Incorporating efficacy data from initial trials into subsequent evaluations of vaccines against respiratory syncytial virus

Joshua L. Warren[1], Maria Sundaram[2], Virginia E. Pitzer[3],
Saad B. Omer[3,4,5] Daniel M. Weinberger[3*]

[1] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[2] Marshfield Clinic Research Institute, Center for Clinical Epidemiology & Population Health, Marshfield, WI, USA.
[3] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA
[4] Yale Institute of Global Health, New Haven, CT, United States;
[5] Yale School of Medicine, New Haven, CT, United States

*Correspondences to: Daniel Weinberger, daniel.weinberger@yale.edu; PO Box 208034, New Haven, CT 06515.

ABSTRACT

Background.

When a randomized controlled trial fails to demonstrate statistically significant efficacy against the primary endpoint, a potentially costly new trial would need to be conducted to receive licensure. Incorporating data from previous trials might allow for the conduct of more efficient follow-up trials to demonstrate efficacy, speeding the availability of effective vaccines.

Methods

Based on the outcomes from a failed trial of a maternal vaccine against respiratory syncytial virus (RSV), we simulated data for a new Bayesian group-sequential trial. The data were analyzed either ignoring data from the previous trial (i.e., weakly informative prior distributions) or using prior distributions that incorporate the historical data into the analysis. We evaluated scenarios where the efficacy in the new trial was the same, greater than, or less than the efficacy in the original trial. For each scenario, we evaluated the statistical power and type I error rate for estimating the vaccine effect following interim analyses.

Results

If a stringent threshold is used to control the type I error rate, the analyses that incorporated historical data had a small advantage over trials that did not. If control of type I error is less important (e.g., in a post-licensure evaluation), the incorporation of historical data can provide a substantial boost in efficiency.

Conclusions

Due to the need to control the type I error rate in trials used to license a vaccine, the incorporation of historical data provides little additional benefit in terms of stopping the trial early. However, these statistical approaches could be promising in evaluations that use real-world evidence following licensure.

INTRODUCTION

When testing new vaccines, there is a need to ensure that the product is safe and effective. Randomized controlled trials, which are required to license new vaccines, have become increasingly large and costly [1] and have stringent criteria for success. This can lead to delays in introducing potentially lifesaving vaccines or even forestall the development of promising vaccines that fail to meet strict guidelines for approval. An example of this comes from a maternal vaccine against respiratory syncytial virus (RSV) that was tested in a phase 3 clinical trial [2]. RSV is a leading cause of hospitalization in infants globally [3]. The vaccine failed to show efficacy against the pre-specified primary outcome of medically significant RSV (efficacy of 39.4%, with a 97.52% confidence interval (CI) that ranged from -1.0% to 63.7%). However, examining several of the secondary clinical outcomes, there was clear evidence of a beneficial effect against RSV hospitalizations, which also appeared to be larger in lower-middle-income study sites compared to high-income sites. Nevertheless, because the CI for the primary outcome included zero, the trial was deemed a failure, and licensure of the vaccine and further testing were not pursued.

Even if a vaccine is successfully licensed based on data from a phase 3 randomized trial, there is often a requirement from regulators that manufacturers monitor safety and effectiveness in real-world studies. In such studies, the goal is to estimate the effects of the vaccine and detect safety signals in a larger, less restricted population.

Both of these types of evaluations typically ignore data from previous trials and evaluations. In the case of a failed phase 3 trial in which there is evidence of a beneficial but not statistically significant effect of vaccination, one possibility to speed licensure of the vaccine would be to incorporate data from the initial trial into the analysis of data from a new, smaller trial [4,5]. This could be conducted as a traditional trial with a fixed sample size or, potentially more efficiently, as a group sequential Bayesian trial where vaccine efficacy (VE) is evaluated on an ongoing basis and the trial is stopped when sufficient evidence of efficacy has been collected [6]. In the case of a post-licensure evaluation of vaccine effectiveness, data on efficacy against different endpoints, including less common but clinically important endpoints, could likewise be partially incorporated into the analysis. This would allow for more rapid and efficient evaluations with real-world evidence.

In this paper, we focus on the application of methods that incorporate data from previous trials into the design of a hypothetical follow-up to the failed phase 3 trial of a maternal vaccine against RSV. We present simulation analyses that demonstrate how such a trial could be devised and explore the advantages and potential pitfalls of incorporating estimates of VE from a previous trial into the analysis of data from a new trial.

METHODS

Data from previous trial

We use data from the completed phase 3 trial of the Novavax maternal RSV vaccine ("Prepare trial") as the historical data for our analyses. We focus on a secondary endpoint in the trial that showed particular promise and that could be considered as a primary endpoint in a hypothetical subsequent trial seeking licensure: VE against hospitalized RSV lower respiratory tract infections (RSV LRTI), measured at day 90 in a per protocol analysis (44.4%, 95% CI: 19.6%, 61.5%; second table in the original paper [2]). This was based on 53 cases out of 1430 participants in the placebo group and 57 cases out of 2765 participants in the vaccinated group.

Simulation study

We simulated data from a hypothetical new trial with 1:1 randomization between the vaccine and placebo groups. We evaluated study populations that ranged in size from 500 to 3000 participants per group. The primary outcome for each individual was a binary variable describing whether or not they were hospitalized with RSV LRTI during the three-month follow-up period. The binary outcome data from a single simulated trial for person j in treatment arm v (i.e., Case_status_{vj}) were generated as

$$\text{Case_status}_{vj} \sim \text{Bernoulli}(\theta_v);\ v=0,1;\ j = 1,\ldots,n_v;$$
$$\ln(\theta_v) = \beta_0 + \beta_1 * v,$$

where $v=0$ corresponds to the placebo group and $v=1$ to the vaccinated group; $n=n_0 + n_1$ is the total number of individuals enrolled in the trial; θ_v is the true probability of hospitalization for RSV LRTI for every individual in treatment arm v; $\exp(\beta_1)$ represents the relative risk of hospitalization for RSV LRTI in the vaccinated group; and $100*(1-\exp(\beta_1))$ is the true percent VE. When generating the data, β_0 is fixed at $\ln(0.037)$ and β_1 is assigned a fixed value as described for the three scenarios below.

We simulated data representing three scenarios by changing the value of β_1 from the data-generating model:

1) There is no effect of the vaccine in the new trial (VE = 0%; $\beta_1 = 0$)
2) The VE in the new trial is the same as in the Prepare trial (VE=44.4%; $\beta_1 = \ln(0.556)$).
3) The VE in the new trial is higher than in the original trial (VE=90%; $\beta_1 = \ln(0.100)$).

For each scenario, we generated 500 simulated datasets, with up to 6000 participants. In all settings, we fix $\beta_0 = \ln(0.037)$ based on the incidence in the placebo arm of the Prepare trial (i.e., 0.037, or 3.7%).

Design of simulated trial

The simulated trial used a Bayesian group-sequential design. A maternal vaccine is given between 27 and 37 weeks gestational age (mean 32 weeks), and the primary endpoint is the number of cases in the first 90 days of life (mirroring the original trial). Therefore, there are ~5 months between vaccine receipt and completion of the primary endpoint. We consider a trial that enrolls in the Northern and Southern Hemispheres prior to the respective RSV seasons, with a maximum of 6000 participants total enrolled over three years. This results in up to six groups of 1000 participants. Interim analyses are planned after each 1000 participants, which

would allow for the trial to be halted before the other Hemisphere begins enrollment for the season.

Analysis

All analyses of the simulated data are carried out in the Bayesian setting. The observed number of cases among all individuals in treatment arm *v* is modeled using Poisson regression such that

$$\text{N_new_cases}_v \sim \text{Poisson}(\lambda_v = n_v * \theta_v);\ v = 0,1;$$
$$\ln(\lambda_v) = \ln(n_v) + \ln(\theta_v)$$
$$= \ln(n_v) + \beta_0 + \beta_1 * v,$$

where N_new_cases_v is the total number of RSV LRTI hospitalizations observed in treatment arm *v* from the new/simulated trial; λ_v is the expected number of cases; and all other terms have been previously described. A Poisson regression model was used to mirror the analysis in the original Prepare trial.

For analyses that incorporate historical data, the VE from the original trial was also re-estimated using Poisson regression such that

$$\text{N_old_cases}_v \sim \text{Poisson}(\mu_v);\ v = 0,1;$$
$$\ln(\mu_v) = \ln(m_v) + \delta_0 + \delta_1 * v,$$

where N_old_cases_v is the total number of RSV LRTI hospitalizations observed in treatment arm *v* from the original Prepare trial; μ_v is the expected number of cases; m_v is the number of individuals in treatment arm *v* from the original Prepare trial; and δ_0, δ_1 represent the risk in the placebo group and the effect of vaccination, respectively, based on data from the original trial. The β_0, β_1, δ_0, and δ_1 parameters are all estimated simultaneously when incorporating historical data into the analysis.

We evaluated the performance of several different prior distributions for β_0 and β_1. The prior distributions we tested were:

a) *Weakly informative prior distributions, no historical data used.* This scenario most closely resembles a typical clinical trial, where there is no prior information incorporated into the new analysis. The prior distributions in this scenario are $\beta_0, \beta_1 \sim \text{Normal}(0, 100^2)$.

b) *Skeptical prior for* β_1*, no historical data used.* This specification is similar to (a), but with a more informative prior distribution selected to penalize extreme VE values (i.e., large beneficial or adverse effects of the vaccine) ([7]). Specifically, the prior distribution is given as $\beta_1 \sim \text{Normal}(0, 3.32)$ which corresponds to a <5% prior probability that the risk ratio is less than 0.05 or greater than 20 (i.e., a risk ratio roughly between -3 and 3 on the log scale). For the intercept, we still use a weakly informative prior distribution: $\beta_0 \sim \text{Normal}(0, 100^2)$.

c) *Commensurate prior for* β_1 *using historical data*. For this method, originally developed by Hobbs [4], the prior distributions for the current trial parameters are centered at the corresponding parameters from the historical trial such that $\beta_0 \sim \text{Normal}(\delta_0, \sigma^2_0)$ and $\beta_1 \sim$

Normal(δ_1, σ^2_1), with weakly informative priors specified for the historical data parameters (i.e., δ_0, δ_1 ~ Normal(0, 100^2)). Estimation of the variance parameters (i.e., σ^2_0, σ^2_1) is driven by comparing the level of agreement between the historical and current datasets with respect to these parameters. A large variance suggests that only data from the current trial are used to estimate the corresponding parameter, while a small variance results in more information-borrowing from the historical data.

We evaluated several prior distributions for these variance parameters, as they control the level of information-sharing between the historical and current trials. Specifically, we tested σ^2_0, σ^2_1 ~ Inverse-Gamma(0.01,0.01), Inverse-Gamma(1,1), and Inverse-Gamma(3,2). Alternatively, we specified the prior distribution on the standard deviation scale for these parameters such that σ_0, σ_1 ~ Uniform(0,2), Uniform(0,100), and half-Cauchy [8].

All models were fit with the rjags package in R, where we collected 10,000 samples from the joint posterior distribution after discarding the first 10,000 prior to convergence of the model.

Operating characteristics of the trial (frequentist power)
For all analyses, the goal was to estimate the true VE. Because we work in the Bayesian setting, we use posterior tail probabilities to describe statistical significance. The main quantity of interest is the posterior probability that the VE > 0%. We also considered a more stringent endpoint with the posterior probability that the VE > 30%. Across the 500 simulated trials, statistical power was estimated as the proportion of trials whose corresponding posterior probability was greater than (1 - α). Without adjustment for type I error, α is set to 0.05 (i.e., one-sided test). To control the type I error for sequential trials, we evaluated the proportion of simulated trials stopped due to efficacy at a given population size or earlier when the true VE was 0%. The same α was used at each time point and was selected so that ≤5% of simulated trials were declared a success when the maximum population size (6000 participants) was reached.

Availability of code
All analyses were completed using R v4.1.2, using the rjags package for analysis [9,10]. All code required to replicate these analyses is available at https://github.com/weinbergerlab/sequential_bayesian_trials.

RESULTS

As data accrue in a group sequential trial, the estimate of efficacy is updated, as shown in an example simulated trial where the VE was evaluated after every 200th participant completes the trial **(Figure 1).** With smaller samples sizes, the prior structure and use of historical information had a larger influence on the inference, resulting in narrower credible intervals when the true VE was 0% or 44.4% **(Figure 1)**. When the true VE was 90% (greater VE than in the historical data), the incorporation of historical information resulted in wider credible intervals at low sample sizes. When there were more than 2000-3000 participants across both treatment arms

(~50 cases, **Figure S1**), the estimates were similar regardless of the choice of prior distribution or the use of historical information (**Figure 1)**.

Examining the expected trial performance across 500 simulated trials, the analyses that incorporated historical data performed only slightly better in terms of power (**Figure 2, Figure S2)** compared with methods that did not use historical data. This is because of the need to use a more stringent threshold to control the type I error when incorporating historical data (**Table S1)**.

Trials tended to end early if the VE was greater than expected (**Figures 2, S2 Table S1)**. If the true VE was 90%, nearly all of the trials would be stopped at the first interim analysis (1000 participants). If the true VE is 44.4%, approximately 25% of the trials would be stopped after 1000 participants were enrolled, 25% would stop with 2000 participants enrolled, and ~25-30% would stop with 3000 participants enrolled. The incorporation of historical data into the analysis resulted in a small but consistent gain in power when the true VE was 44.4% (**Figures 2, S2**), despite the use of a more conservative threshold (**Table S1)**. These numbers can be compared with a trial that uses a single pre-determined sample size. We estimate that, depending on the prior distribution used, 2400-2800 individuals would need to be enrolled in a new trial to achieve 80% power to detect a VE of 44.4%.

Using a more conservative stopping rule (i.e., posterior probability of VE > 30%) resulted in a relatively small number of trials reaching the stopping criteria even when there is a true effect of 44.4% (**Figure S3).** With a true VE of 90%, almost all trials were still declared a success with just 1000 individuals enrolled.

In the context of a licensure trial, control of the type I error rate is essential. However, if the goal is estimation of real-world effectiveness, and there is strong prior evidence of efficacy, it might be appropriate to use a less stringent threshold (e.g., in post-licensure evaluations). In this situation, the models using historical information had a clear advantage in efficiency, with 72% of evaluations having a high posterior probability of an effect (P(VE>0%|data)>0.95), with just 1000 participants, compared with 43% if data from previous studies are ignored **(Figure S4**).

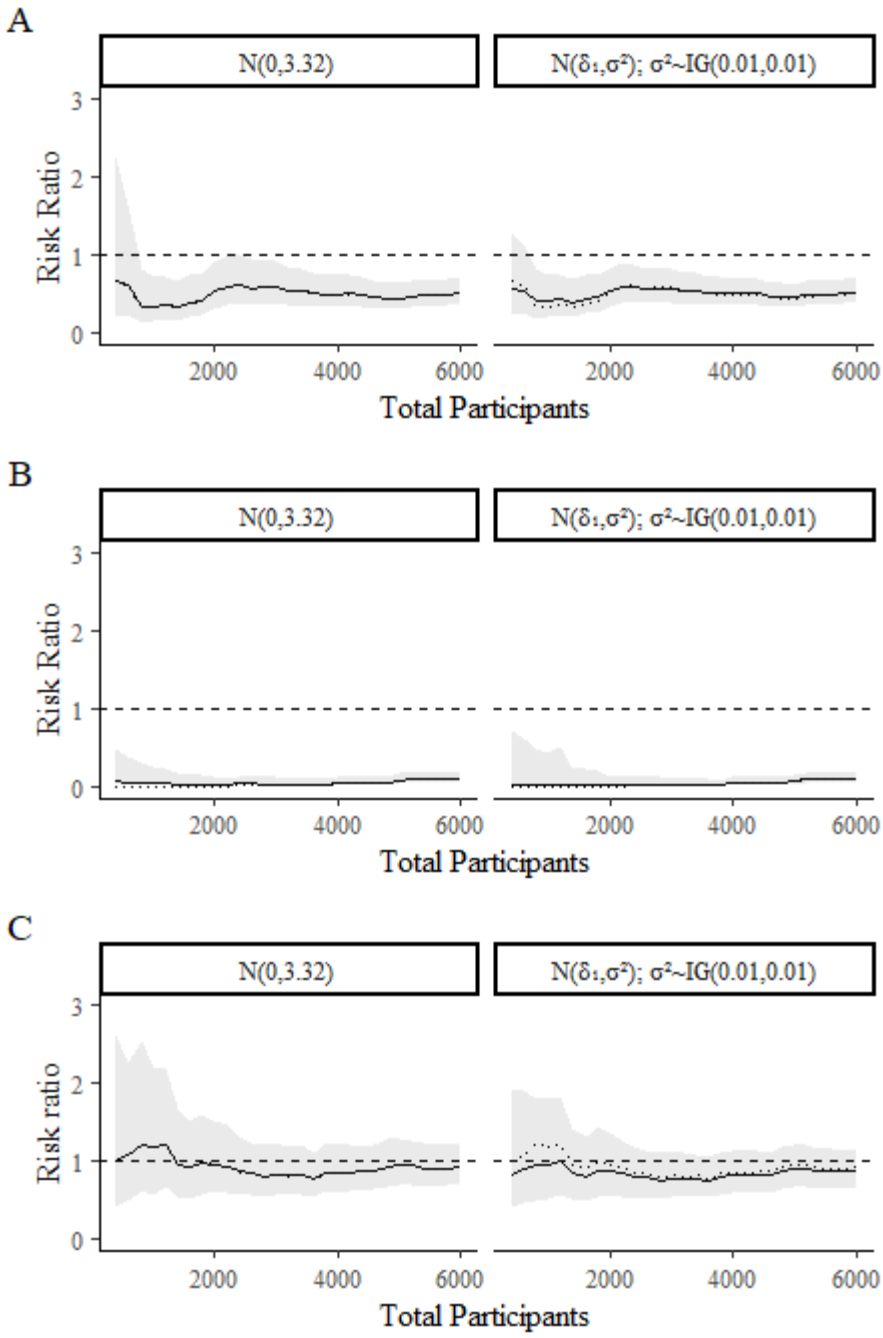

Figure 1. Evolution of the estimated risk ratio in a single trial as more participants complete the trial. In this example, the VE is calculated on a continuous basis after every 200th participant completes the study, starting with the 400th participant. The posterior median estimated values for the risk ratio (100 - VE)/100 are shown with the black line, the 95% credible intervals are represented by the gray shaded area. The dotted line shows the risk ratio calculated empirically based on the observed data only. The dashed line indicates no effect (VE = 0%, risk ratio = 1). In Panel A, the true VE is 44.4 (concordant with the historical data), in panel B, the true VE is 90%, and in panel C, the true VE is 0%. Each sub-panel

represents a different prior structure: skeptical prior with no historical information (N(0,3.32)); commensurate prior with a minimally informative inverse gamma prior on the variance.

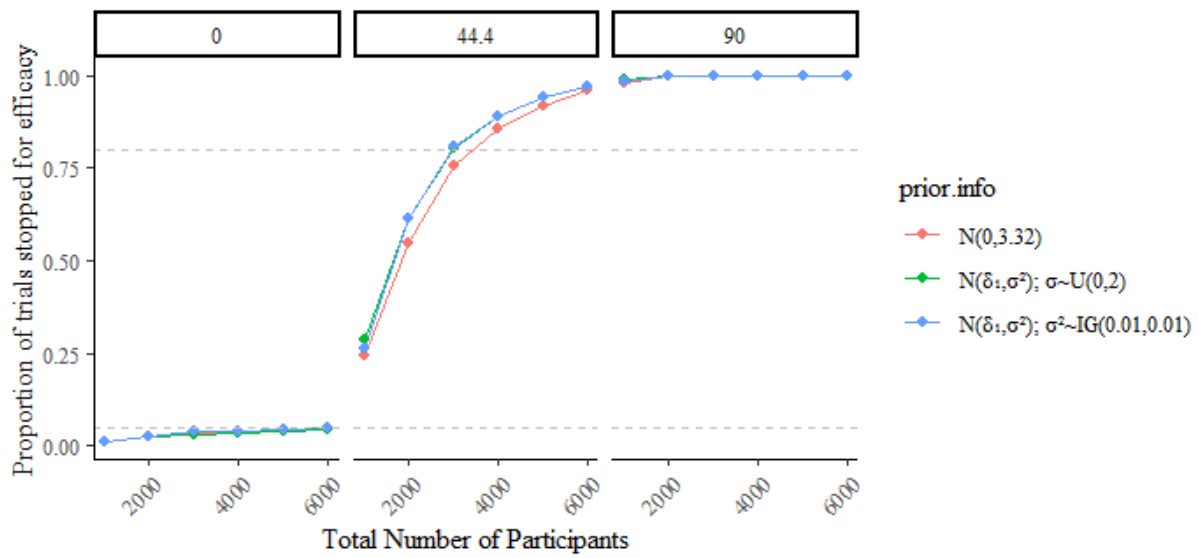

Figure 2. Cumulative proportion of trials that are declared a success (VE > 0%) using a threshold adjusted to control for type I error. VE is evaluated after every 1000 participants complete the follow-up period. From left to right, we consider scenarios with a “true” vaccine efficacy of 0%, 44.4%, and 90%. The horizontal dashed lines are at 0.05 and 0.80.

Table S1. Average final population size when using a stopping criterion that controls type I error, evaluating P(VE>0|data).

Prior on vaccine effect (β_1)*	Vaccine efficacy in new trial	Mean Population size when stopping	Posterior probability threshold for stopping
N(0, 100^2)	0%	5670	0.985
N(0, 3.32)		5680	0.983
N(δ_1, σ^2_1); σ_1 ~ Unif(0, 2)		5730	0.991
N(δ_1, σ^2_1); σ^2_1 ~ IG(0.01, 0.01)		5680	0.993
N(0, 100^2)	44.4%	2800	0.985
N(0, 3.32)		2730	0.983
N(δ_1, σ^2_1); σ_1 ~ Unif(0,2)		2520	0.991
N(δ_1, σ^2_1); σ^2_1 ~ IG(0.01, 0.01)		2550	0.993
N(0, 100^2)	90%	1020	0.985
N(0, 3.32)		1020	0.983
N(δ_1, σ^2_1); σ_1 ~ Unif(0,2)		1010	0.991
N(δ_1, σ^2_1); σ^2_1 ~ IG(0.01, 0.01)		1010	0.993

Vaccine efficacy is calculated as 100(1-exp(β_1))

DISCUSSION

Late-stage clinical trials are costly and time consuming. There are potentially substantial benefits to introducing even a moderately effective intervention into the broader population. It is therefore critical to design trials that can be stopped early once there is strong evidence of efficacy or of futility but that can also be flexible in allowing for continued enrollment if more evidence is needed. Bayesian sequential trials can achieve both of these goals, while also incorporating prior information from previous trials when appropriate. Flexible Bayesian designs should therefore be considered more widely for late-stage vaccine trials by sponsors and regulators.

We considered two complementary approaches for conducting trials more efficiently: (1) The incorporation of data from previous trials, and/or (2) the use of Bayesian group sequential trials. If there is a need to control type I error, the incorporation of data from previous trials yields only a modest improvement in power, a phenomenon that has been noted elsewhere [11]. However, the use of Bayesian group sequential designs that have frequent interim analyses, with or without historical data, could lead to the conduct of more efficient trials with fewer participants, particularly when the intervention effect is strong.

Implementing this type of analysis in practice could be challenging. Regulators are often skeptical of the use of Bayesian methods for licensure trials. There could also be a desire to

have data from the trial stand alone, without incorporation of strong prior information. The general methods could still be applied in other settings, for example by incorporating data from the trials into post-licensure studies that monitor vaccine effectiveness and safety. In such evaluations, the benefits of incorporating the data from prior evaluations is greater (**Figure S4)**, because there might not be as strong of a need to guard against type I error. Conceptually, in a post-licensure evaluation, there is stronger prior information (from phase 3 trials), so it is appropriate to incorporate such information without penalty. This may be particularly valuable for quantifying vaccine effectiveness against rarer but more clinically important outcomes, e.g. severe disease or death.

Trial efficiency (i.e., stopping as early as possible due to demonstrated efficacy) is not the only consideration that needs to be accounted for when designing a trial. Ensuring the trial is sufficiently powered to detect safety signals is also an important consideration. Moreover, stopping earlier might make it more difficult to evaluate efficacy in particular subgroups or geographic regions, which might be important in the post-licensure context when advisory bodies are making recommendations for use.

There is precedent for using novel designs in vaccine trials, particularly in the context of public health emergencies, such as for trials of vaccines against SARS-CoV-2 and Ebola. A trial conducted of the BNT162b2 vaccine (Pfizer-BioNTech) against SARS-CoV-2 used a group sequential Bayesian design for the phase 3 portion of the trial that led to licensure [12]. In the study protocol, the investigators provided detailed justification of the criteria for success and futility based on a variety of expected levels of vaccine efficacy and sample sizes. Similarly, a trial conducted of the Ad26.COV2.S vaccine (Janssen) used a frequentist group sequential design [13]. The original trial of the Novavax maternal RSV vaccine also had originally planned to use a group sequential design, which was abandoned following the interim analyses [2].

Group sequential trials that control frequentist type I error rates have some tradeoffs due to the need to use a more stringent cutoff value for the posterior probability. There is a possibility that the trial could be stopped after demonstrating efficacy with a relatively small number of participants. However, with a modest effect size, there is a possibility that more participants would be needed. It is possible that the number of subsequent looks at the data and the cutoffs could be adjusted after the initial look to mitigate this issue for interventions with a moderate effect size.

There has been some debate about the need to account for type I error with Bayesian trials [14]. Some have argued that the Bayesian paradigm is not compatible with the concept of having a multitude of theoretical trials over which performance is evaluated. Similarly, with a Bayesian approach, it is natural to report and make decisions based on a continuous value of the posterior probability rather than a binary hypothesis test. While these arguments are valid, FDA guidance documents clearly state that Bayesian trials need to evaluate frequentist operating characteristics (power and type I error) [15]. If the thresholds used to declare ‘success’ of the trial are not adjusted, a large proportion of trials would declare success when there is no real effect when performing group sequential trials. Our analyses use a simple adjustment to reduce

the type I error rate, using the same, more conservative threshold at each evaluation point [6]. More sophisticated adjustments that use more conservative thresholds at earlier time points could also be implemented [16], or a less stringent threshold could be used at the final evaluation point if the trial was not stopped at an interim analysis [2].

Implementing a group sequential trial of an intervention against RSV would have some challenges. RSV is typically a highly seasonal pathogen, with epidemics occurring in the winter months in most locations. Thus, enrollment periods would have to be timed to the RSV season. This could be partially mitigated by including study sites in both Northern and Southern hemisphere locations, where seasons are offset, or by enrolling individuals in locations where circulation of the virus has less marked periodicity (such as in tropical locations), where continuous enrollment would be more fruitful.

In these analyses, data from a previous phase 3 trial of a maternal RSV vaccine candidate were used to specify the prior distribution for parameters corresponding to a new phase 3 trial. However, case data from a phase 2 trial, if collected, could also be used for building a commensurate prior for a phase 3 trial. There should be some consideration given to whether the efficacy from the new trial would be expected to be similar to the efficacy in the original trial. The simulations performed here used a simple comparison of the incidence of disease in the two study arms, mirroring the analysis in the Prepare trial. These types of approaches can be readily adapted for other types of analyses, such as survival analysis and analyses based on case accrual rather than population size.

Inequities in vaccine access during the pandemic have led to initiatives to develop, produce, and evaluate vaccines in low- and middle-income countries. Efficient yet methodologically robust trial designs, such as the ones proposed here, will help reduce the cost of vaccine development and evaluation in low- and middle-income countries.

In conclusion, group sequential analyses with frequent evaluations of efficacy, potentially with the incorporation of historical data, could allow for more efficient evaluations of vaccines and increase the chances that a trial is successful compared to a trial with a single pre-specified evaluation point. Given the high burden of disease caused by RSV and the recent track record of having a failed trial, these alternative designs should be considered for future evaluations of interventions against RSV as well as other endemic diseases.

Acknowledgements. Thanks to the extensive resources of Frank Harrell's websites, which inspired some of the analyses here: https://hbiostat.org/bayes/; https://www.fharrell.com/post/bayes-seq

Funding. This work is funded by grants from the Bill and Melinda Gates Foundation and partially funded by grants from the NIH/NIAID (R01AI137093) and Yale Clinical and Translational Science Award (UL1 TR001863). This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest. DMW has received consulting fees from Pfizer, Merck, GSK, and Affinivax for work unrelated to this manuscript, and research grants from Pfizer and Merck for work unrelated to this manuscript. JLW has received consulting fees from Pfizer for work unrelated to this manuscript. VEP is a member of the WHO Immunization and Vaccine-related Implementation Research Advisory Committee (IVIR-AC). The other authors declare no conflicts.

Supplementary Figures

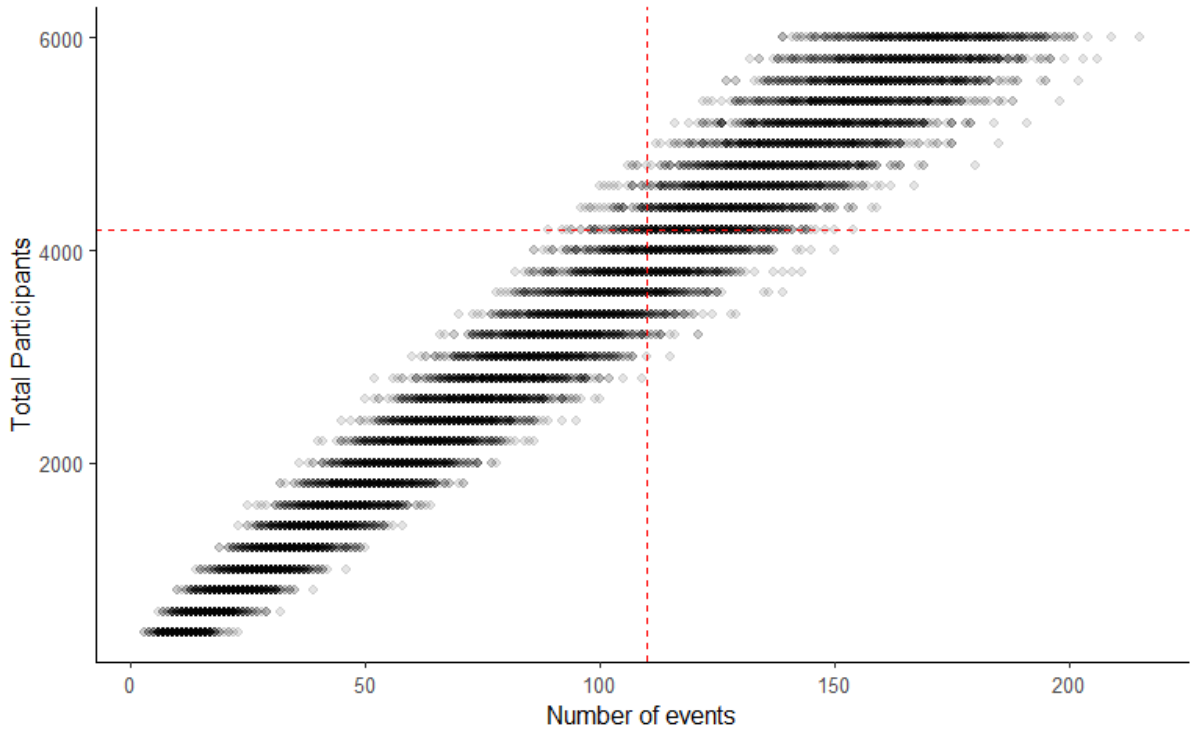

Figure S1. Relationship between average number of participants enrolled across both arms and the number of events in the trial. The red lines show the observed values used in the analysis from the Prepare trial.

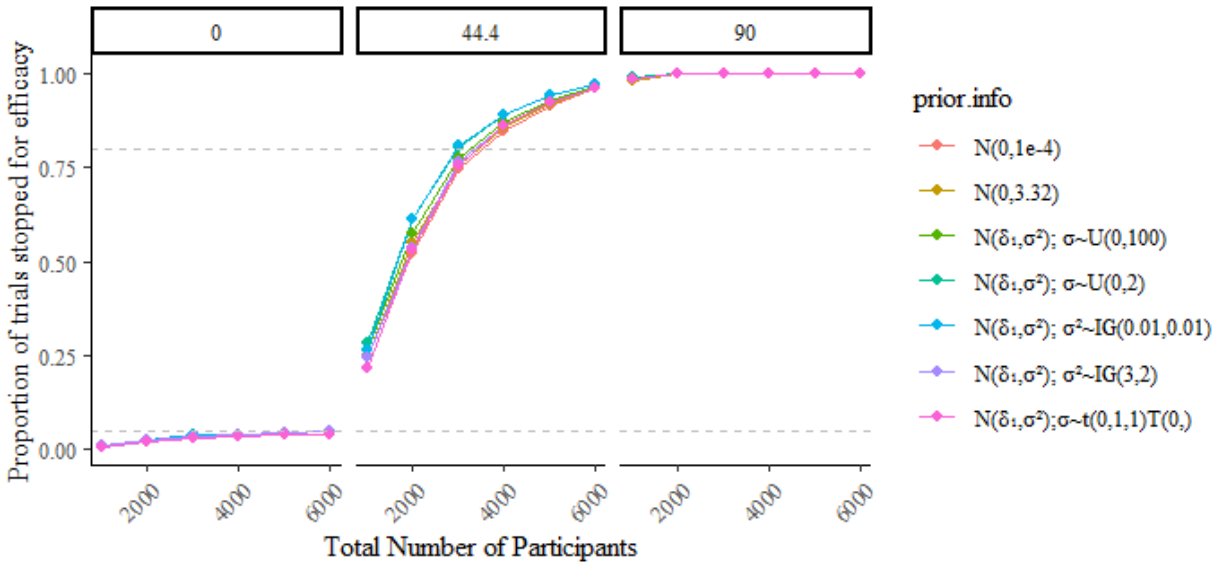

Figure S2. Cumulative proportion of trials that are declared a success (VE>0%) using a threshold adjusted to control for type I error, showing the full set of models that were evaluated. VE is evaluated after every 1000 participants complete the follow-up period. From left to right, we consider scenarios with a “true” vaccine efficacy of 0%, 44.4%, and 90%. The horizontal dashed lines are at 0.05 and 0.8.

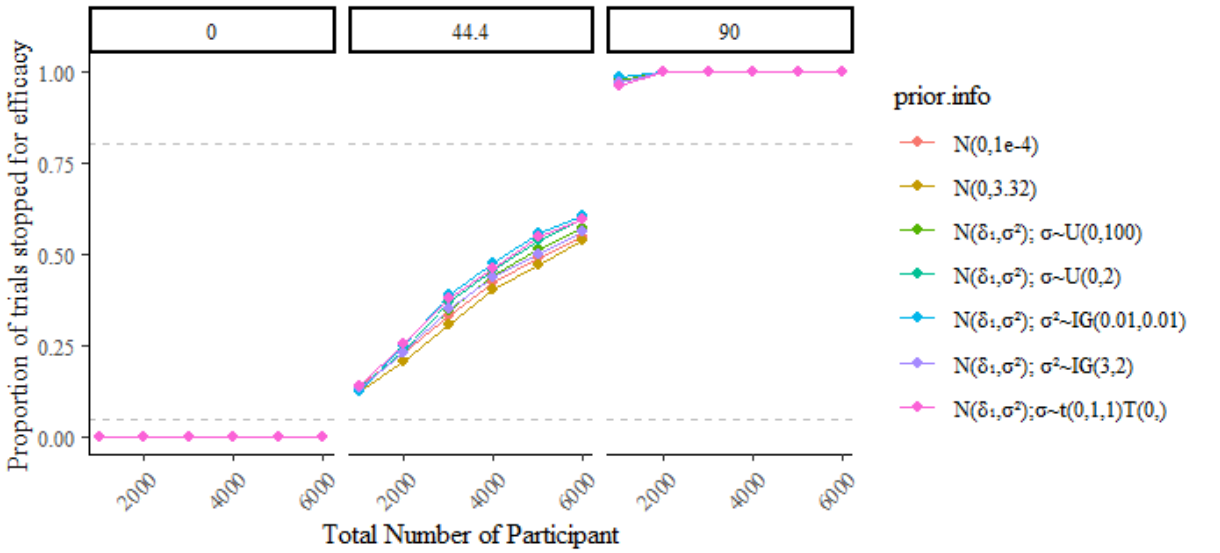

Figure S3. Cumulative proportion of trials that are declared a success, using a more stringent definition of success (VE>30%), with a threshold adjusted to control for type I error. (A) VE evaluated after every 1000 participants complete the follow up period. From left to right, we consider scenarios with a “true” vaccine efficacy of 0%, 44.4%, and 90%. The horizontal dashed lines are at 0.05 and 0.8.

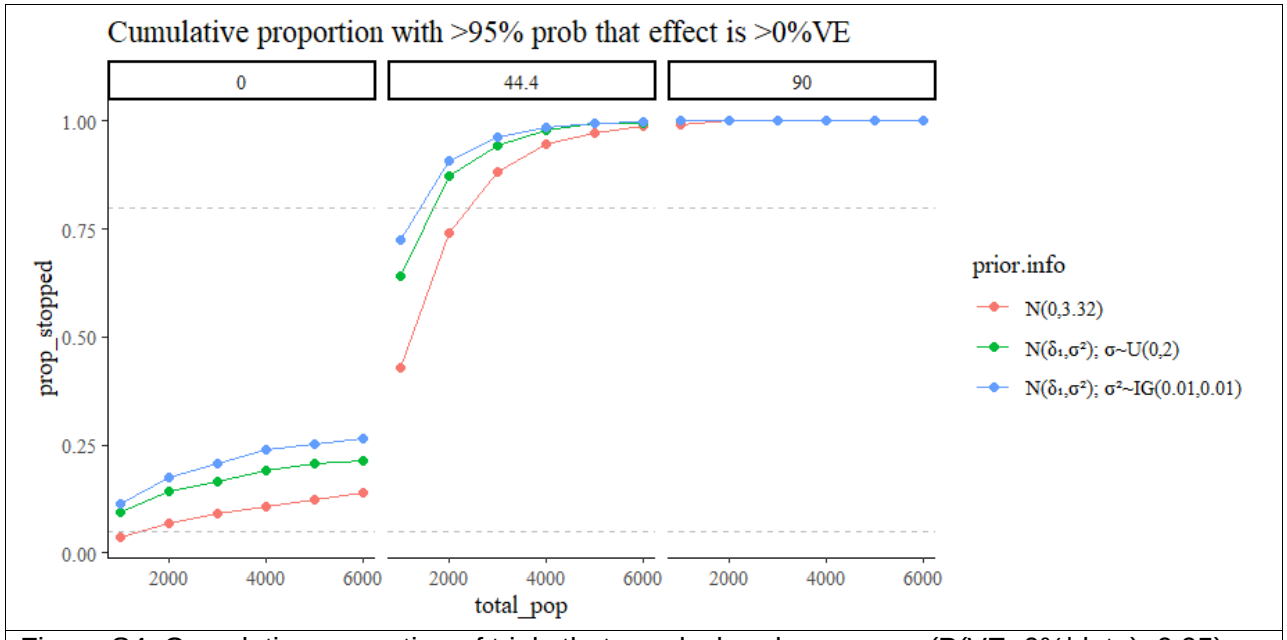

Figure S4. Cumulative proportion of trials that are declared a success (P(VE>0%|data)>0.95), with no adjustment for Type I error. VE is evaluated after every 1000 participants complete the follow-up period. From left to right, we consider scenarios with a "true" vaccine efficacy of 0%, 44.4%, and 90%. The horizontal dashed lines are at 0.05 and 0.8.

REFERENCES

1. The costs and effectiveness of large Phase III pre- licensure vaccine clinical trials.

2. Madhi SA, Polack FP, Piedra PA, Munoz FM, Trenholme AA, Simões EAF, et al. Respiratory syncytial virus vaccination during pregnancy and effects in infants. N Engl J Med. 2020;383: 426–439.

3. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390: 946–958.

4. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics. 2011;67: 1047–1056.

5. Ibrahim JG, Chen M-H, Gwon Y, Chen F. The power prior: theory and applications. Stat Med. 2015;34: 3724–3749.

6. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977;64: 191–199.

7. Harrell F, Chris Lindsell VICTR Research Methods Program Department of Biostatistics Vanderbilt University School of Medicine. Statistical design and analysis plan for sequential parallel-group RCT for COVID-19. 6 Nov 2021 [cited 13 Jan 2023]. Available: https://hbiostat.org/proj/covid19/bayesplan.html

8. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 2006;1: 515–534.

9. R Core Team (2021). R Foundation for Statistical Computing. R: A language and environment for statistical computing. 1-2021. Available: https://www.R-project.org/.

10. Plummer M. rjags: Bayesian graphical models using MCMC. R package version 3-14. 2014. Available: http://CRAN.R-project.org/package=rjags

11. Kopp-Schneider A, Calderazzo S, Wiesenfarth M. Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. Biom J. 2020;62: 361–374.

12. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N Engl J Med. 2020;383: 2603–2615.

13. A study of Ad26.COV2.S for the Prevention of SARS-CoV-2-Mediated COVID-19 in Adult Participants - Full Text View - Clinicaltrials.gov. [cited 13 Jan 2023]. Available: https://clinicaltrials.gov/ct2/show/study/NCT04505722

14. Ryan EG, Brock K, Gates S, Slade D. Do we need to adjust for interim analyses in a Bayesian adaptive trial design? BMC Med Res Methodol. 2020;20: 150.

15. Guidance for industry and FDA staff guidance for the use of Bayesian statistics in medical device clinical trials. 2010 [cited 7 Feb 2023]. Available: https://www.fda.gov/media/71512/download

16. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35: 549–556.