s+t)\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)\frac{1}{\sigma_f}\phi\left(\frac{s}{\sigma_f}\right)ds\ dt}\right) \\
&= 100(1-\rho) \\
&=: \Omega_R(\rho)
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density. Note that a feed-day-specific TRA gives the same result:

$$\Omega_{Ri} = 100\left(1 - \frac{E(Y_{ijk}|X_{ij}=1, \text{feed-day}=i)}{E(Y_{ij'k'}|X_{ij'}=0, \text{feed-day}=i)}\right) = 100(1-\rho) = \Omega_R(\rho).$$

This allows TRA to be estimated across feed-days or within feed-days, depending on the application.

4.2 Estimating TRA

The TRA estimator of the model's TRA estimand $\Omega_R(\rho)$ only requires an estimate of the ratio of means, ρ. Thus we use $\hat{\rho} = \hat{\mu}_c/\hat{\mu}_t$ making $\hat{\Omega}_R = r$ (from the Introduction). This simple estimate uses no data from other samples, ensuring the independence of TRA values between different samples. We can consider the case of sample replicates. Suppose we have k replicates of a sample (each with its own control replicate) over many feed-days. Let $\hat{\mu}_{ic}$ and $\hat{\mu}_{it}$ be the control and treatment mean for the ith replicate, and let $\hat{\rho}_i = \hat{\mu}_{it}/\hat{\mu}_{ic}$. Then two simple estimates of ρ are either a mean of ratios ($\frac{1}{k}\sum\hat{\rho}_i$) or a ratio of means ($\bar{\mu}_t/\bar{\mu}_c$, where $\bar{\mu}_c = \frac{1}{k}\sum\hat{\mu}_{ic}$ and $\bar{\mu}_t = \frac{1}{k}\sum\hat{\mu}_{it}$). We simulated two replicates with $n_t = n_c = 20$ mosquitoes per COM using the RCM with $\hat{\psi}^{RCM}$

and Ω_R values of 10%, 50% and 90%, and 10^5 simulations. In practice, feed-days with $\hat{\mu}_{ic} = 0$ would not be used, but for the simulation we replace those values with $0.5/n_c$. Using the mean squared error (MSE) on ρ, we find that neither is uniformly better (simulated MSE for the ratio of means over simulated MSE for the mean of ratios is 1.03 [$\Omega_R = 10\%$], 1.02 [$\Omega_R = 50\%$], and 0.89 [$\Omega_R = 90\%$]).

4.3 Confidence Intervals on a Single Sample TRA

In practice we only run one feed-day and observe the oocyst counts from n_t mosquitoes from one COM using the test sample and n_c from a different COM using the control sample. In this case, we cannot estimate the standard deviation of the random effects for feed-day, σ_f, since there is only one feed-day observed. Further, we cannot estimate the standard deviation of the COMs, σ_c, since there are only two COMs sampled, and they have different samples.

We consider 6 methods for calculating confidence intervals on the TRA. The first 4 methods assume that there is no variability in the COMs (implying that $\sigma_c = 0$). The fifth and sixth methods assume that the data follow the RCM with $\sigma_c > 0$ and uses a simulation method to calculate the confidence intervals. We undertake a simulation study to evaluate coverage probabilities under different scenarios for the six methods.

For the first three methods, we start by taking the log transformation of the counts, replacing zero counts with 0.5. Then on the transformed responses, we either do a t-test with the associated confidence intervals (denoted t-Welch, using Welch's version of the t-test which allows different variances for the two groups), or the Wilcoxon-Mann-Whitney (WMW) test with the Hodges-Lehmann CIs. These are the default confidence intervals for `t.test` and `wilcox.test` in R. Then we transform the difference in log means (for t-test) or the shift in distribution (for WMW test) to the TRA scale using

$$\widetilde{\Omega}_R \;=\; 100\left(1 - \exp(\widetilde{\beta})\right)$$

where $\widetilde{\beta}$ is the estimated difference in means or shift. A third method is a delta method approach. The third method considers the variance of $\log(\hat{\rho})$ as approximated by the delta method:

$$\widehat{Var}(\log(\hat{\rho})) = \left(\frac{1}{n_t\hat{\mu}_t} + \frac{1}{n_c\hat{\mu}_c}\right) + \frac{1+\hat{\pi}\hat{\theta}}{(1-\hat{\pi})\hat{\theta}}\left(\frac{1}{n_t} + \frac{1}{n_c}\right),$$

constructing the 95% CI for $\log(\rho)$ by $\log(\hat{\rho}) \pm 1.96\sqrt{\widehat{Var}(\log(\hat{\rho}))}$ using $\hat{\rho} = \hat{\mu}_t/\hat{\mu}_c$ with $\hat{\pi}$ and $\hat{\theta}$ from the RCM. The 95% CI for $\log(\rho)$ are converted to the TRA scale. The fourth method uses a zero-inflated negative binomial model on the untransformed counts using the normal theory confidence intervals (ZINB-Norm). We used the R package pscl (Zeileis et al., 2008; Jackman, 2015). The fifth method (ZINBRE-Sim1) uses a simulation method that does not assume $\sigma_c = 0$. This method involves simulated data with the means estimated from the data while using the RCM parameter values $\hat{\psi}^{RCM}$. We specifically estimate the means for the fixed effects from the negative binomial part of the model (e.g., for $\exp(\gamma + \beta)$ for the test sample) using $\hat{\lambda}_t = \hat{\mu}_t/(1-\hat{\pi})$ and $\hat{\lambda}_c = \hat{\mu}_c/(1-\hat{\pi})$ for the test and control samples respectively. Then we simulate data 2000 times using those means to define the parameters in the simulation model, adding random effects for the feed-day and COM, and including zero inflation. We then recalculate the TRA 2000 times, once for each data set, using the $100(1-\tilde{\mu}_t/\tilde{\mu}_c)$ estimator, where for $a \in \{c,t\}$ we let $\tilde{\mu}_a = \hat{\mu}_a$ unless all n_a mosquitoes had 0 oocysts, then $\tilde{\mu}_a = 0.5/n_a$. We use the middle 95% of the simulated TRA values to get the confidence interval. The sixth method (ZINBRE-Sim2) builds upon the fifth, replacing the static $\hat{\sigma}_c = 0.2306$ with a random σ_c that follows a log-normal distribution with mean 0.2306 and standard deviation 0.0071505 (the standard error of $\hat{\sigma}_c$ estimated from the RCM).

We evaluate the 6 methods by simulation under 12 scenarios. The (TRA, σ_c) pair denote the simulation scenarios. We carry out 4 levels of TRA (0, 20, 50, 80) and 3 levels of COM variability (σ_c from the RCM, and the lower and upper 95% bootstrap CI bounds for σ_c using the 9804 control mosquitoes and 2000 bootstrap samples) for a total of 12 simulation scenarios. Each scenario had a 1000 iterations, and the percent of iterations where the calculated 95% CI covered the true TRA are displayed in Table 1. Although assuming $\hat{\sigma}_c$ is known and fixed is a crude

							ZINBRE	
TRA	$Q(\sigma_c)$	σ_c	t-Welch	WMW-HL	Delta	ZINB-Norm	Sim1	Sim2
0	est	0.2306	86.3	83.4	80.1	75.6	94.7	95.9
20	est	0.2306	86.9	84.3	80.8	74.5	95.9	96.7
50	est	0.2306	85.0	82.5	81.3	78.6	94.5	95.6
80	est	0.2306	74.4	76.1	81.3	81.3	94.7	96.3
0	2.5%	0.1587	91.0	89.0	88.6	84.9	98.7	99.2
20	2.5%	0.1587	90.7	87.2	88.1	84.8	98.0	98.7
50	2.5%	0.1587	88.9	87.8	88.7	84.7	98.5	98.7
80	2.5%	0.1587	78.7	81.1	87.6	87.0	98.3	98.2
0	97.5%	0.3006	79.3	75.3	71.8	69.2	90.4	91.8
20	97.5%	0.3006	82.0	77.4	74.3	69.1	92.2	93.2
50	97.5%	0.3006	78.8	74.9	73.3	68.5	90.8	92.4
80	97.5%	0.3006	73.2	73.4	75.6	72.7	91.2	91.9

Table 1: Coverage Probabilities of 1 Feed-Day Confidence Intervals.

approximation (ZINBRE-Sim1), Table 1 shows that both simulation approaches perform better in terms of simulated coverage than implicitly assuming $\sigma_c = 0$ as is done for the first 4 methods. Allowing variability on $\hat{\sigma}_c$ (ZINBRE-Sim2) gives slightly higher coverage than $\hat{\sigma}_c$ fixed.

5 TBA

The TRA estimand afforded by the model has no dependency on the control mean or random effects. Thus, the TRA for a test sample can be estimated from a single feed-day (with a COM for test and a COM for control), or it can be estimated by dividing the test sample into multiple aliquots that can be measured on different feed-days (each with their own control COM) and using

the ratio of the mean of the test COMs over the mean of the control COMs. We show in this section the expected value of the naive TBA estimate, b, depends on the TRA, control mean, and the standard deviations of the feed-day and COM random effects. We develop a feed-day specific TBA estimand assuming no COM random effects. This estimand depends on the feed-day specific control mean, which can be replaced with a target mean as a possible way of standardization, as demonstrated by the standardized TBA estimators in Section 5.2.

5.1 Defining the TBA Estimand

Using the fact that if Z is a zero-inflated negative binomial random variable,

$$E(I(Z>0)) = (1-\pi)\left\{1-\left(\frac{\theta}{\mu+\theta}\right)^{\theta}\right\},$$

and π does not depend on X_{ij}, the overall TBA would be derived from the model as:

$$\begin{aligned}\Omega_B^{overall} &= 100\left(1-\frac{E(I(Y_{ijk}>0)|X_{ij}=1)}{E(I(Y_{i'j'k'}>0)|X_{i'j'}=0)}\right)\\ &= 100\left(1-\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{1-\left(\frac{\theta}{\exp(\gamma+\beta+s+t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)\frac{1}{\sigma_f}\phi\left(\frac{s}{\sigma_f}\right)ds\ dt}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{1-\left(\frac{\theta}{\exp(\gamma+s+t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)\frac{1}{\sigma_f}\phi\left(\frac{s}{\sigma_f}\right)ds\ dt}\right),\end{aligned}$$

which has no closed form. Aside from being mathematically intractable, the result depends on nuisance COM effects (the ideal assay should have no COM effect ($\sigma_c = 0$)). The feed-day effect is also a nuisance component of the model (but with a more substantial variance component $\hat{\sigma_f}^2 = 1.08$). Given these points, the nuisance of feed-day effects will be handled via conditional expectation and then the result taken as a limit of $\sigma_c \to 0$.

Consider a feed-day-specific TBA,

$$\begin{aligned}
\Omega_B^f &= 100\left(1 - \frac{E(I(Y_{ijk} > 0)|X_{ij} = 1, \text{feed-day} = i)}{E(I(Y_{ij'k'} > 0)|X_{ij'} = 0, \text{feed-day} = i)}\right) \\
&= 100\left(1 - \frac{\int_{-\infty}^{\infty}\left\{1 - \left(\frac{\theta}{\exp(\gamma+\beta+f_i+t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)dt}{\int_{-\infty}^{\infty}\left\{1 - \left(\frac{\theta}{\exp(\gamma+f_i+t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)dt}\right) \\
&= 100\left(1 - \frac{\int_{-\infty}^{\infty}\left\{1 - \left(\frac{\theta}{\rho\mu_{ci}\exp(t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)dt}{\int_{-\infty}^{\infty}\left\{1 - \left(\frac{\theta}{\mu_{ci}\exp(t)+\theta}\right)^{\theta}\right\}\frac{1}{\sigma_c}\phi\left(\frac{t}{\sigma_c}\right)dt}\right) \\
&=: \Omega_B^f\left(\mu_{ci}, \rho, \theta, \sigma_c\right)
\end{aligned}$$

where $\mu_{ci} = \exp(\gamma + f_i)$. In the limit when σ_c goes to zero, we have

$$\Omega_B^f(\mu_{ci}, \rho, \theta, 0) \equiv \lim_{\sigma_c \to 0} \Omega_B^f(\mu_{ci}, \rho, \theta, \sigma_c) \quad = \quad 100\left(1 - \frac{1 - \left(\frac{\theta}{\rho\mu_{ci}+\theta}\right)^{\theta}}{1 - \left(\frac{\theta}{\mu_{ci}+\theta}\right)^{\theta}}\right).$$

Note that Ω_B^f depends on the feed-day i, even with very small random effects for COM. There is little practical interest in a TBA for a particular feed-day. For a general TBA, μ_{ci} could be set to some target value, say μ_c^*, to give the targeted TBA estimand:

$$\Omega_B(\mu_c^*, \rho, \theta) \quad \coloneqq \quad \Omega_B^f(\mu_{ci} = \mu_c^*, \rho, \theta, 0),$$

which is our estimand of interest for TBA.

5.2 Two estimators of $\Omega_B(\mu_c^*)$

We propose two estimators of $\Omega_B(\mu_c^*, \rho, \theta)$, a restricted TBA estimate that discards samples that have a control mean outside of an envelope around a target mean, and a standardized TBA that uses the model to utilize any sample regardless of the observed control mean.

Restricted TBA

For the ith feed-day let $\hat{\mu}_{ic}$ and $\hat{p}_{ic}$ be the mean oocyst count and the sample proportion for the control COM, and similarly define $\hat{\mu}_{it}$ and $\hat{p}_{it}$ for the test COM. For a target control mean μ_c^*, define two restriction bounds $\tau_L < \mu_c^*$ and $\tau_U > \mu_c^*$ and calculate TBA only if $\hat{\mu}_{ic} \in (\tau_L, \tau_U)$, otherwise discard the feed-day. The restricted TBA is then

$$\widehat{\Omega}_{B1,i} = 100\left(1 - \frac{\hat{p}_{it}}{\hat{p}_{ic}}\right),$$

provided $\hat{\mu}_{ic} \in (\tau_L, \tau_U)$.

Standardized TBA

For a target control mean μ_c^*, a standardized TBA for the ith feed-day is

$$\begin{aligned}\widehat{\Omega}_{B2,i} &= \widehat{\Omega}_B(\mu_c^*, \hat{\rho}_i, \hat{\theta}) \\ &= 100\left(1 - \frac{1 - \left(\frac{\hat{\theta}}{\hat{\rho}_i\mu_c^* + \hat{\theta}}\right)^{\hat{\theta}}}{1 - \left(\frac{\hat{\theta}}{\mu_c^* + \hat{\theta}}\right)^{\hat{\theta}}}\right).\end{aligned}$$

where $\hat{\rho}_i = \hat{\mu}_{it}/\hat{\mu}_{ic}$ and $\hat{\theta} = 1.93$ is estimated from the RCM. Note $\widehat{\Omega}_{B2,i}$ is calculated regardless of the value of $\hat{\mu}_{ic}$, and it discards no data, unlike $\widehat{\Omega}_{B1,i}$.

5.3 Demonstration of the Need for TBA Standardization or Restriction

In this section we use monoclonal antibodies to create 100 ostensively identical test samples and compare unstandardized and standardized TBA estimators. We show that the unstandardized TBA estimators are highly dependent on the control mean, while the standardized ones are not.

We take 100 test samples (from 98 feed-days) of 4B7 monoclonal antibody each prepared at a concentration of 94 $\mu g/ml$. We randomly assign each test sample to a unique control COM from the same feed-day; that is we opt to omit a few test COMs to ensure that each COM

in the analysis is used once. For the ith pair we calculate an unstandardized TBA estimate, $b_i = 100\,(1 - \hat{p}_{it}/\hat{p}_{ic})$, where $\hat{p}_{it}$ and $\hat{p}_{ic}$ are the proportion of infected mosquitoes from the test and control COM, respectively. On the left three panels of Figure 6 we plot these b_i by the associated control mean, $\hat{\mu}_{ci}$, using circles, together with 95% confidence intervals calculated using melded binomial confidence intervals on the ratio, $\frac{\hat{p}_{it}}{\hat{p}_{ic}}$ (Fay et al., 2015), and transforming the confidence limits on the ratios to limits on the TBA. The three left panels are 'low' ($\hat{\mu}_{ic} \in (0, 6]$), 'med' ($\hat{\mu}_{ic} \in (6, 30]$), and 'high' ($\hat{\mu}_{ic} \in (30, 81]$). On the right three panels we plot the standardized TBA estimates with circles, using $\hat{\Omega}_{B2,i}(\mu_c^* = 2, \hat{\rho}_i, \hat{\theta})$, where $\hat{\rho}_i = \hat{\mu}_{it}/\hat{\mu}_{ic}$ and $\hat{\theta} = 1.93$. Actual μ values in the malaria endemic fields can be very low and also vary depending on many factors, such as endemicity, season, and mosquito species. For the sake of argument, $\mu_c^* = 2$ was chosen as a reasonable, representative value in the field. The associated 95% confidence intervals are the single-subject TRA confidence intervals (ZINBRE-Sim2, Section 4.3) converted to the TBA scale. In other words, we take the single-subject TRA confidence intervals, convert them to limits on ρ_i, then insert those limits into the function $\hat{\Omega}_{B2,i}(2, \hat{\rho}_i, \hat{\theta})$ to get 95% confidence limits on $\Omega_B(2)$.

We summarize the TBA estimates within each panel using a one sample t-test on the transformed values $\log(1 - TBA/100)$. We use that transformation so that the values are closer to normally distributed. The means and 95% confidence intervals on the transformed values are back-transformed onto the TBA-scale, and are presented as squares and vertical bars on the right side of each panel (ignore the corresponding x-value). (The meta-regression estimates (triangles) and bars are an alternate way of summarizing the data are discussed in Sections 6 and Supplementary Section 10.)

In Figure 6 we see from the square symbols that the unstandardized TBA estimates decrease as the control mean increases, while the standardized TBA estimates do not. The standardization tends to shift up the estimates of the medium and high control mean groups to a similar level of the low control mean group, which is expected because the target mean of 2 is in the range of the low group. We can think of the unstandardized TBA estimates in the leftmost panel as restricted

TBA estimates, restricted to $\hat{\mu}_{ic} \in (0, 6]$. We expect the unstandardized TBA estimates in the med and high panels to be poor estimators of $\Omega_B(2)$, because their control means are far from the target of $\mu_c^* = 2$. The right two panels of Figure 6 clearly show that we can estimate $\Omega_B(\mu_c^*)$ by standardized TBA fairly well.

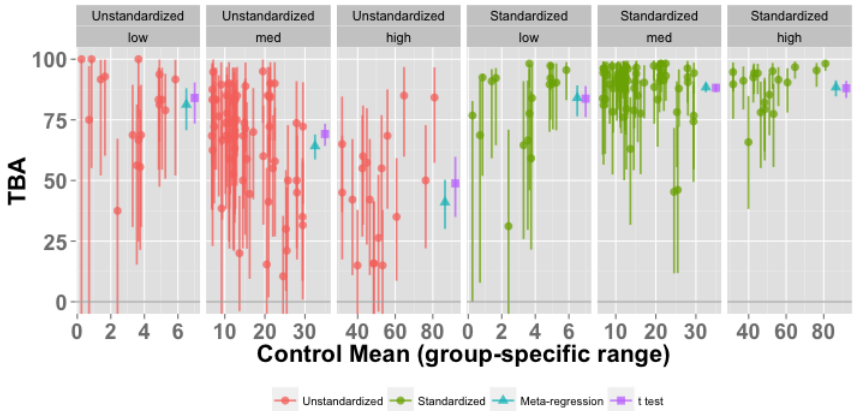

Figure 6: Comparing Unstandardized and Standardized TBAs for a 4B7 monoclonal antibody 94 $\mu g/ml$ dose, with meta-estimates, stratified by control mean groups (low, med, high). For a target mean of 2, the Unstandardized-low panel corresponds to the empirical restricted TBA estimator, the Unstandardized-med and Unstandardized-high would be discarded. The 3 right panels show the Standardized TBA estimator; note how standardization tends to shift Unstandardized TBA estimates of the med and high group up to resemble those of the low group (which contains the target mean), effectively removing the dependence of unstandardized TBA on the control mean. The meta-regression and t-test meta-estimates are very similar due to the sample sizes being equal for every TBA estimate.

5.4 Comparing the Two estimators of $\Omega_B(\mu_c^*)$

If the single-subject TBA is desired, the estimand and estimators imply some kind of restriction or standardization is in order. In the previous section we showed how this might take place with

a real-data demonstration. Here we compare the restricted and standardized TBA estimators in a variety of conditions. Using $\hat{\psi}^{RCM}$ (Section 3.2), we simulated 10,000 data sets for each of 80 simulation scenarios (10 TRAs $\times$ 4 levels of mosquitoes in a COM $\times$ 2 target control mean oocysts).

Let I_+ be the number of iterations where $\hat{\mu}_{ic} \in (\mu_c^* - 2, \mu_c^* + 2)$, the simulated mean squared error (sMSE) of $\widehat{\Omega}_{B1}$ is calculated as:

$$\frac{1}{I_+} \sum \left(\widehat{\Omega}_{B1,i} - \Omega_B(\mu_c^*, \rho, \hat{\theta}) \right)^2$$

where the summation is only over the I_+ iterations that are not discarded.

To give equal footing, we compare each restricted TBA to an average of standardized TBAs, averaged over the iteration that calculated the restricted TBA and all subsequent iterations that were discarded before the next restricted TBA was calculated. In other words, we compare $\widehat{\Omega}_{B1,j}$ with $\widehat{\Omega}_{B2avg,j} = \frac{1}{g_j} \sum_{i=j}^{k} \widehat{\Omega}_{B2,i}$ where the $(k+1)$st iteration is the next iteration that is not discarded, and $g_j = k - j + 1$. The associated sMSE is

$$\frac{1}{I_+} \sum \left(\widehat{\Omega}_{B2avg,i} - \Omega_B(\mu_c^*, \rho, \hat{\theta}) \right)^2$$

Simulated mean squared errors are calculated for each estimator and displayed in Figure 7.

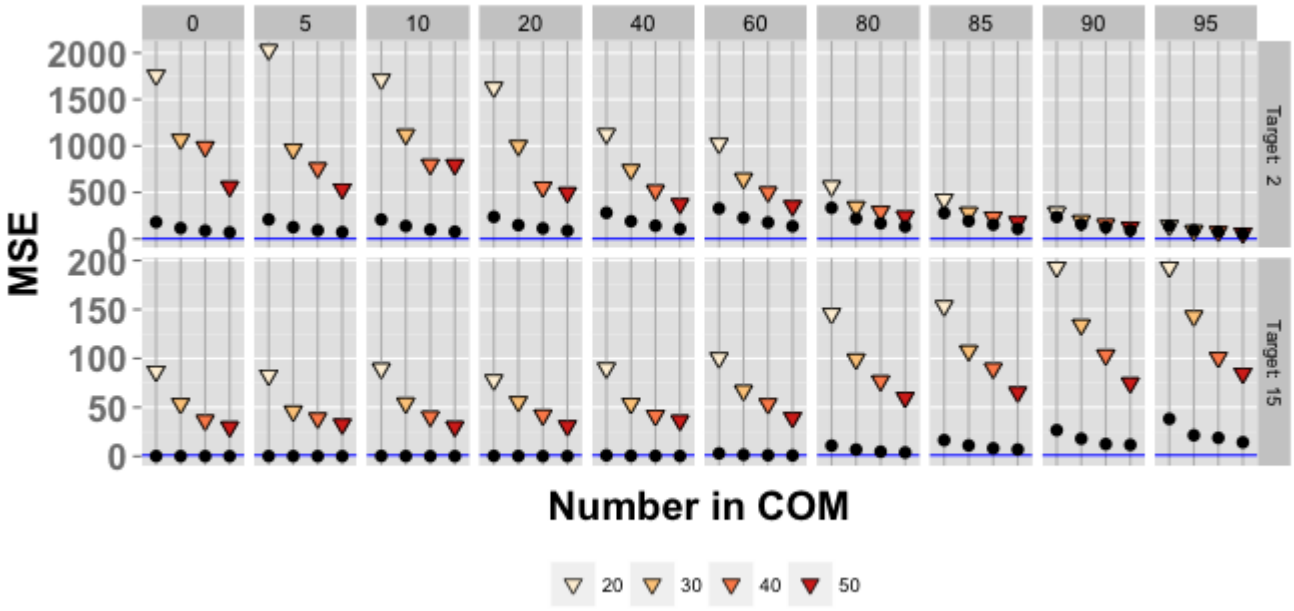

Figure 7: Simulated MSEs of the restricted TBA (shaded triangles) and standardized TBA (black dots) approaches for targets of $\mu_c^* = 2$ and $\mu_c^* = 15$ for different number of mosquitoes in each COM ($M \in \{20, 30, 40, 50\}$) and different True TRA values $\Omega_R \in \{0, 5, 10, 20, 40, 60, 80, 85, 90, 95\}$.

The standardized TBA is more precise across all levels except for the highest TRA of control mean 2. For a given target control mean and TRA, increasing the number of mosquitoes in each COM will lower the MSE.

5.5 Confidence Interval for single-subject Standardized TBA

In addition to being more efficient than the restricted TBA, as shown in the previous section, the standardized TBA provides a simple mechanism for single-subject confidence intervals. If (TRA_L, TRA_U) are the lower and upper bound for the 95% CI for a TRA estimate, then the 95%

CI of a TBA estimate for a target control mean μ_c^* is:

$$\left(100\left(1-\frac{1-\left(\frac{\hat{\theta}}{\rho_L u_s+\hat{\theta}}\right)^{\hat{\theta}}}{1-\left(\frac{\hat{\theta}}{u_s+\hat{\theta}}\right)^{\hat{\theta}}}\right), 100\left(1-\frac{1-\left(\frac{\hat{\theta}}{\rho_U u_s+\hat{\theta}}\right)^{\hat{\theta}}}{1-\left(\frac{\hat{\theta}}{u_s+\hat{\theta}}\right)^{\hat{\theta}}}\right)\right)$$

where $\rho_K = 1 - TRA_K/100$. Since these are monotonic increasing functions for ρ given a fixed μ_c^*, the coverage probabilities for the targeted TBA would match those of the corresponding to the (TRA_L, TRA_U) for the true TRA as reported in the Section 4.3. Additionally, any assay that showed TRA significantly greater than 0 would induce targeted TBA significantly greater than 0, regardless of the target control mean. For example, if a TRA estimate in the lab was 85 (95% CI (82, 88)), we would estimate the TBA in the lab ($\mu_c^* = 15$) to be 21 (95% CI (17, 27)) whereas in the field ($\mu_c^* = 2$) to be 67 (95% CI (62, 73)). The discrepancy of TBA standardized to two different μ_c^* values underscores the importance of standardization as part of the decision making process in the TBV development pipeline.

6 Combining Information from Different Test Samples

In many situations, the assay results for many test samples will need to be combined. For example, consider an animal study or a human phase II trial of a transmission blocking vaccine compared to placebo. The primary efficacy outcome for each subject may be the TRA at one specific time post vaccination. We use TRA instead of TBA since the TRA does not depend on the true unknown control means. Additionally, the ranks of the TRA and the standardized TBA will be the same anyway. For a two sample test, we can simply use a Wilcoxon-Mann-Whitney (WMW) test on the sample TRA (r, from equation 1). Even if our model induces some bias due the study design and the lack of fit of the model to the true one, then because of randomization, the WMW test will still produce valid inferences.

If we wanted estimates of effect size, we can use t-test (or Hodges-Lehman) confidence intervals. For these confidence intervals we first transform the responses, using $\ell_k = \log(\hat{\rho}_k) =$

$\log(1 - \widehat{TRA}_k/100)$ as the outcome for the kth subject, since these transformed values will have a distribution that is closer to a normal distribution. When we observe complete blockage (all mosquitoes in the test COM have 0 oocysts), then we add 0.5 to the first mosquito before taking the log transformation. The Welch's t-test will naturally lead to estimates and confidence intervals on the mean difference in the ℓ_k between the groups. Some have used meta-analysis regression to combine information across many assays (Da et al., 2013), but upon closer examination, there is little practical advantage of using meta-analysis regression over t-tests in the SMFA setting (see Supplement Section 10).

Let us interpret that difference within the context of our model. Previously, we used μ_{ij} to represent the mean oocyst count in the ith feed-day and the jth COM within that feed-day. Now add super scripts to denote the subject and whether the COM is the test or control COM associated with that subject. For example, μ_{ij}^{kt} and $\mu_{ij'}^{kc}$ are the mean oocyst counts for the kth subject's test COM and associated control COM, respectively. Let the TRA effect for the kth subject be $100(1 - \exp(\beta_k))$. Then according to our model,

$$\begin{aligned} E(\ell_k) &= E\left\{\log(\mu_{ij}^{kt}) - \log(\mu_{ij'}^{kc})\right\} \\ &= E(\gamma + \beta_k + f_i + c_{ij} - \gamma - f_i - c_{ij'}) = \beta_k + E(c_{ij} - c_{ij'}) = \beta_k. \end{aligned} \tag{3}$$

Assume that if subject k got the experimental vaccine (or placebo), then β_k comes from a distribution with mean B_v (or B_p). Then the expected difference in means of the ℓ_k values would be $B_v - B_p$. We can interpret $100(1 - \exp(B_v - B_p)) \equiv TRA_{vp}$ as the estimand of the vaccine's TRA effect compared to placebo, and call this the transmission reducing vaccine activity (TRVA). In terms of the model, letting ℓ_k^v and $\ell_{k'}^p$ be typical responses from the vaccine (ijth feed-COM) and placebo (abth feed-COM) group, we get,

$$\begin{aligned} E(\ell_k^v - \ell_{k'}^p) &= E\left\{\log(\mu_{ij}^{kt}) - \log(\mu_{ij'}^{kc}) - \left[\log(\mu_{ab}^{k't}) - \log(\mu_{ab'}^{k'c})\right]\right\} \\ &= E\left\{\log(\mu_{ij}^{kt}) - \log(\mu_{ab}^{k't})\right\} + E\left\{\log(\mu_{ab'}^{k'c}) - \log(\mu_{ij'}^{kc})\right\} \\ &= E\left\{\log(\mu_{ij}^{kt}) - \log(\mu_{ab}^{k't})\right\}. \end{aligned} \tag{4}$$

By comparing equations 3 and 4, we see that in the TRVA effect, the placebo effect is acting like the control in the single sample TRA. For the TBA effects, there is no such simple vaccine effect, since the TBA effect for the kth subject depends on the target control mean and the overdispersion parameter. Thus, for simplicity of reporting, we recommend presenting inferences in terms of TRVA for two sample situations, rather than defining an analogous estimand for TBA. Alternatively, we can simply estimate the expected value of individual standardized TBA values within group using the t-test (see Section 5.3).

7 Designing Multiple Assays for a Study

Consider a two sample trial comparing a novel vaccine to a standard vaccine. If our concern was only with the TRVA, then we need not use control COMs at all. From the TRVA definition (see equation 4) the control COM effects cancel out and need not be measured. Ideally, we would randomly assign an equal number of new vaccine sample COMs and standard vaccine sample COMs on each assay feed-day. Then the TRVA could be estimated using a linear model on the log transformed means for those COMs (with suitable adjustments for zero counts), with vaccine-type and feed-day main effects. Alternatively, a stratified Wilcoxon-Mann-Whitney test could compare the means from the new vaccine COMs and the standard vaccine COMs. In either of these cases, we automatically account for the random COM effects, since those effects would be part of the random response error, where the subjects' responses are their COM means.

Practically speaking, it is likely that subject level estimates of TRA or standardized TBA will be desired, so it makes sense to include at least one control COM within each feed so that those estimates can be calculated. In this latter case, the TRA estimates (and standardized TBA estimates that depend on them) would be correlated, so they would only be used as descriptive statistics.

8 Discussion

In this paper we have developed a statistical model for SMFAs used for the development of transmission blocking vaccines in malaria. This assay measures counts of oocysts from two sets of mosquitoes, those fed from a control sample and those fed from a test sample. Of the two possible readouts, TRA and TBA, we have demonstrated that TRA is usually preferred due to its generalizability among different assays.

In some cases a TBA estimate is desired. Users of this assay have realized that the TBA measurement (equation 2) depends on the mean oocyst counts from the control sample, which is difficult to control by experimental design. An expensive solution is to restrict TBA measurements to only those values associated with control means within a narrow range, and discard many assays outside of that range. Our statistical model addresses this problem by defining a TBA estimand from the model. The TBA estimand depends on a target control mean, the TRA, and the overdispersion parameter. With this model, we define a standardized TBA estimator based on the sample TRA that can be standardized to any pre-specified target control mean. For a single sample assay run, this requires using $\hat{\theta}$, an estimated overdispersion from extensive SMFA replicate data. We have shown through simulations that this standardized TBA estimator is much more efficient than using a restricted TBA estimator (restricting to only control means within a range).

Ideally, the assay readout from a test sample should have variability small enough that it is ignorable. But unfortunately, that is not feasible in this situation. Therefore, it is desirable to have confidence intervals for each test readout, to determine if each test sample shows significant transmission blocking (or reducing) activity. We have shown that simple methods which treat each mosquito as independent do not perform well and have developed a simulation-based method to calculate individual assay readout confidence intervals for both TRA and TBA. These confidence intervals also depend on $\hat{\theta}$.

We have demonstrated a pressing need for standardization of the TBA estimate (see Section 5.3), and proposed one standardized TBA estimator. A potential drawback of our standardized TBA

estimate is its dependency on one specific model, and on the specific value, $\hat{\theta}$. A concern is that the model may not fit SMFA data well from other laboratories or on slightly modified assays. We have partially addressed this question by showing that the model developed and fit using control polyclonal antibody data fits test polyclonal and monoclonal antibody data quite well (with R^2 values greater than 0.90). Churcher et al. (2012) used a similar but more complicated model for SMFA data, which allows the zero inflation and the overdispersion parameters to depend on the mean oocyst counts through non-linear functions. Although they have not used their model for standardizing TBA estimates, their model does relate the TBA, control mean, and the TRA (see Churcher et al., 2012, Figure 2B), and it provides a similar relationship as our simpler model (Reference Blinded). Although Churcher et al. (2012) show that the added complexity of their model is better for their data, it is not clear how their model will perform on data from other laboratories. As with any model building, we need to balance the closeness of the fit to the model-building data (which can be improved with more complex models) and expected closeness of fit to other data sets from slightly modified assay designs (which may be diminished with overly complex models). We have shown that our models fit the independent polyclonal and monoclonal data well, but further work can be done to explore the use of our model (or similar models, e.g., Churcher et al., 2012) on different data (e.g., from other laboratories). Especially of interest is how much the overdispersion parameter estimate changes between laboratories, and how this will affect inferences.

Despite this potential drawback, one advantage of the proposed standardized TBA, which bases the TBA confidence intervals on the associated confidence intervals for TRA, is that the TBA and TRA inferences will match in the sense that test samples that have TRA statistically significantly different from 0 will also have statistically significant standardized TBA different from 0 and vice versa. This matching property does not depend on the target control mean, the goodness-of-fit of our model, or its generalizability. The fit of the model is more important when a TBA estimate is needed to use in some other prediction model (see e.g., Smith et al., 2011). Nonetheless, we have proposed a model for malaria researchers to use to address the issues with TBA, and rigorously

demonstrated whether restriction or standardization is used, *some* kind of adjustment must be administered if TBA estimates are used. Future applications of the model will determine its ability to enhance the utility of the the SMFA in vaccine development.

Acknowledgements

We thank Erin Gabriel for discussions on the statistical analysis of standard membrane-feeding assays. This study was supported in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH and also by the PATH Malaria Vaccine Initiative. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)

References

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C., and Biggeri, A. (2002). Some general points in estimating heterogeneity variance with the dersimonian–laird estimator. *Biostatistics*, 3(4):445–457.

Bousema, T., Churcher, T. S., Morlais, I., and Dinglasan, R. R. (2013). Can field-based mosquito feeding assays be used for evaluating transmission-blocking interventions? *Trends in parasitology*, 29(2):53–59.

Bousema, T., Dinglasan, R. R., Morlais, I., Gouagna, L. C., van Warmerdam, T., Awono-Ambene, P. H., Bonnet, S., Diallo, M., Coulibaly, M., Tchuinkam, T., et al. (2012). Mosquito feeding assays to determine the infectiousness of naturally infected plasmodium falciparum gametocyte carriers. *PLoS ONE*, 7(8):e42821.

Churcher, T. S., Blagborough, A. M., Delves, M., Ramakrishnan, C., Kapulu, M. C., Williams, A. R., Biswas, S., Da, D. F., Cohuet, A., and Sinden, R. E. (2012). Measuring the blockade of

malaria transmission–an analysis of the standard membrane feeding assay. *International journal for parasitology*, 42(11):1037–1044.

Da, D. F., Dixit, S., Sattabonkot, J., Mu, J., Abate, L., Ramineni, B., Ouedraogo, J. B., MacDonald, N. J., Fay, M. P., Su, X.-z., et al. (2013). Anti-pfs25 human plasma reduces transmission of plasmodium falciparum isolates that have diverse genetic backgrounds. *Infection and immunity*, 81(6):1984–1989.

Delrieu, I., Leboulleux, D., Ivinson, K., Gessner, B. D., et al. (2015). Design of a phase iii cluster randomized trial to assess the efficacy and safety of a malaria transmission blocking vaccine. *Vaccine*, 33(13):1518–1526.

Fay, M. P., Proschan, M. A., and Brittain, E. (2015). Combining one-sample confidence procedures for inference in the two-sample case. *Biometrics*, 71(1):146–156.

Jackman, S. (2015). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Department of Political Science, Stanford University, Stanford, California. R package version 1.4.9.

malERA Consultative Group on Vaccines et al. (2011). A research agenda for malaria eradication: vaccines. *PLoS Med*, 8(1):e1000398.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

Miura, K., Deng, B., Tullo, G., Diouf, A., Moretz, S. E., Locke, E., Morin, M., Fay, M. P., and Long, C. A. (2013). Qualification of standard membrane-feeding assay with plasmodium falciparum malaria and potential improvements for future assays. *PloS one*, 8(3):e57909.

Nunes, J. K., Woods, C., Carter, T., Raphael, T., Morin, M. J., Diallo, D., Leboulleux, D., Jain, S., Loucq, C., Kaslow, D. C., et al. (2014). Development of a transmission-blocking malaria vaccine: Progress, challenges, and the path forward. *Vaccine*, 32(43):5531–5539.

RTS,S Clinical Trials Partnership (2014). Efficacy and safety of the rts, s/as01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 african sites. *PLoS medicine*, 11.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

Sinden, R. E., Blagborough, A. M., Churcher, T., Ramakrishnan, C., Biswas, S., and Delves, M. J. (2012). The design and interpretation of laboratory assays measuring mosquito transmission of plasmodium. *Trends in parasitology*, 28(11):457–465.

Smith, T. A., Chitnis, N., Briët, O. J., and Tanner, M. (2011). Uses of mosquito-stage transmission-blocking vaccines against plasmodium falciparum. *Trends in parasitology*, 27(5):190–196.

Viechtbauer, W. et al. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

White, M. T., Verity, R., Griffin, J. T., Asante, K. P., Owusu-Agyei, S., Greenwood, B., Drakeley, C., Gesase, S., Lusingu, J., Ansong, D., et al. (2015). Immunogenicity of the rts, s/as01 malaria vaccine and implications for duration of vaccine efficacy: secondary analysis of data from a phase 3 randomised controlled trial. *The Lancet Infectious Diseases*.

WHO (2015). Malaria vaccine rainbow tables. `http://www.who.int/vaccine_research/links/Rainbow/en/index.html`. [accessed 04-Sep-2015].

WHO (2015). *World Malaria Report 2015*. World Health Organization.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8).

9 Model Selection Statistics (online supplement)

We consider a zero inflated negative binomial model with random COM and feed-day effects. The negative binomial part of the model can be parameterized to have a linear mean-variance relation (mean: λ, variance: $\lambda\theta$) or a quadratic relation (mean: λ, variance: $\lambda + \lambda^2/\theta$), which we refer to as NB1 and NB2 respectively (McCullagh and Nelder, 1989). We list 12 versions of the model in Table 2, each version choses one of 3 count components (NB1,NB2, or PO for Poisson), zero inflation (ZI) or not, and random effects (RE) or not. If there were no random effects, then each COM from the model-building data would have the same mean since that data are all control COMs, and we could not differentiate between NB1 and NB2. So for that model-building data we can only differentiate the two negative binomial models within the context of the random effects. Of the two zero-inflated negative binomial models with random effects (Models 1 and 5) which have the same number of parameters, Model 1 (ZINB2-RE) has the larger log-likelihood (-34798.4 vs. -35062.8). Thus, we prefer NB2 to NB1. We show that NB2 is significantly better than Poisson by comparing Model 1 (ZINB2-RE) to Model 9 (ZIPO-RE) by the likelihood ratio test (LRT) of $\theta = \infty$. Although this LRT (and the others that follow) is on the boundary of the parameter space, we calculate a naive p-value ignoring this fact since this naive p-value is conservative and all our results are highly significant by this method anyway (see Self and Liang, 1987). Model 1 is statistically significantly better than Model 9 (-34798.4 vs. -53031.8, naive p-value < 0.001 on 1 d.f.). We show that zero inflation is needed by comparing Model 1 (ZINB2-RE) to Model 2 (NB2-RE) by the LRT $\pi = 0$. Model 1 is statistically significantly better (-34798.4 vs. -35065.5, naive p-value < 0.001 on 1 d.f.). We show that the random effects are needed comparing Model 1 (ZINB2-RE) to Model 3 (ZINB2) by the LRT $\sigma_f = \sigma_c = 0$. Model 1 is statistically significantly better (-34798.4 vs. -38177.5, naive p-value < 0.001 on 2 d.f.).

Model	Label	γ	θ	π	σ_f^2	σ_c^2	logL
1	ZINB2-RE	2.57	1.93	0.056	1.085	0.053	-34798.4
2	NB2-RE	2.51	1.39	-	1.103	0.057	-35065.5
3	ZINB2	2.98	0.76	0.052	-	-	-38177.5
4	NB2	2.93	0.63	-	-	-	-38217.5
5	ZINB1-RE	2.66	11.75	0.031	0.692	0.034	-35062.8
6	NB1-RE	2.65	14.13	-	0.620	0.037	-35336.8
7	ZINB1	2.98	26.96	0.052	-	-	-38177.5
8	NB1	2.93	30.40	-	-	-	-38217.5
9	ZIPO-RE	2.31	-	0.089	3.415	0.095	-53031.8
10	PO-RE	2.49	-	-	1.107	0.096	-66480.9
11	ZIPO	2.57	-	0.153	-	-	-72583.4
12	PO	2.93	-	-	-	-	-129840.0

Table 2: The 12 models fitted to 9804 Control Mosquitoes (in 492 COMs). Note, the θ parameter has a different interpretation in the NB1 models than in the NB2 models.

10 Using Meta Regression to Combine Assay Results (online supplement)

One approach to inferences on TRVA using the ℓ_k is to use the information from the model about the variance of the ℓ_k in a meta regression approach (see Da et al., 2013). However, this meta regression approach does not necessarily increase efficiency of inferences. Let s_k be an estimate of the standard deviation of ℓ_k. We can estimate s_k from the simulation that is used for the kth confidence interval by using the simulated standard deviation of ℓ_k. Alternatively, we can estimate

s_k using a $100(1-\alpha)\%$ (typically 95%) confidence interval for for the TRA for subject k, say (L_k, U_k), using

$$s_k = \frac{\log(1-\frac{L_k}{100}) - \log(1-\frac{U_k}{100})}{2\Phi^{-1}(1-\alpha)} \quad (5)$$

where $\Phi^{-1}(q)$ is the qth quantile from the standard normal. Then we can use the (ℓ_k, s_k) pairs in a meta regression analysis. Specifically, let X_k be a treatment indicator. Then we assume that

$$\ell_k = \xi + \phi X_k + \epsilon_k$$

where ϵ_k are normal independent errors with mean 0 and variance $\tau^2 + s_k^2$, where s_k^2 is the within-assay variance and τ^2 represents the between-test-sample variance. Typically, we use the estimated s_k^2 and assume that it is estimated without error (see Viechtbauer et al., 2010). Although this may be reasonable for meta analyses where sample sizes vary widely, for the SMFA data most assays will be measured on a similar number of mosquitoes. Besides ignoring the variability of s_k^2, another problem is that the true individual variances may depend in a complex way on the mean (Böhning et al., 2002). Thus, we recommend combining information across different assays using equal weights (i.e., using the unweighted mean of the ℓ_k), and simply using the t confidence intervals described above.