Site-specific biases in phase-III clinical trials underestimate the effect of radical cure against *Plasmodium vivax* hypnozoites

John H. Huber¹*, Cristian Koepfli¹, Guido España¹, Narimane Nekkab², Michael T. White², T. Alex Perkins¹

¹Department of Biological Sciences & Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, USA
²Unité Malaria: Parasites et Hôtes, Département Parasites et Insectes Vecteur, Institut Pasteur, Paris, FRA

Corresponding Author: jhuber3@nd.edu (JHH)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Plasmodium vivax relapses caused by reactivating hypnozoites are a major barrier for elimination and control of this form of malaria. Radical cure is a form of therapy capable of addressing this problem. Recent clinical trials of radical cure have yielded efficacy estimates ranging from 65% to 94%, with substantial variation across trial sites. We performed an analysis of simulated trial data using a transmission model to demonstrate that variation in efficacy estimates across trial sites can arise from differences in the conditions under which trials are conducted. Our analysis revealed that differences in transmission intensity, heterogeneous exposure, and relapse rate can yield efficacy estimates ranging as wide as 11-82%, despite simulating trial data under the uniform assumption that treatment had a 75% chance of clearing hypnozoites. We show that vector control and parasite genotyping offer two potential means to yield more standardized efficacy estimates that better reflect protection against relapse.
INTRODUCTION

Plasmodium vivax is the most geographically widespread cause of human malaria, and its burden in 2017 was estimated at 14.3 million clinical cases globally (1). Control of P. vivax is challenging due to a unique life stage of the parasite, known as the hypnozoite, which latently infects the livers of individuals with recent P. vivax blood-stage infections (2). Hypnozoites activate to cause successive relapsing infections following the initial blood-stage infection, and relapses comprise an estimated 79 – 96% of all P. vivax infections (3,4). The prevention of relapses is therefore an ongoing priority for P. vivax control (5), and clearance of hypnozoites can be achieved through radical cure treatment with an 8-aminoquinoline, such as primaquine (PQ) or tafenoquine (TFQ).

Recent phase-III clinical trials for PQ and TFQ have been conducted in Latin America, sub-Saharan Africa, and Southeast Asia (6–9). The DETECTIVE trial estimated that the recurrence-free efficacies of PQ and TFQ were 74% and 70%, respectively (6), suggesting a large potential impact of radical cure as a first-line P. vivax treatment (10). However, each trial noted substantial geographical variation in efficacy estimates. Although potentially reflective of differences in hypnozoite clearance with 8-aminoquinoline treatment among the distinct trial populations (11), the geographical variation in efficacy may have instead been attributable to features of each transmission setting, which would bias our efficacy estimates and limit the interpretability of trial results when evaluating the extent to which radical cure prevents relapse.

The primary endpoint used in recent phase-III clinical trials is freedom from P. vivax recurrent infection (6–8), which encompasses relapses caused by hypnozoite activation, reinfections caused by mosquito biting, and recrudescences caused by blood-stage therapeutic
failure (12). Only relapses caused by hypnozoite batches (i.e., a group of hypnozoites that
derives from one infectious mosquito bite) present prior to treatment directly reflect the extent to
which 8-aminoquinoline treatment prevents relapse, or the hypnozoite batch clearance
probability (hereafter, clearance probability) (Fig. 1). Because there remains no reliable way to
distinguish these relapses from all other recurrent infections using epidemiological data alone,
efficacy against recurrent infection is not equal to efficacy against relapse, and the measured
efficacy will likely vary depending upon the site in which the trial is conducted. Obtaining
standardized estimates of the effect of each 8-aminoquinoline against hypnozoite batches will
therefore depend upon the proportion of recurrent infections identified during follow-up that are
relapses caused by hypnozoite batches acquired prior to 8-aminoquinoline treatment.

An incongruence between the primary endpoint and the effect that the clinical trial
intends to measure has previously been noted as a source of bias in measuring the therapeutic
cure rates for *P. falciparum* malaria and tuberculosis (13–15). We propose that the magnitude of
this bias could depend upon features of the trial location and the trial design. Although most or
all clinical *P. vivax* patients enrolled in clinical trials are expected to carry hypnozoites, the rates
at which participants experience reinfections and relapses may vary across trial sites due to
transmission intensity, heterogeneous mosquito biting patterns (16), and the relapse phenotype of
the *P. vivax* parasite (17). The number of recorded relapses also depends upon the duration of
follow-up and the duration of prophylaxis provided by the treatment regimen used in the trial.
The extent to which these different features of the trial location and the trial design contribute to
bias in efficacy estimates—and impact our ability to measure the clearance probability—has not
been explored.
Here, we extend an existing individual-based model of *P. vivax* transmission (10,18) to simulate phase-III clinical trials for radical cure. To maintain consistency with previous trials, we modeled our phase-III clinical trials after the recent DETECTIVE trials for PQ and TFQ (6–8). By varying different features of the trial location and trial design, we sought to understand how site-specific biases in efficacy estimates may arise. We also explored different approaches, such as the use of vector control and parasite genotyping methods, that could be implemented to mitigate site-specific biases and standardize estimates of the effect of radical cure against hypnozoite batches.

**Figure 1. Schematic of potential infection outcomes during trial follow-up.** Participants are enrolled in the trial and initially harbor hypnozoite batches (colored squares). The colors represent the genetic identity of each hypnozoite batch. Hypnozoite batches are cleared (unfilled dotted squares) with 8-aminoquinoline treatment according to the per-hypnozoite batch probability of clearance. Treatment with an 8-aminoquinoline provides each participant with a period of prophylaxis during which reinfections (solid line) and relapses (dotted line) are suppressed. Following prophylaxis, trial participants can experience reinfections and relapses, and these reinfections and relapses may be detected during the remainder of follow-up. Only relapses that arise from the activation of hypnozoite batches that were present before 8-aminoquinoline treatment reflect the per-hypnozoite batch clearance probability.
RESULTS

Effect of Transmission Intensity and Heterogeneity in Biting

We first examined how transmission intensity and the level of heterogeneity in mosquito biting patterns in the trial location interact to affect efficacy estimates. We varied transmission intensity by setting the entomological inoculation rate (EIR) from 0.1 to 100 infectious bites per person-year and varied the level of heterogeneity in mosquito biting by changing the variance in the distribution of individual-level exposure to mosquito biting.

When exposure to mosquito biting was the same for all trial participants (Fig. 2B), we observed a downward bias in efficacy estimates that increased with transmission intensity (Fig. 2A). For clinical trials of an 8-aminoquinoline with a clearance probability equal to 0.75, light microscopy (LM)-detectable recurrence-free efficacies ranged from 0.80 (IQR: 0.79 - 0.82) at an EIR of 0.1 to 0.35 (0.30 - 0.40) at an EIR of 100. This downward bias in efficacy occurred because reinfection by mosquito biting, an infection event that does not reflect the clearance probability, was more frequent at higher EIR. When EIR was 100, 99% (SD: 0.22%) of participants were reinfected on average by mosquitoes during the 180-day follow-up period, compared to only 0.73% (SD: 0.18%) at an EIR of 0.1.

At low transmission intensities when reinfection rates were low, our efficacy estimates exceeded the clearance probability, revealing a positive bias. This positive bias occurred because we assumed that radical cure completely cleared hypnozoite batches in a subset of trial participants and did not clear hypnozoite batches in the remainder of trial participants. In trials measuring time to first recurrent infection, those participants that completely clear hypnozoite batches increase the total person-time at risk in the treatment arm, leading to a competing positive bias that was present in all trial simulations. This bias was eliminated if efficacy was
calculated based upon the proportion at risk using the complete record of recurrent infections (Fig. S1).

In addition to the effect of transmission intensity itself, the effect of heterogeneous biting on efficacy estimates was modulated by the transmission intensity of the trial location. At a low EIR of 0.1, efficacy estimates were insensitive to changes in the level of heterogeneity in biting (Fig. 2A), because the upper bound of trial participants reinfected during follow-up was 0.73% (SD: 0.18%). By contrast, at intermediate EIRs of 1 and 10, the downward bias was reduced with greater heterogeneity in biting. Efficacy measured at an EIR of 10 ranged from 0.58 (0.52 - 0.62) under homogeneous biting to 0.74 (0.72 - 0.76) under the highest level of heterogeneity in biting simulated. Heterogeneous biting reduced the downward bias caused by frequent reinfection, because fewer infections observed during follow-up were reinfections from mosquito biting. For example, at an EIR of 10, the proportion of individuals re-infected during follow-up ranged from 74% (SD: 0.94%) under homogeneous biting to 18% (0.78%) at the highest level of heterogeneity in biting simulated.

At an EIR of 100, efficacy estimates had a non-monotonic relationship with the extent of heterogeneous biting (Fig. 2A). At high transmission intensities, greater heterogeneity in biting contributed to greater variation in anti-parasite immunity among trial participants. Trial participants that were more frequently exposed to mosquito biting were expected to have greater levels of anti-parasite immunity, thereby reducing the detectability of each *P. vivax* infection during follow-up. Therefore, although the percentage of participants reinfected during follow-up decreased monotonically with greater heterogeneity in mosquito biting, the percentage of participants with a detectable reinfection (e.g., clinical or subclinical/LM-detectable) varied non-
monotonically and was 36% (1.0%), 38% (1.0%), 34% (1.0%), and 23% (1.0%) when the respective variance in exposure to biting was 0, 1, 2, and 3 (Fig. 2B-E).

Figure 2. Effect of transmission intensity and heterogeneity in biting on efficacy estimates. (A) LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIR) and levels of heterogeneity in biting. Each point represents the median of 200 simulations, and each bar is the interquartile range. The color represents the degree of heterogeneity in individual-level exposure to biting, corresponding to the distributions in (B-E). Darker colors indicate greater heterogeneity in individual-level exposure to biting, and the dotted line is the clearance probability. The distributions of biting propensities are shown from representative simulated trials where the variance in the logged biting propensity was equal to (B) zero, (C) one, (D) two, or (E) three.

Effect of the Rate of Relapse and Duration of Follow-up

We next examined how the rate of relapse of *P. vivax* hypnozoites and the duration of follow-up of trial participants interact to affect efficacy estimates. We varied the mean time to relapse from 30 to 180 days and simulated phase-III clinical trials in which participants were followed for 90,
To assess the effect of transmission intensity, all trials were simulated at EIRs of 0.1, 1, 10, and 100, assuming homogeneous biting.

For a given duration of follow-up, efficacy estimates decreased with a longer mean time to first relapse (Fig. 3A). In simulated trials in which participants were followed for 180 days at an EIR of 1, we estimated that the efficacy of an 8-aminoquinoline with a clearance probability of 0.75 decreased from 0.70 (IQR: 0.68 - 0.72) to 0.59 (0.56 - 0.62) as the mean time to first relapse increased from 30 to 180 days. This bias toward the null occurred because, as the mean time to first relapse increased, fewer trial participants were expected to relapse within the duration of follow-up (Fig. 3B), causing the distribution of times to first recurrent infection to appear more similar across trial arms. Only 63.2% of participants were expected to relapse by 180 days when the mean time to first relapse was 180 days, compared to 99.8% of participants when the time to first relapse was 30 days.

For a given mean time to first relapse, our efficacy estimates decreased with a longer duration of follow-up (Fig. 3A). With an EIR equal to 1 and a mean time to first relapse of 30 days, we estimated that the efficacy of an 8-aminoquinoline was 0.73 (0.70 - 0.75) when trial participants were followed for 90 days, compared to 0.54 (0.52 - 0.56) when trial participants were followed for 730 days. Under an alternative scenario in which the EIR was 10, the estimated efficacies were 0.70 (0.65 - 0.74) and 0.39 (0.37 - 0.43) at the respective durations of follow-up. The downward bias occurred because a longer duration of follow-up ensured that more trial participants were reinfected by mosquito biting (Fig. 3C), similarly causing the distribution of times to first recurrent infection to appear more similar across the trial arms. By 90 days of follow-up, 12% and 71% of trial participants were expected to have been reinfected.
when the respective EIRs were 1 and 10. By 730 days of follow-up, 63% and 100% of trial participants were now expected to have been reinfected at respective EIRs of 1 and 10.

**Figure 3. Effect of rate of relapse and duration of follow-up on efficacy estimates.** (A) LM-detectable recurrence-free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIRs), mean times to relapse, and durations of follow-up. Each point is the median of 200 simulations, and each bar is the interquartile range. The color represents the mean time to relapse, and each vertical strip for each EIR is the duration of follow-up (90, 180, 365, or 730 days). The dotted line is the clearance probability simulated in each trial. (B) The percentage of trial participants expected to relapse as a function of the duration of follow-up is shown for different mean times to relapse. (C) The percentage of trial participants expected to be reinfected as a function of the duration of follow-up is shown for different EIRs.
Effect of Vector Control

Higher transmission intensity biased our efficacy estimates downward, because trial participants in both the treatment and control arms were frequently reinfected (Fig. 2A). We assessed the potential of long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) to protect trial participants from reinfection and therefore reduce the downward bias due to transmission intensity. Because the effects of LLINs and IRS depend upon vector bionomics, we varied the proportion of bites that occurred indoors and the proportion of bites that occurred while trial participants were in bed.

Vector control interventions were effective in reducing the downward bias due to transmission intensity in trial locations where the vector was predominantly endophagic (i.e., feeds indoors) (Fig. 4). With an endophagic vector with no preference in biting time and at an EIR of 10, the estimated efficacies improved from 0.52 (0.46 - 0.57) in the absence of vector control to 0.57 (0.53 - 0.62) with LLINs, 0.61 (0.56 - 0.66) with IRS, and 0.65 (0.60 - 0.69) with the combination of LLINs and IRS (Fig. 4A). By comparison, with an exophagic vector with no preference in biting time and at an EIR of 10, efficacy estimates improved only slightly from 0.52 (0.48 - 0.57) in the absence of vector control and 0.52 (0.47 - 0.57) with LLINs to 0.53 (0.49 - 0.58) with IRS and 0.53 (0.48 - 0.58) with the combination of LLINs and IRS (Fig. S2D).

Across all mosquito biting behaviors simulated, the combination of LLINs and IRS was generally more effective than either intervention in isolation in reducing the downward bias in efficacy estimates (Fig. S2). When considering interventions in isolation, IRS was more effective than LLINs in reducing the downward bias, and the distribution of LLINs most improved efficacy estimates under scenarios in which an appreciable number of mosquito bites were taken at night while trial participants were in bed.
Figure 4. Effect of vector control on efficacy estimates. The impact of LLIN distribution (red), IRS administration (green), and combined LLIN distribution and IRS administration (teal) on resolving biases in LM-detectable recurrence free efficacy estimates is compared to a no-intervention scenario (dark blue) across a range of entomological inoculation rates (EIR). Each point represents the median of 200 simulations, and each bar is the interquartile range. The dotted line is the clearance probability simulated in each trial. The absolute proportion of bites occurring indoors ($\Phi_I$) was 0.90 and in bed ($\Phi_B$) was 0.45.

Effect of Parasite Genotyping

Unlike vector control interventions that reduce the downward bias in efficacy by reducing the number of reinfection events among trial participants (Fig. S3), parasite genotyping could correct or reduce the downward bias by distinguishing recurrent infection events that directly reflect the clearance probability (i.e., relapses arising from hypnozoite batches acquired prior to treatment) from all other recurrent infections. We assessed the potential of a generic genotyping method to correct the bias in efficacy estimates at different transmission intensities and considered how performance characteristics of the genotyping method affected the extent to which the bias was corrected by varying the sensitivity and specificity of the genotyping method. For the purpose of
In this simulation study, we defined sensitivity as the probability of correctly identifying relapses caused by hypnozoite batches acquired prior to treatment and specificity as the probability of correctly identifying all other recurrent infections (e.g., reinfections and relapses caused by hypnozoite batches acquired after treatment).

At lower and intermediate transmission intensities (i.e., EIRs of 0.1 and 1), efficacy estimates changed more with improved sensitivity than improved specificity of the genotyping method. By contrast, at higher transmission intensities, a genotyping method with high specificity was needed to reduce the downward bias in efficacy. At an EIR of 100 and assuming 100% sensitivity of the genotyping method, the estimated efficacy of an 8-aminoquinoline with clearance probability equal to 0.75 improved from 0.38 (0.33 - 0.44) at 25% specificity to 0.67 (0.62 - 0.71) at 100% specificity. Under the alternative scenario in which specificity of the genotyping method was 100%, the estimated efficacy at the respective EIR improved from 0.48 (0.35 - 0.58) at 25% sensitivity to 0.67 (0.62 - 0.71) at 100% sensitivity. The specificity of the genotyping method was more important at higher EIR, because, at an EIR of 100, 100% of participants were expected to be reinfected during follow-up (Fig. 3C). Misclassifying reinfection events as failures of 8-aminoquinoline treatment (i.e., relapses caused by hypnozoites batches acquired prior to treatment) lead to a greater downward bias, so a highly specific method was needed in settings where trial participants were frequently reinfected. These effects were less pronounced at low transmission intensities, given that fewer trials participants were reinfected during follow-up.
Figure 5. Effect of parasite genotyping on efficacy estimates. The impact of genotyping recurrent infections to estimate the efficacy of radical cure is shown for different sensitivities and specificities of the genotyping method across a range of entomological inoculation rates (EIR).

Sensitivity of the genotyping method is the probability of correctly identifying a relapse caused by hypnozoites acquired prior to treatment, and specificity of the genotyping method is the probability of correctly identifying all other recurrent infections. For a given EIR, the vertical panels correspond to the specificity of the genotyping method, and the colors of the points within a given panel denote the sensitivity of the genotyping method. Each point represents the median of 200 simulations, and each bar is the interquartile range. The dotted line is the clearance probability simulated in each trial.

Effect of the 8-aminoquinoline

The previous analyses considered phase-III trials of an 8-aminoquinoline that provided prophylaxis for 28 days, a duration of time less than the 32-day period of left-censoring used to calculate efficacy. Consequently, prophylaxis did not impact our efficacy estimates. To examine whether a longer duration of prophylaxis biased our efficacy estimates, we simulated phase-III trials for two different 8-aminoquinolines, PQ and TFQ, that provide prophylactic effects for fixed periods of 28 and 45 days, respectively. Because the benefit of prophylaxis increases with transmission intensity, we simulated trials at EIRs equal to 0.1, 1, 10, and 100, assuming homogeneous biting.
A longer duration of prophylaxis biased our efficacy estimates upward, though the magnitude of the bias depended upon transmission intensity (Fig. 6). In a low-transmission setting (i.e., EIR of 0.1), our efficacy estimates did not change with an increased duration of prophylaxis, because, on average, only 0.058% (SD: 0.081%) of participants in the control arm were reinfected between days 32 and 45 post-enrollment—the time period over which TFQ provided prophylactic effects not accounted for by left-censoring. Consequently, at an EIR of 0.1, we estimated that the efficacies of TFQ and PQ were both 0.80 (0.79 - 0.81). At higher EIRs, we estimated a greater efficacy for TFQ than PQ, because participants in the control arm were frequently reinfected whereas participants in the treatment arm were protected from reinfection. On average, the percentage of participants in the control arm who were reinfected between 32- and 45-days post-enrollment was 1.5% (SD: 0.38%), 9.4% (SD: 1.0%), and 52% (2.2%) at EIRs of 1, 10 and 100, respectively. Consequently, we estimated that the respective efficacies of TFQ and PQ were 0.69 (0.66 - 0.72) and 0.67 (0.65 - 0.70) at an EIR of 1, 0.60 (0.55 - 0.64) and 0.56 (0.52 - 0.61) at an EIR of 10, and 0.42 (0.38 - 0.47) and 0.35 (0.30 - 0.40) at an EIR of 100.
Figure 6. Effect of 8-aminoquinoline on efficacy estimates. LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown for primaquine (purple) and tafenoquine (blue) at different entomological inoculation rates (EIR). Each point is the median of 200 simulations, and each bar is the interquartile range. The dotted line is the clearance probability simulated in each trial.

Effect of the efficacy metric and infection endpoint

To test whether the efficacy estimates that we obtained were sensitive to the choice of efficacy metric and the infection endpoint, we simulated phase-III clinical trials in which trial participants were assessed for PCR-detectable, LM-detectable, or clinical infections during follow-up. Trials were simulated at EIRs of 0.1, 1, 10, and 100 assuming homogeneous biting, and efficacy was calculated using three metrics: (1) Cox proportional hazards model, (2) incidence rates, and (3) the proportion at risk.

For all three efficacy metrics considered, increasing transmission intensity caused a downward bias in efficacy estimates, because, at higher transmission intensities, most
participants in both the treatment and control arms experienced a reinfection or relapse during follow-up (Fig. 7). For PCR-detectable infections, estimated efficacy was highest if calculated using the Cox proportional hazards model and lowest if calculated using the proportion at risk. This suggested that, in the absence of parasite genotyping (Fig. S1), the magnitude of bias caused by reinfection from mosquito biting was more severe when basing efficacy upon incidence rates or the proportion at risk than upon proportional hazards.

The sensitivity of the assay by which trial participants were assessed for recurrent infections also affected our efficacy estimates, though the direction of the effect depended upon the chosen efficacy metric. For efficacy estimates based upon incidence rates or the proportion at risk, shifting from a more sensitive assay (i.e., PCR-based detection) to a less sensitive assay (i.e., monitoring for clinical infections) reduced the downward bias due to reinfection, particularly at higher transmission intensities. This was due to the fact that the fraction of recurrent infections ascertained increased with a more sensitive assay, causing us to detect more recurrent infections during follow-up and making the incidence rates and proportions at risk appear more similar across the treatment and control arms. By contrast, for efficacy estimates based upon the Cox proportional hazards model, efficacy estimates were lower for a less sensitive assay at EIRs of 0.1, 1, and 10. A less sensitive assay resulted in the time to the first detected recurrent infection occurring later during follow-up for participants in both the treatment and control arm, causing the observed downward bias. At a very high transmission intensity (i.e., EIR of 100), a less sensitive assay minimized bias, because effectively all trial participants were expected to be reinfected during follow-up (Fig. 2C).
**Figure 7. Effect of efficacy metric and infection endpoint on efficacy estimates.** Efficacy estimates obtained from simulated clinical trials is shown when calculated using the Cox proportional hazards model, incidence rates, or the proportion at risk. Efficacy calculations were performed using clinical (maroon), LM-detectable (red), and PCR-detectable (orange) recurrent infections identified during follow-up as the infection endpoint. Each point is the median of 200 simulations, and each bar is the interquartile range. The dotted line is the clearance probability simulated in each trial.

**DISCUSSION**

Obtaining standardized estimates of the effect of radical cure on *P. vivax* hypnozoites is challenging, because the recurrent infection endpoint used in phase-III clinical trials includes infection events, such as reinfections by mosquito biting, that do not reflect the effect of radical cure on *P. vivax* hypnozoites. In this simulation study, we identified features of the trial setting, including transmission intensity, heterogeneous feeding patterns, and the relapse rate of the *P. vivax* parasite, that affected our estimates of efficacy and the utility of clinical trial data to assess the extent to which radical cure prevents relapse. We demonstrated that the use of vector control...
and genotyping methods are two approaches that can reduce and, in some cases, even correct these site-specific biases and yield more standardized estimates of efficacy against relapse.

In the recent GATHER, DETECTIVE, and IMPROV trials (6–9), efficacy estimates varied across trial sites. Interpreting site-specific differences in efficacy is important for past and future clinical trials (19), and our simulation results suggest that the differences in efficacy estimates could be caused by site-specific biases that arise when the infection endpoint used to measure efficacy does not directly reflect the action of the therapeutic being trialed. If the trial sites vary in transmission intensity, our results suggest that efficacy estimates could be lower in high-transmission settings, because trial participants are more frequently reinfected during follow-up. Greater heterogeneity in mosquito biting reduces this downward bias by reducing the number of trial participants who are bitten and possibly reinfected during follow-up, though this effect is non-monotonic at very high transmission intensities due to the predicted interaction with anti-parasite immunity. Trial sites may also differ in the rates at which P. vivax hypnozoites relapse (17). A longer time to relapse implies that an appreciable number of trial participants will not yet have relapsed during follow-up, thereby biasing efficacy estimates downward (20). Increasing the duration of follow-up further compounds the bias rather than corrects it, because a long duration of follow-up causes more participants in both the treatment and control arms to become reinfected by mosquito biting, particularly at higher transmission intensities. Finally, frequent reinfection at higher transmission intensities makes the comparison of treatment regimens with different durations of prophylaxis challenging, because a longer duration of prophylaxis prevents more individuals in the treatment arm from reinfection. This results in an upward bias that may cause us to overestimate the effect against hypnozoites of treatment...
regimens that provide a longer period of prophylaxis, such as tafenoquine co-administered with chloroquine.

In general, there was greater bias in high-transmission settings, because of the higher risk of recurrent infections being due to reinfections from new mosquito bites. In low-transmission settings, this bias is reduced, although there is the concomitant challenge of recruiting sufficient trial participants. Our simulation results suggest that trial investigators can reduce bias by prospectively preventing reinfections with vector control or by retrospectively accounting for reinfections by parasite genotyping. LLINs and IRS were predicted to be most effective in trial settings with an endophagic vector that takes most bites indoors, so novel vector control interventions, such as spatial repellents, may be needed in settings with an exophagic vector to reduce peri-domestic biting (21). Entomological data collected prior to trial enrollment could characterize the local vector bionomics and inform the selection of interventions (22,23), and blood meal analysis (24–26) or serological assays (27,28) performed as part of the trial could quantify heterogeneous feeding, a factor that directly affects the magnitude of the site-specific bias.

Our results demonstrate that parasite genotyping could be used to generate less biased efficacy estimates that better reflect the effect of the 8-aminoquinoline against hypnozoite batches. Although in practice the sensitivity of a genotyping method may be reduced by not observing infections that occurred prior to trial enrollment, the extent to which genotyping reduced bias in our analysis was robust to changes in sensitivity and depended most upon the specificity. High specificity can likely be achieved given the high expected heterozygosity of microsatellite and amplicon deep sequencing panels, as well as the ability to detect minority clones in multiclonal *P. vivax* infections (29–33). Statistical frameworks leveraging genotyping
data to distinguish relapses from reinfections have been successfully developed and could be readily integrated into the analysis of future clinical trial data (12,33).

Beyond the factors that make resolving differences in efficacy challenging within the context of a single trial, we identified features of trial design that could limit the comparability of efficacy estimates across trials. Specifically, the details of how the efficacy endpoint was calculated (e.g., time to first infection vs. incidence ratio) and the sensitivity of the assay used to detect recurrent infections resulted in considerable variation in efficacy estimates, particularly at higher transmission intensities. These results suggest that future studies performing meta-analyses of the efficacy of radical cure should consider differences in the trial designs of the clinical trials included.

There are a number of limitations of this analysis. First, we did not calibrate our model to clinical trial data, so our results are not representative of any specific trial settings. However, our simulations encompassed a wide range of epidemiological settings, so the site-specific biases identified in our analysis should reflect variation possible across trial sites. Second, there remains much about hypnozoite biology that is not well understood (34), so our simplified representation of hypnozoite activation and death in our transmission model may not fully capture reality. Finally, we have an imperfect understanding of the acquisition and loss of anti-parasite and clinical immunity (35). Underestimating the level of anti-parasite and clinical immunity among trial participants may lead to a greater fraction of recurrent infections detected during follow-up than would occur in an actual trial context. This could have led us to overestimate site-specific biases, in which case more refined mathematical representations of immunity could be beneficial for future studies building on ours (10,36).
Our analysis predicted that site-specific biases are likely to occur in phase-III trials for radical cure, and these results suggest that care should be taken in the planning of future trials and the interpretation of trial data. Mathematical modeling that accounts for site-specific biases can aid in the interpretation of clinical trial data and may be useful for informing the choice of efficacy metric and identifying future trial sites where biases may be less severe (37). As always, the utility of the insights from mathematical models is improved with additional data, so modeling should be integrated into clinical trial design to identify data needs and inform data collection in order to reduce these biases and improve understanding of radical cure’s potential to control *P. vivax* malaria.

**METHODS**

**Transmission Model**

To simulate *Plasmodium vivax* transmission, we used a stochastic, individual-based model developed by White *et al.* (10) and Nekkab *et al.* (18). This model extends the Ross-MacDonald framework of *P. falciparum* transmission to incorporate relapses (38), a characteristic feature of *P. vivax* transmission (39). Calibrated to epidemiological surveys conducted in Papua New Guinea, the Solomon Islands, and Brazil, the model by White *et al.* (10) and Nekkab *et al.* (18) reproduces *P. vivax* transmission dynamics across a range of epidemiological settings.

The rates at which individuals in the population are reinfected and relapse depend upon the respective number of infectious mosquito bites received (i.e., the entomological inoculation rate, or EIR) and the number of hypnozoite batches present in each individual’s liver. The average EIR within the population is determined by the mosquito population dynamics, and individual variation therein is due to heterogeneity in exposure and age-dependent differences in
biting. Following successful inoculation with *P. vivax* sporozoites from an infectious mosquito bite, an individual experiences a primary bloodstream infection and accumulates an additional batch of hypnozoites within the liver. In the absence of treatment, each batch of hypnozoites can activate to cause relapses or be cleared naturally from the liver according to fixed rates (Table S1).

The nature of each blood-stage infection (e.g., reinfection or relapse) is determined by the level of anti-parasite and clinical immunity in the infected individual. In the model, the level of anti-parasite immunity determines the probability that a blood-stage infection is detectable by light microscopy (LM) and also increases the rate at which low-density infections are cleared. The level of clinical immunity determines the probability that an individual with a blood-stage infection exhibits symptoms.

The levels of anti-parasite and clinical immunity interact to determine the detectability of each blood-stage *P. vivax* infection. The model by White *et al.* (10) and Nekkab *et al.* (18) considers three types of blood-stage infections: (1) sub-microscopic infections detectable by PCR; (2) subclinical infections detectable by LM and PCR; and (3) clinical infections detectable by LM and PCR. Following White *et al.* (10) and Nekkab *et al.* (18), we assumed that each clinical infection was characterized by a fever exceeding 38°C within the last 48 hours and a parasite density greater than 500/µL. Individuals with clinical *P. vivax* infections then seek treatment with anti-malarial drugs according to a probability of treatment-seeking behavior.

Simulation of the *P. vivax* transmission model occurs in two steps. First, the population is initialized at equilibrium according to the analogous set of deterministic compartmental differential equations. Then, the stochastic, individual-based model is simulated with the
initialized population. For further description of the model and its assumptions, please refer to Tables S1-S3 and the documentation provided in White et al. (10) and Nekkab et al. (18).

**Trial Overview**

We constructed phase-III clinical trials for a generic 8-aminoquinoline (e.g., primaquine or tafenoquine). We simulated trials in order to quantify the recurrence-free efficacy as a measure of the therapeutic effect of radical cure against the hypnozoite batches present in each individual presenting with a clinical *P. vivax* infection confirmed by light microscopy.

We designed our simulated phase-III clinical trials to be comparable to previous trials of PQ and TFQ (6–8). Participants were randomly assigned to the treatment and control arms, and all participants received chloroquine (CQ) upon enrollment to clear the *P. vivax* blood-stage infection. Consistent with the DETECTIVE trial (6,8), participants were followed for 180 days and assessed for asexual parasitemia by light microscopy at days 8, 15, 22, 29, 60, 90, 120, and 180 post-enrollment. For trials simulated with longer duration of follow-up, additional time points were included. See supplement for further details.

**Radical Cure Model**

We modeled the action of radical cure in each individual receiving treatment as the per-hypnozoite batch probability of clearance. To allow for heterogeneous action of the intervention (40), we assumed that a proportion $p_i$ of the population belonged to each stratum $i$ with per-hypnozoite batch clearance probability, $c_i$. For a population consisting of two strata, the mean per-hypnozoite batch probability of clearance was
\( \bar{c} = p_1 c_1 + p_2 c_2 \), \hspace{1em} (1)

where \( p_2 = 1 - p_1 \). This model of heterogeneous action of radical cure generalized the intervention actions commonly considered in the evaluation of clinical trials. Under the “leaky” model (40–42), the action of radical cure is the same across all individuals in the population \( (p_1 = 1) \) with \( 0 \leq c_1 \leq 1 \). Under the “all-or-none” model (40–42), radical cure completely clears hypnozoites in a subset of the population \( (0 < p_1 < 1; c_1 = 1) \) and has no effect on the hypnozoites in the remainder of the population \( (p_2 = 1 - p_1; c_2 = 0) \). These two models of intervention action could manifest due to host-specific factors, most notably the inherited blood disorder glucose-6-phosphate dehydrogenase deficiency (G6PDd) and cytochrome P-450 isoenzyme 2D6 (CYP2D6) genotype, an enzyme involved in metabolizing PQ. Rather than directly affect PQ or TFQ efficacy, G6PDd requires individuals to be excluded from clinical trials due to the risk of hemolysis. For individuals with a low CYP2D6 metabolism phenotype, it has been shown that PQ does not prevent \( P. \) vivax relapses (11). Despite advances in the understanding of CYP2D6 metabolism (43), its association with 8-aminoquinoline efficacy remains poorly understood. For example, under an “all-or-none” action, CYP2D6 metabolization could manifest as a binary phenotype with high metabolizers effectively clearing hypnozoite batches and low metabolizers failure to clear hypnozoite batches. Alternatively, under a “leaky” action, all treated individuals partially clear hypnozoite batches. Supplementary analyses revealed that the choice of intervention action did not substantially affect our efficacy estimates generated using the Cox proportional hazards model (Fig. S4). Therefore, we assumed by default that each 8-aminoquinoline had an “all-or-none” action.
Under the model of heterogeneous action of radical cure, the number of batches of hypnozoites following treatment was distributed as

\[ H_{post} \sim \text{Binomial}(H_{pre}, 1 - c_i), \forall i \in \{1,2\}. \]  

(2)

In eq. (2), \( H_{pre} \) is the number of batches of hypnozoite present in the liver prior to treatment, and \( c_i \) is the per-hypnozoite batch clearance probability for an individual in stratum \( i \).

**Trial Design**

We constructed our simulated phase-III clinical trials for radical cure to match the trial design used in past trials for PQ and TFQ (6–8). Accordingly, we considered four phases of the randomized control trial design: (1) recruitment, (2) treatment, (3) vector control, and (3) follow-up.

**Recruitment**

We enrolled individuals in the simulated clinical trial if they presented at a health clinic with febrile illness and were diagnosed by light microscopy with a \( P. \) vivax infection. Consistent with the DETECTIVE trial (6,8), participants were enrolled if they were at least 16 years of age and had a measured glucose-6-phosphate dehydrogenase (G6DP) activity greater than 70% of the median value of the trial location. Following Nekkab et al. (18), we assumed that G6PD activity was measured using the SD Biosensor STANDARD G6PD test (44). This test qualitatively classifies G6PD activity as normal (>70% activity), intermediate (30-70% activity), and low (<30% activity).
Individuals that met the criteria for enrollment were then randomly assigned to the treatment or control arm of the clinical trial. Allocation of participants to each trial arm occurred with equal probability, provided that the current number of participants in each arm was less than the desired sample size of 1,000.

Treatment

Upon enrollment, all trial participants were treated with a 3-day course of CQ to clear the blood-stage, asexual parasites (6,8). We assumed that treatment with CQ was 100% effective and that there were no recrudescences among trial participants. Participants in the treatment arm were also treated with an 8-aminoquinoline to clear the hypnozoite reservoir. For individuals in the treatment arm, the number of hypnozoite batches following radical cure was calculated according to eq. (2). We further assumed that treatment with the 8-aminoquinoline provided both blood-stage and liver-stage prophylaxis. Consistent with White et al. (10) and Nekkab et al. (18), we assumed that the duration of prophylaxis was 28 days for PQ when co-administered with CQ and 45 days for TFQ when co-administered with CQ.

Vector Control

As in the DETECTIVE trial (6,8), all participants were provided with a long-lasting insecticidal net (LLIN) to prevent reinfection from mosquitoes during follow-up and therefore decrease the potential bias that may arise in calculating efficacy based on a recurrent infection endpoint. Following Griffin et al. (36) and White et al. (10), we assumed that vector control measures decrease the probabilities of mosquito biting and successful feeding and increase the probabilities of repellency and death. The magnitude of the effect depends upon the proportion
of mosquito bites that occur while individuals are indoors ($\Phi_I$) and in bed ($\Phi_B$), the duration of usage of the LLIN, and the duration of insecticidal activity. We assumed that the mean duration of usage of LLINs was 3 years and that the mean duration of insecticidal activity was 2.5 years.

We also considered whether the use of indoor residual spraying (IRS) in each participant’s house administered in isolation or in combination with LLINs decreased the potential bias in our efficacy estimates. As with LLINs, we assumed that the use of IRS decreases the probabilities of mosquito biting and successful feeding and increases the probabilities of repellency and death (10,36). To account for the waning effect of IRS over time, we assumed that the mean duration of insecticidal activity was 6 months.

**Follow-up**

We followed each participant for 180 days following enrollment in the clinical trial and treatment with anti-malarial drugs. Consistent with the DETECTIVE trial (6,8), for each participant, we recorded the date of each clinical *P. vivax* infection that occurred within the duration of follow-up. Furthermore, we tested each participant for *P. vivax* asexual parasites using light microscopy on days 8, 15, 22, 29, 60, 90, 120, 160, and 180 post-enrollment.

To examine potential biases that arise from trial surveillance methods, we kept a complete record of all recurrent infections that occurred within the duration of follow-up, including those that would not have been detected under the trial protocol. For each recurrent infection, we recorded the cause (i.e., reinfection or relapse) and the type of blood-stage infection (e.g., sub-microscopic, sub-clinical, or clinical). We further distinguished between relapses caused by hypnozoite batches acquired prior to treatment and relapses caused by hypnozoites
batches acquired following treatment. Only relapses caused by hypnozoite batches acquired prior
to treatment reflect the action of 8-aminoquinoline treatment against hypnozoites batches.

**Statistical Analyses**

We calculated the efficacy of the 8-aminoquinoline used in each clinical trial from the output
collected in each respective simulation. We calculated efficacy using multiple metrics in order to
examine how different data collected during trial follow-up resolved biases in the efficacy
estimates.

Consistent with previous phase-III clinical trials (6,8), we used freedom from LM-
detectable recurrent infection as the default efficacy metric in our simulation studies. We
calculated efficacy using the Cox proportional hazards model (45), which computes the hazard
ratio between the treatment and control arms based on the times to first LM-detectable recurrent
infection. Following the DETECTIVE trial protocol (6,8), recurrent infections that occurred
before 32 days post-enrollment were not included. Given that the timing of each recurrent
infection was unobserved, all trial participants were interval censored.

We calculated efficacy using incidence rates computed from the recurrent infections that
occurred within the duration of follow-up. We calculated efficacy based on incidence rates as

\[
E_{IR} = 1 - \frac{e_t / y_t}{e_c / y_c}
\]  

(3)
In eq. (3), \( e_t \) and \( e_c \) are the respective number of infection events in the treatment and control arms, and \( y_t \) and \( y_c \) are the respective number of person-years of follow-up in the treatment and control arms (42,45).

The magnitude of the measured efficacy also depends upon the action of the radical cure therapeutic. In absence of other sources of bias, we will overestimate efficacy for 8-aminoquinolines that elicit an “all-or-none” response if we calculate efficacy based on incidence rates. This phenomenon is caused by the subset of individuals in the treatment arm who completely clear the hypnozoites from their livers and increase the number of person-years of follow-up (42). To resolve this bias that arises from the assumption of intervention action, we calculated efficacy based on risk as

\[
E_R = 1 - \frac{p_t}{p_c}, \quad (4)
\]

where \( p_t \) and \( p_c \) are the proportion of individuals in the treatment and control arms that experience a particular infection event within the duration of follow-up.

**Simulation Experiments**

We performed simulation experiments to quantify biases in the efficacy estimates that may arise under different transmission settings and other features of the clinical trial design. First, we considered how our efficacy estimates varied with transmission intensity and heterogeneity in mosquito biting patterns. Second, we considered how differences in the rate of relapse of the \( P. vivax \) hypnozoites interact with the duration of the trial follow-up and the transmission intensity to bias our estimates of efficacy. After characterizing these biases, we assessed whether the
allocation of vector control measures, such as LLINs and IRS, to trial participants could prevent reinfections and thereby reduce the bias in the efficacy estimates. We next considered whether a genotyping method capable of distinguishing between different types of *P. vivax* recurrent infections at different sensitivities and specificities could reduce or correct the bias in the efficacy estimates. Then, we measured the extent to which the duration of prophylaxis provided by the treatment regimen biases our estimates of efficacy by preventing reinfections in the treatment arm only. Finally, we considered how the efficacy estimates that we obtained varied with the choice of efficacy metric and infection endpoint.

For each simulation setting, we simulated a phase-III clinical trial with 1,000 participants in each arm. This sample size was chosen to ensure that the power of our clinical trials exceeded 95% for each transmission intensity and efficacy metric considered (Fig. S5). Each trial was simulated in a population of 200,000 individuals, and the maximum duration of each trial was 15 years in order to ensure that complete enrollment was attained, even at low transmission intensities. By default, PQ was the 8-aminoquinoline provided to the participants in the treatment arm and when co-administered with CQ provided 28 days of blood-stage and liver-stage prophylaxis during which trial participants could not become reinfected or relapse. The mean per-hypnozoite batch probability of clearance was equal to 0.75. For each simulation setting, we simulated trials in which the therapeutic had an “all-or-none” response. Because the transmission model was stochastic, we ran 200 simulations for each simulation setting and computed the mean and interquartile range for the efficacy estimates across these simulations. By default, we defined efficacy as freedom from LM-detectable recurrent infection calculated using the Cox proportional hazards model.
**Table 1. Parameters Varied during the Simulation Analyses.** The parameters that were varied during each simulation analysis are reported. All other parameters in the transmission model are set to default values and are consistent with the values reported in Tables S1 and S2.

<table>
<thead>
<tr>
<th>Effect of Transmission Intensity and Heterogeneity in Biting</th>
<th>Time to Relapse (d)</th>
<th>Follow-up (d)</th>
<th>Prophylaxis (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in Mosquito Biting</td>
<td>65</td>
<td>180</td>
<td>28</td>
</tr>
<tr>
<td>Effect of Rate of Relapse and Duration of Follow-up</td>
<td>0</td>
<td>30, 60, 90, 180</td>
<td>90, 180, 365, 730</td>
</tr>
<tr>
<td>Effect of Vector Control / Genotyping / Efficacy Metric and Infection Endpoint</td>
<td>0</td>
<td>65</td>
<td>180</td>
</tr>
<tr>
<td>Effect of the 8-aminoquinoline</td>
<td>0</td>
<td>65</td>
<td>180</td>
</tr>
</tbody>
</table>

**Availability of Code**

All code to reproduce the analyses in this study is available on Github at https://github.com/johnhhuber/Radical_Cure_Uncertainty.

**ACKNOWLEDGMENTS**

JHH acknowledges funding from a National Science Foundation Graduate Research Fellowship and a Richard and Peggy Notebaert Premier Fellowship from the University of Notre Dame. CK received funding from NIH R21-AI137891. NN and MTW were supported by a grant from Medicine for Malaria Venture. NN also received funding from a Pasteur Roux-Cantarini Fellowship. This publication was made possible with partial support from Grant Numbers TL1 TR002531 and UL1TR002529 (A. Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award.

**COMPETING INTERESTS**

The authors have no competing interests to declare.
**Methods**

**Transmission Model**

*Table S1. Default Transmission Setting Parameters for the Human Component of the Transmission Model.* A description of the default parameters used in the individual-based model to simulate clinical trial data are provided along with the numeric values and the corresponding references.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Age of Population</td>
<td>31.8 years</td>
<td>Assumed</td>
</tr>
<tr>
<td>Max Age of Population</td>
<td>80 years</td>
<td>Assumed</td>
</tr>
<tr>
<td>Proportion of Pregnant Women</td>
<td>7.5%</td>
<td>Assumed</td>
</tr>
<tr>
<td>G6PD Deficiency Prevalence</td>
<td>0%</td>
<td>Assumed</td>
</tr>
<tr>
<td>CYP-2D6 Phenotype Prevalence</td>
<td>0%</td>
<td>Assumed</td>
</tr>
<tr>
<td><strong>Exposure to Mosquito Biting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of Age-Dependent Biting</td>
<td>0.85</td>
<td>(46,47)</td>
</tr>
<tr>
<td>Reference Age of Biting</td>
<td>8 years</td>
<td>(46,47)</td>
</tr>
<tr>
<td>Mosquito-to-Human Transmission Probability</td>
<td>0.50</td>
<td>(18)</td>
</tr>
<tr>
<td><strong>Hypnozoite Dynamics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate of Relapse</td>
<td>1/65 day⁻¹</td>
<td>(18)</td>
</tr>
<tr>
<td>Rate of Liver-Stage Hypnozoite Clearance</td>
<td>1/383 day⁻¹</td>
<td>(10)</td>
</tr>
</tbody>
</table>
**Table S2. Default Infection Dynamics Parameters for the Human Component of the Transmission Model.** A description of the default parameters used in the individual-based model to simulate clinical trial data are provided along with the numeric values and the corresponding references.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human Infection Duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent Period</td>
<td>10 days</td>
<td>(48)</td>
</tr>
<tr>
<td>Max PCR-Detectable Infection (No Immunity)</td>
<td>70 days</td>
<td>(10)</td>
</tr>
<tr>
<td>Min PCR-Detectable Infection (Full Immunity)</td>
<td>10 days</td>
<td>(10)</td>
</tr>
<tr>
<td>LM-Detectable Infection</td>
<td>10 days</td>
<td>(18)</td>
</tr>
<tr>
<td>Clinical Disease (Untreated)</td>
<td>5 days</td>
<td>(10)</td>
</tr>
<tr>
<td>Clinical Disease (Treated)</td>
<td>1 day</td>
<td>(10)</td>
</tr>
<tr>
<td><strong>Immunity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Parasite Immune Boosting Refractory Period</td>
<td>19.77 days</td>
<td>(18)</td>
</tr>
<tr>
<td>Duration of Anti-Parasite Immunity</td>
<td>10 years</td>
<td>(10)</td>
</tr>
<tr>
<td>Probability of LM-Detectable Infection (No Immunity)</td>
<td>0.8918</td>
<td>(18)</td>
</tr>
<tr>
<td>Probability of LM-Detectable Infection (Full Immunity)</td>
<td>0.0043</td>
<td>(18)</td>
</tr>
<tr>
<td>Anti-Parasite Immunity for 50% Reduction in Duration of LM-Detectable Infection</td>
<td>27.52</td>
<td>(18)</td>
</tr>
<tr>
<td>Shape Parameter for LM-Detectable Infection</td>
<td>2.403</td>
<td>(18)</td>
</tr>
<tr>
<td>Anti-Parasite Immunity for 50% Reduction in Duration of PCR-Detectable Infection</td>
<td>9.9</td>
<td>(18)</td>
</tr>
<tr>
<td>Shape Parameter for Duration of PCR-Detectable Infection</td>
<td>4.602</td>
<td>(18)</td>
</tr>
<tr>
<td>Clinical Immune Boosting Refractory Period</td>
<td>7.85 days</td>
<td>(18)</td>
</tr>
<tr>
<td>Duration of Clinical Immunity</td>
<td>30 years</td>
<td>(10)</td>
</tr>
<tr>
<td>Probability of Clinical Episode (No Immunity)</td>
<td>0.8605</td>
<td>(18)</td>
</tr>
<tr>
<td>Probability of Clinical Episode (Full Immunity)</td>
<td>0.018</td>
<td>(18)</td>
</tr>
<tr>
<td>Clinical Immunity for 50% Reduction in Clinical Episode</td>
<td>11.538</td>
<td>(18)</td>
</tr>
<tr>
<td>Shape Parameter for Probability of Clinical Episode</td>
<td>2.25</td>
<td>(18)</td>
</tr>
<tr>
<td>Newborn Immunity Relative to Mother’s</td>
<td>0.421</td>
<td>(18)</td>
</tr>
<tr>
<td>Duration of Maternal Immunity</td>
<td>35.148</td>
<td>(18)</td>
</tr>
<tr>
<td><strong>Infectiousness to Mosquitoes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During PCR-Detectable Infection</td>
<td>0.035</td>
<td>(49)</td>
</tr>
<tr>
<td>During LM-Detectable Infection</td>
<td>0.1</td>
<td>(49)</td>
</tr>
<tr>
<td>During Clinical Disease</td>
<td>0.8</td>
<td>(49)</td>
</tr>
<tr>
<td>During Treated Clinical Disease</td>
<td>0.4</td>
<td>(49)</td>
</tr>
</tbody>
</table>
Table S3. Default Parameters for the Mosquito Component of the Transmission Model. A description of the default parameters used in the individual-based model to simulate clinical trial data are provided along with the numeric values and the corresponding reference.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mosquito Bionomics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mosquito Life Expectancy</td>
<td>6 days</td>
<td>(18)</td>
</tr>
<tr>
<td>Duration of Sporogony</td>
<td>8.4 days</td>
<td>(18)</td>
</tr>
<tr>
<td>Human Blood Index</td>
<td>0.50</td>
<td>Assumed</td>
</tr>
<tr>
<td>Proportion of Bites Occurring in Bed</td>
<td>0.45</td>
<td>Assumed</td>
</tr>
<tr>
<td>Proportion of Bites Occurring Indoors</td>
<td>0.90</td>
<td>Assumed</td>
</tr>
<tr>
<td><strong>Mosquito Larval Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development Time of Early Larval Instars</td>
<td>6.64 days</td>
<td>(50)</td>
</tr>
<tr>
<td>Development Time of Late Larval Instars</td>
<td>3.72 days</td>
<td>(50)</td>
</tr>
<tr>
<td>Development Time of Pupae</td>
<td>0.64 days</td>
<td>(50)</td>
</tr>
<tr>
<td>Per Capita Mortality Rate of Early Larval Instars</td>
<td>0.034 day(^{-1})</td>
<td>(50)</td>
</tr>
<tr>
<td>Per Capita Mortality Rate of Late Larval Instars</td>
<td>0.035 day(^{-1})</td>
<td>(50)</td>
</tr>
<tr>
<td>Per Capita Mortality Rate of Pupae</td>
<td>0.25 day(^{-1})</td>
<td>(50)</td>
</tr>
<tr>
<td>Daily Number of Eggs Laid per Mosquito</td>
<td>21.19</td>
<td>(50)</td>
</tr>
<tr>
<td>Effect of Density Dependence on Late Larval Instars Relative to Early Larval Instars</td>
<td>13.25</td>
<td>(50)</td>
</tr>
</tbody>
</table>

**Trial Design**

**Power Calculations**

To ensure that our simulated clinical trials were appropriately powered, we performed a retrospective power analysis. We took the outputs of our simulated clinical trials performed in transmission settings with homogeneous biting and computed the median and interquartile range of sample sizes in each arm necessary to achieve a specific level of power at a given EIR. The incidence rate and the proportion at risk in the control arm were computed using the full record of LM-detectable/subclinical and clinical relapses caused by hypnozoites acquired prior to treatment. For each power calculation, we assumed a type-I error rate of 0.05 and an efficacy of 75%.

Using the Cox proportional hazards model (51), we first calculated the total number of events \(N\) needed to have power \(\beta\) and type-I error rate \(\alpha\) as
\[ N = \frac{(z_\beta + z_{1-\alpha/2})^2}{P_t P_c \log^2(E_{Cox})}, \quad (S1) \]

where \( z_\beta \) and \( z_{1-\alpha/2} \) are the z-scores, \( P_t \) and \( P_c \) are the respective proportion of participants in the treatment and control arms, and \( E_{Cox} \) is the expected efficacy. We then calculated the number of participants in the control arm by dividing the total number of events needed by the expected total incidence rate computed from the trial output. The number of participants in the treatment arm was equal to the number of participants in the control arm.

Based on incidence rates, we followed Smith et al. (52) and computed the number of participants in each arm as

\[ n = \left( z_\beta + z_{1-\alpha/2} \right)^2 \frac{r_c + r_t}{(r_c - r_t)^2}, \quad (S2) \]

where \( z_\beta \) and \( z_{1-\alpha/2} \) are the z-scores, \( r_c \) is the expected incidence rate in the control arm, and \( r_t \) is the expected incidence rate in the treatment arm.

Based on proportion at risk, we computed the number of participants in each arm as

\[ n = \left( z_\beta + z_{1-\alpha/2} \right)^2 \frac{2p(1-p)}{(p_c - p_t)^2}, \quad (S3) \]
In eq. (S3), $z_\beta$ and $z_{1 - \alpha}$ are the z-scores, $p_t$ and $p_c$ are the expected proportion of participants at risk in the treatment and control arms, and $p$ is the average of $p_t$ and $p_c$ (52).

**Simulation Scenarios**

*Effect of Transmission Intensity and Heterogeneity in Biting*

The transmission intensity at the trial location may determine the frequency, cause, and detectability of recurrent infections that occur within the duration of follow-up. The number of reinfection events experienced by each trial participant increases with EIR. Thus, at higher EIR, the proportion of recurrent infections that are relapses caused by hypnozoites acquired prior to treatment may be small. Moreover, the levels of anti-parasite and clinical immunity in each trial participant is a function of exposure and transmission intensity, and, as EIR increases, fewer recurrent infections may be detected by light microscopy. Therefore, the transmission intensity at the trial location may bias efficacy estimates, likely downward, due to the rate at which trial participants are reinfected by mosquitoes.

Heterogeneity in mosquito biting contributes to observed heterogeneous patterns of *Plasmodium* transmission (16). For *P. falciparum*, heterogeneity in biting biases efficacy estimates towards the null, because fewer participants in the treatment and control arms are exposed within the duration of follow-up (42). For *P. vivax*, however, heterogeneity in biting may reduce the bias due to transmission intensity, because fewer participants are reinfected by mosquitoes within the duration of follow-up.

We varied the transmission intensity and the degree of heterogeneity in mosquito biting to quantify their effect on efficacy estimates. We varied transmission intensity by simulating a population-level EIR of 0.1, 1, 10, or 100 infectious bites per person-year (*ibppy*). To account for
heterogeneity in biting, we varied the variance of the distribution of logged biting propensities from zero to three. This distribution determines the variation in exposure to mosquito biting across individuals, and a zero variance implies homogeneous biting.

**Effect of Rate of Relapse and Duration of Follow-up**

The rate of *P. vivax* relapse varies geographically, with the estimated mean time to first relapse ranging from 41-65 days for tropical phenotypes to 299 days for temperate phenotypes (17). If the duration of follow-up is too short relative to the mean time to first relapse, then an appreciable number of trial participants may not have relapsed by the duration of follow-up. A longer duration of follow-up may capture a great proportion of relapses but also increases the temporal window over which trial participants can become reinfected by mosquito biting. The probability of reinfection during the duration of follow-up will vary with the EIR.

We examined the interaction between the rate of relapse, the duration of follow-up, and transmission intensity and quantified its effect on the efficacy estimates in simulated clinical trials. We fixed the mean time to relapse at 30, 60, 90, or 180 days and simulated clinical trials in which the duration of follow-up was 90, 180, 365, or 730 days. Trial participants were assessed for asexual parasites using light microscopy according to the following schedule: days 8, 15, 22, 29, 60, and 90 for the 90-day follow-up period; days 8, 15, 22, 29, 60, 90, 120, and 180 for the 180-day follow-up period; days 8, 15, 22, 29, 60, 90, 120, 180, 240, 300, and 365 for the 365-day follow-up period; and days 8, 15, 22, 29, 60, 90, 120, 180, 240, 300, 360, 420, 480, 540, 600, 660, and 730 for the 730-day follow-up period. Moreover, to examine the effect of transmission intensity, we set the population-level EIR at 0.1, 1, 10, or 100. To isolate the joint effect of the
rate of relapse and duration of follow-up across transmission intensities, we assumed homogeneous biting.

Effect of Vector Control

The implementation of vector control measures may protect trial participants against mosquito biting and reduce the number of reinfection events during follow-up. As a result, vector control may reduce the bias due to frequent reinfection of trial participants at higher transmission intensities.

To evaluate whether vector control measures are effective in reducing the bias in the efficacy estimates, we simulated clinical trials under four different vector control scenarios: (1) no vector control; (2) distribution of LLINs; (3) administration of IRS; and (4) distribution of LLINs and administration of IRS. The extent to which LLINs and IRS prevent mosquito biting depends upon the biting behavior of the mosquito population. We simulated different biting behaviors by varying the proportion of bites that occurred while individuals were indoors ($\Phi_I$) and while individuals were in bed ($\Phi_B$). We considered four biting behaviors: (1) endophagic with no preference in biting time ($\Phi_I = 0.9; \Phi_B = 0.45$); (2) endophagic with nighttime biting ($\Phi_I = 0.9; \Phi_B = 0.675$); (3) endophagic with daytime biting ($\Phi_I = 0.9; \Phi_B = 0.225$); and (4) exophagic with no preference in biting time ($\Phi_I = 0.1; \Phi_B = 0.05$). We simulated each combination of vector control scenario and mosquito biting behavior at population-level EIRs of 0.1, 1, 10, and 100. By default, we assumed homogeneous biting.

Effect of Genotyping
Although vector control may decrease the bias by reducing the number of reinfection events, a method capable of distinguishing relapses caused by hypnozoites acquired prior to treatment from all other recurrent infections is needed to completely correct the bias in efficacy estimates. Genotyping methods have been successfully used to determine the cause of each *P. vivax* recurrent infection (12,33) and therefore could correct biases in phase-III clinical trials.

We evaluated the potential for a genotyping method at varying sensitivities and specificities to correct the bias in the efficacy estimates. We considered a genotyping method intended to distinguish relapses caused from hypnozoite batches acquired prior to treatment from all other recurrent infections. Therefore, the sensitivity of the genotyping method was the probability of correctly identifying a relapse caused by a hypnozoite batch acquired prior to treatment, and the specificity of the genotyping method was the probability of correctly identifying all other recurrent infections. We simulated clinical trials at population-level EIRs of 0.1, 1, 10, and 100, assuming homogeneous biting. We then analyzed the simulated outputs of the clinical trials using a hypothetical genotyping method with sensitivities and specificities of 25%, 50%, 75%, or 100%.

**Effect of the 8-aminoquinoline**

By default, we simulated clinical trials for an 8-aminoquinoline with prophylactic effect lasting 28 days, consistent with treatment with chloroquine and primaquine. Under this assumption, the duration of prophylaxis is less than the 32-day period of left-censoring, so the suppression of blood-stage parasites during prophylaxis will not bias our efficacy estimates. Treatment with tafenoquine when co-administered with chloroquine, however, provides prophylaxis for 45 days and, therefore, may bias upwards our efficacy estimates (6).
To examine whether the duration of prophylaxis affects our estimates of efficacy, we simulated and compared phase-III clinical trials for primaquine and tafenoquine with prophylactic effects lasting 28 and 45 days, respectively. Because the benefit of prophylaxis may vary with transmission intensity, we simulated each trial at population-level EIRs of 0.1, 1, 10, and 100 with homogeneous biting.

Effect of the efficacy metric and infection endpoint

The combination of efficacy metric and infection endpoint used to calculate efficacy varied across recent phase-III clinical trials (6–9). Different combinations of efficacy metrics and infection endpoints could make comparison of efficacy estimates across different trials challenging if efficacy estimates are sensitive to this feature of the trial design.

To test whether our efficacy estimates are sensitive to the choice of the efficacy metric and the infection endpoint used, we simulated phase-III clinical trials at EIRs of 0.1, 1, 10, and 100 and calculated efficacy using different combinations of efficacy metrics and infection endpoints. The efficacy metrics examined were the Cox proportional hazards model, incidence rates, and the proportion at risk. The infection endpoints examined were PCR-detectable recurrent infections, LM-detectable recurrent infections, and clinical recurrent infections. Because we wanted to isolate the effect of the efficacy metric and the infection endpoint, we simulated each trial assuming homogeneous biting.
Results

**Figure S1. Comparison of efficacy metrics and infection types.** Efficacy estimates calculated using different methods and using different infection types with perfect genotyping are compared across different entomological inoculation rates (EIR). The color of the point denotes the recurrent infection endpoint used, where orange is PCR-detectable, red is LM-detectable, and maroon is clinical. Because we assumed perfect genotyping, all infections were relapses caused by hypnozoite batches acquired prior to 8-aminoquinoline treatment. Each point is the median of 200 simulations, and the bar is the interquartile range. The dotted line is the true clearance probability simulated in each clinical trial.
Figure S2. Effect of vector control on efficacy estimates across different vector bionomics. The impact of LLIN distribution (red), IRS administration (green), and combined LLIN distribution and IRS administration (teal) on resolving biases in LM-detectable recurrence free efficacy estimates is compared to a no-intervention scenario (dark blue) across a range of entomological inoculation rates (EIR). Each point represents the median of 200 simulations, and each bar is the interquartile range. The dotted line is the clearance probability simulated in each trial. The absolute proportion of bites occurring indoors ($\Phi_I$) and in bed ($\Phi_B$) varied with (A) $\Phi_I = 0.9$ and $\Phi_B = 0.45$; (B) $\Phi_I = 0.9$ and $\Phi_B = 0.675$; (C) $\Phi_I = 0.9$ and $\Phi_B = 0.225$; and (D) $\Phi_I = 0.1$ and $\Phi_B = 0.05$. 
Figure S3. The effect of vector control measures on reducing reinfections across different vector bionomics. The percentage of recurrent infections in the control arm that are treatment failures (i.e., relapses arising from hypnozoites acquired prior to enrollment) is compared across different entomological inoculation rates (EIR) and under different vector control scenarios. Dark blue bars represent trials in which no vector control measures were provided, red bars are trials in which LLINs were distributed, green bars are trials in which IRS was administered, and teal bars are trials in which both LLINs and IRS were administered. The height of each bar is the median of 200 simulations, and the segment is the interquartile range.
Figure S4. Comparison of “all-or-none” and “leaky” actions on efficacy estimates. LM-detectable recurrence-free efficacy estimates obtained from simulated clinical trials are shown for “all-or-none” (circles) and “leaky” (triangles) interventions at different entomological inoculation rates (EIR) and levels of heterogeneity in biting. Each point represents the median of 200 simulations, and each bar is the interquartile range. The color represents the degree of heterogeneity in individual-level exposure to mosquito biting with dark colors representing greater heterogeneity. The dotted line is the clearance probability simulated in the trials.
Figure S5. Sample sizes necessary to achieve a specific level of power. For a given EIR, the sample size in each arm is reported to achieve a specific level of power when calculating efficacy using (A) the Cox proportional hazards model, (B) incidence rates, and (C) the proportion at risk. Incidence rates were computed using the number of relapses caused by hypnozoites acquired prior to treatment in the control arm. Power calculations were performed for each of 200 simulations assuming homogeneous biting. The points and line denote the median sample size, and the shaded region is the interquartile range. Efficacy of the intervention was assumed to be 75%, the type-I error rate was set to 0.05, and the allocation ratio between trial arms was one.
REFERENCES


