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

John H. Huber[1*], Cristian Koepfli[1], Guido España[1], Narimane Nekkab[2], Michael T. White[2], T. Alex Perkins[1]

[1]Department of Biological Sciences & Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, USA

[2]Unité Malaria: Parasites et Hôtes, Département Parasites et Insectes Vecteur, Institut Pasteur, Paris, FRA

Corresponding Author: jhuber3@nd.edu (JHH)

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

45 failure (12). Only relapses caused by hypnozoite batches (i.e., a group of hypnozoites that
46 derives from one infectious mosquito bite) present prior to treatment directly reflect the extent to
47 which 8-aminoquinoline treatment prevents relapse, or the hypnozoite batch clearance
48 probability (hereafter, clearance probability) (Fig. 1). Because there remains no reliable way to
49 distinguish these relapses from all other recurrent infections using epidemiological data alone,
50 efficacy against recurrent infection is not equal to efficacy against relapse, and the measured
51 efficacy will likely vary depending upon the site in which the trial is conducted. Obtaining
52 standardized estimates of the effect of each 8-aminoquinoline against hypnozoite batches will
53 therefore depend upon the proportion of recurrent infections identified during follow-up that are
54 relapses caused by hypnozoite batches acquired prior to 8-aminoquinoline treatment.

55 An incongruence between the primary endpoint and the effect that the clinical trial
56 intends to measure has previously been noted as a source of bias in measuring the therapeutic
57 cure rates for *P. falciparum* malaria and tuberculosis (13–15). We propose that the magnitude of
58 this bias could depend upon features of the trial location and the trial design. Although most or
59 all clinical *P. vivax* patients enrolled in clinical trials are expected to carry hypnozoites, the rates
60 at which participants experience reinfections and relapses may vary across trial sites due to
61 transmission intensity, heterogeneous mosquito biting patterns (16), and the relapse phenotype of
62 the *P. vivax* parasite (17). The number of recorded relapses also depends upon the duration of
63 follow-up and the duration of prophylaxis provided by the treatment regimen used in the trial.
64 The extent to which these different features of the trial location and the trial design contribute to
65 bias in efficacy estimates—and impact our ability to measure the clearance probability—has not
66 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.

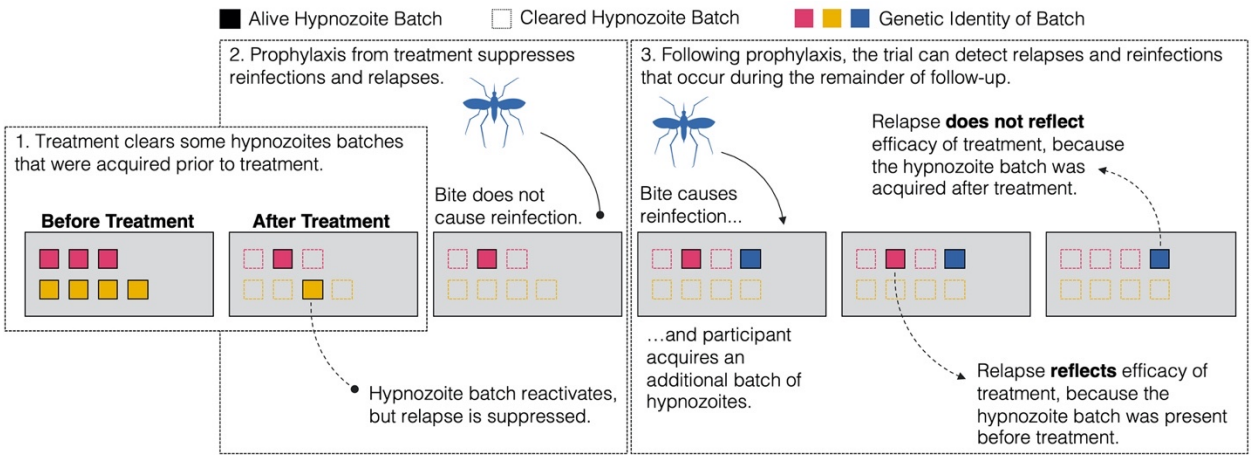


Figure 1. Schematic of potential infection outcomes during trial follow-up. *Participants are enrolled in the trial and initially harbor hypnozoite batches (colored squares). The colors represent the genetic identity of each hypnozoite batch. Hypnozoite batches are cleared (unfilled dotted squares) with 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.*

87 **RESULTS**

88 *Effect of Transmission Intensity and Heterogeneity in Biting*

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

94 When exposure to mosquito biting was the same for all trial participants (Fig. 2B), we
95 observed a downward bias in efficacy estimates that increased with transmission intensity (Fig.
96 2A). For clinical trials of an 8-aminoquinoline with a clearance probability equal to 0.75, light
97 microscopy (LM)-detectable recurrence-free efficacies ranged from 0.80 (IQR: 0.79 - 0.82) at an
98 EIR of 0.1 to 0.35 (0.30 - 0.40) at an EIR of 100. This downward bias in efficacy occurred
99 because reinfection by mosquito biting, an infection event that does not reflect the clearance
100 probability, was more frequent at higher EIR. When EIR was 100, 99% (SD: 0.22%) of
101 participants were reinfected on average by mosquitoes during the 180-day follow-up period,
102 compared to only 0.73% (SD: 0.18%) at an EIR of 0.1.

103 At low transmission intensities when reinfection rates were low, our efficacy estimates
104 exceeded the clearance probability, revealing a positive bias. This positive bias occurred because
105 we assumed that radical cure completely cleared hypnozoite batches in a subset of trial
106 participants and did not clear hypnozoite batches in the remainder of trial participants. In trials
107 measuring time to first recurrent infection, those participants that completely clear hypnozoite
108 batches increase the total person-time at risk in the treatment arm, leading to a competing
109 positive bias that was present in all trial simulations. This bias was eliminated if efficacy was

110 calculated based upon the proportion at risk using the complete record of recurrent infections
111 (Fig. S1).
112 In addition to the effect of transmission intensity itself, the effect of heterogeneous biting
113 on efficacy estimates was modulated by the transmission intensity of the trial location. At a low
114 EIR of 0.1, efficacy estimates were insensitive to changes in the level of heterogeneity in biting
115 (Fig. 2A), because the upper bound of trial participants reinfected during follow-up was 0.73%
116 (SD: 0.18%). By contrast, at intermediate EIRs of 1 and 10, the downward bias was reduced with
117 greater heterogeneity in biting. Efficacy measured at an EIR of 10 ranged from 0.58 (0.52 - 0.62)
118 under homogeneous biting to 0.74 (0.72 - 0.76) under the highest level of heterogeneity in biting
119 simulated. Heterogeneous biting reduced the downward bias caused by frequent reinfection,
120 because fewer infections observed during follow-up were reinfections from mosquito biting. For
121 example, at an EIR of 10, the proportion of individuals re-infected during follow-up ranged from
122 74% (SD: 0.94%) under homogeneous biting to 18% (0.78%) at the highest level of
123 heterogeneity in biting simulated.
124 At an EIR of 100, efficacy estimates had a non-monotonic relationship with the extent of
125 heterogeneous biting (Fig. 2A). At high transmission intensities, greater heterogeneity in biting
126 contributed to greater variation in anti-parasite immunity among trial participants. Trial
127 participants that were more frequently exposed to mosquito biting were expected to have greater
128 levels of anti-parasite immunity, thereby reducing the detectability of each *P. vivax* infection
129 during follow-up. Therefore, although the percentage of participants reinfected during follow-up
130 decreased monotonically with greater heterogeneity in mosquito biting, the percentage of
131 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).

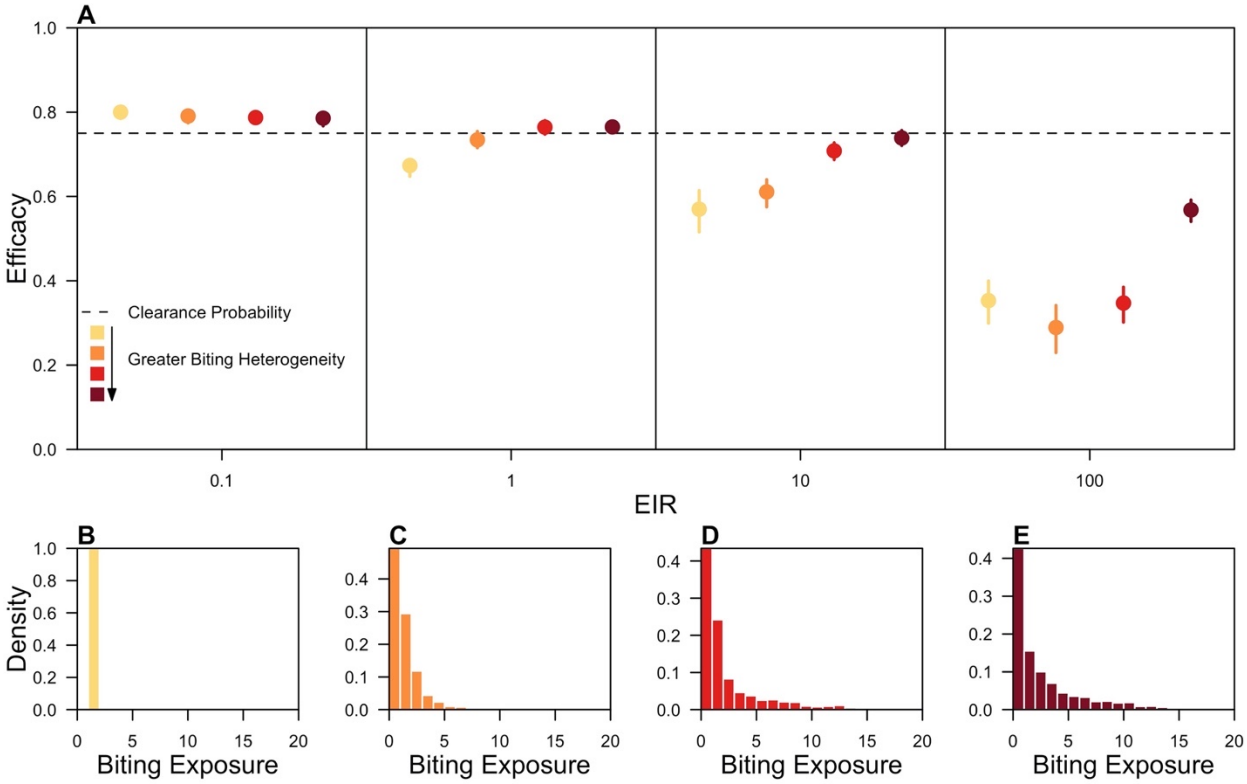


Figure 2. Effect of transmission intensity and heterogeneity in biting on efficacy estimates. *(A) LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIR) and levels of heterogeneity in biting. Each point represents the 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,

150 180, 365, or 730 days. To assess the effect of transmission intensity, all trials were simulated at
151 EIRs of 0.1, 1, 10, and 100, assuming homogeneous biting.

152 For a given duration of follow-up, efficacy estimates decreased with a longer mean time
153 to first relapse (Fig. 3A). In simulated trials in which participants were followed for 180 days at
154 an EIR of 1, we estimated that the efficacy of an 8-aminoquinoline with a clearance probability
155 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
156 relapse increased from 30 to 180 days. This bias toward the null occurred because, as the mean
157 time to first relapse increased, fewer trial participants were expected to relapse within the
158 duration of follow-up (Fig. 3B), causing the distribution of times to first recurrent infection to
159 appear more similar across trial arms. Only 63.2% of participants were expected to relapse by
160 180 days when the mean time to first relapse was 180 days, compared to 99.8% of participants
161 when the time to first relapse was 30 days.

162 For a given mean time to first relapse, our efficacy estimates decreased with a longer
163 duration of follow-up (Fig. 3A). With an EIR equal to 1 and a mean time to first relapse of 30
164 days, we estimated that the efficacy of an 8-aminoquinoline was 0.73 (0.70 - 0.75) when trial
165 participants were followed for 90 days, compared to 0.54 (0.52 - 0.56) when trial participants
166 were followed for 730 days. Under an alternative scenario in which the EIR was 10, the
167 estimated efficacies were 0.70 (0.65 - 0.74) and 0.39 (0.37 - 0.43) at the respective durations of
168 follow-up. The downward bias occurred because a longer duration of follow-up ensured that
169 more trial participants were reinfected by mosquito biting (Fig. 3C), similarly causing the
170 distribution of times to first recurrent infection to appear more similar across the trial arms. By
171 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.

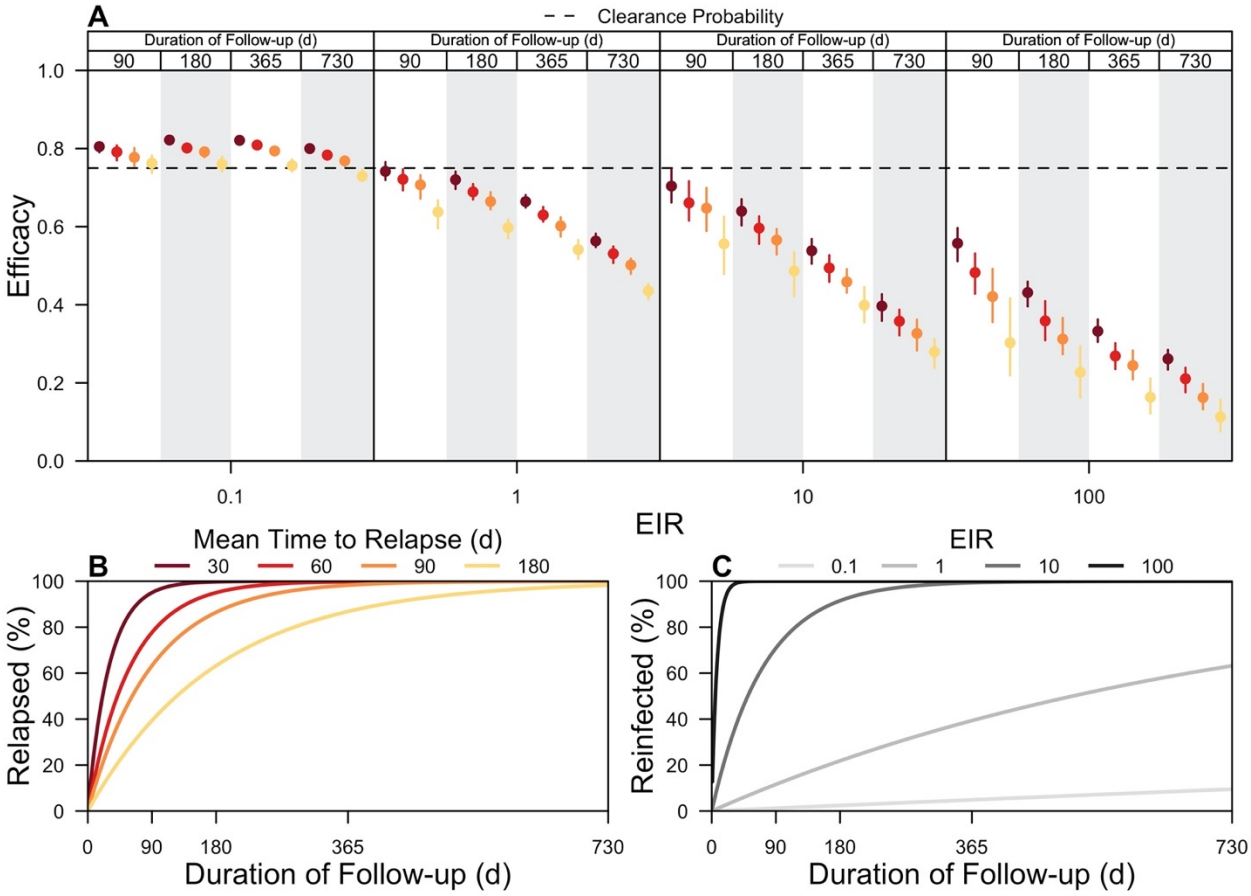


Figure 3. Effect of rate of relapse and duration of follow-up on efficacy estimates. *(A) LM-detectable recurrence-free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIRs), mean times to relapse, and durations of follow-up. Each point is the 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.*

189 *Effect of Vector Control*

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

197 Vector control interventions were effective in reducing the downward bias due to
198 transmission intensity in trial locations where the vector was predominantly endophagic (i.e.,
199 feeds indoors) (Fig. 4). With an endophagic vector with no preference in biting time and at an
200 EIR of 10, the estimated efficacies improved from 0.52 (0.46 - 0.57) in the absence of vector
201 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
202 the combination of LLINs and IRS (Fig. 4A). By comparison, with an exophagic vector with no
203 preference in biting time and at an EIR of 10, efficacy estimates improved only slightly from
204 0.52 (0.48 - 0.57) in the absence of vector control and 0.52 (0.47 - 0.57) with LLINs to 0.53
205 (0.49 - 0.58) with IRS and 0.53 (0.48 - 0.58) with the combination of LLINs and IRS (Fig. S2D).

206 Across all mosquito biting behaviors simulated, the combination of LLINs and IRS was
207 generally more effective than either intervention in isolation in reducing the downward bias in
208 efficacy estimates (Fig. S2). When considering interventions in isolation, IRS was more effective
209 than LLINs in reducing the downward bias, and the distribution of LLINs most improved
210 efficacy estimates under scenarios in which an appreciable number of mosquito bites were taken
211 at night while trial participants were in bed.

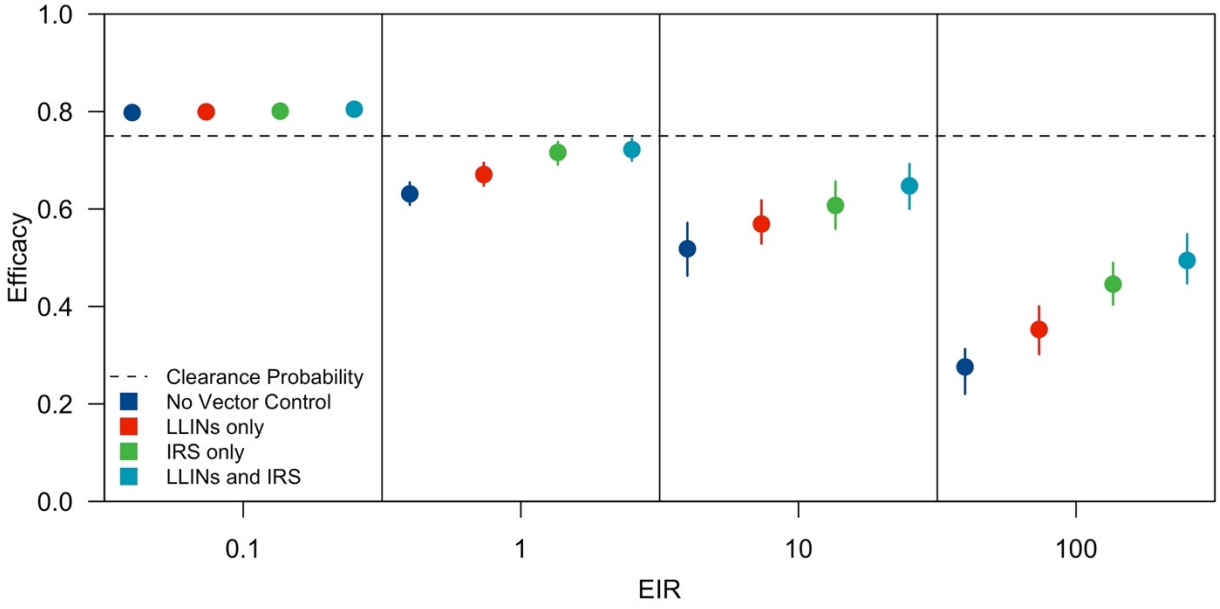


Figure 4. Effect of vector control on efficacy estimates. *The impact of LLIN distribution (red), IRS administration (green), and combined LLIN distribution and IRS administration (teal) on resolving biases in LM-detectable recurrence free efficacy estimates is compared to a no-intervention scenario (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 (Φ_I) was 0.90 and in bed (Φ_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

231 this simulation study, we defined sensitivity as the probability of correctly identifying relapses
232 caused by hypnozoite batches acquired prior to treatment and specificity as the probability of
233 correctly identifying all other recurrent infections (e.g., reinfections and relapses caused by
234 hypnozoite batches acquired after treatment).

235 At lower and intermediate transmission intensities (i.e., EIRs of 0.1 and 1), efficacy
236 estimates changed more with improved sensitivity than improved specificity of the genotyping
237 method. By contrast, at higher transmission intensities, a genotyping method with high
238 specificity was needed to reduce the downward bias in efficacy. At an EIR of 100 and assuming
239 100% sensitivity of the genotyping method, the estimated efficacy of an 8-aminoquinoline with
240 clearance probability equal to 0.75 improved from 0.38 (0.33 - 0.44) at 25% specificity to 0.67
241 (0.62 - 0.71) at 100% specificity. Under the alternative scenario in which specificity of the
242 genotyping method was 100%, the estimated efficacy at the respective EIR improved from 0.48
243 (0.35 - 0.58) at 25% sensitivity to 0.67 (0.62 - 0.71) at 100% sensitivity. The specificity of the
244 genotyping method was more important at higher EIR, because, at an EIR of 100, 100% of
245 participants were expected to be reinfected during follow-up (Fig. 3C). Misclassifying
246 reinfection events as failures of 8-aminoquinoline treatment (i.e., relapses caused by hypnozoites
247 batches acquired prior to treatment) lead to a greater downward bias, so a highly specific method
248 was needed in settings where trial participants were frequently reinfected. These effects were less
249 pronounced at low transmission intensities, given that fewer trials participants were reinfected
250 during follow-up.

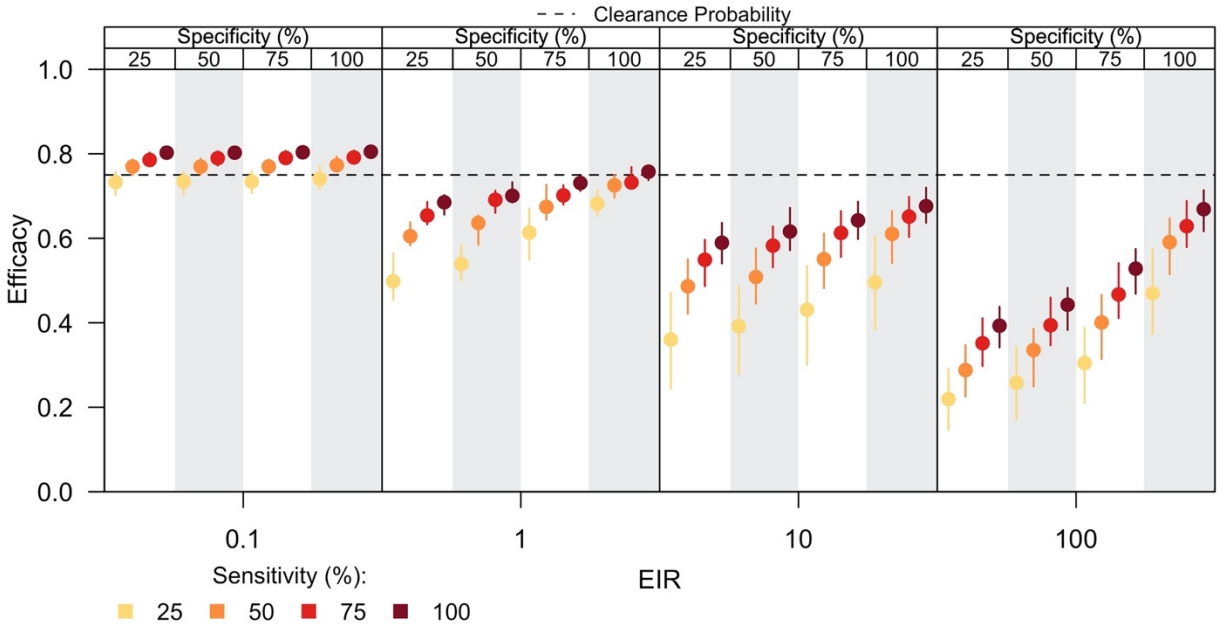


Figure 5. Effect of parasite genotyping on efficacy estimates. *The impact of genotyping recurrent infections to estimate the efficacy of radical cure is shown for different sensitivities and specificities of the genotyping method across a range of entomological inoculation rates (EIR). Sensitivity of 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.

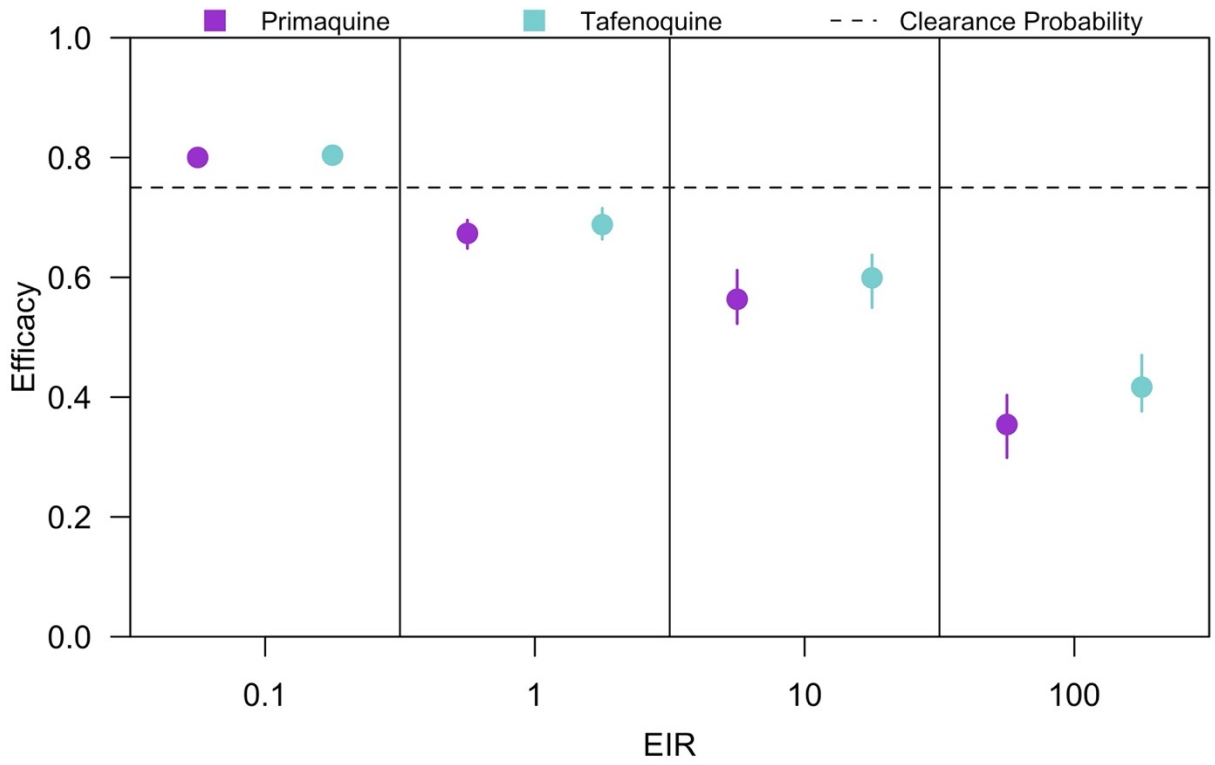


Figure 6. Effect of 8-aminoquinoline on efficacy estimates. *LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown for primaquine (purple) and tafenoquine (blue) at different entomological inoculation rates (EIR). Each point is the median of 200 simulations, and each 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

303 participants in both the treatment and control arms experienced a reinfection or relapse during
304 follow-up (Fig. 7). For PCR-detectable infections, estimated efficacy was highest if calculated
305 using the Cox proportional hazards model and lowest if calculated using the proportion at risk.
306 This suggested that, in the absence of parasite genotyping (Fig. S1), the magnitude of bias caused
307 by reinfection from mosquito biting was more severe when basing efficacy upon incidence rates
308 or the proportion at risk than upon proportional hazards.

309 The sensitivity of the assay by which trial participants were assessed for recurrent
310 infections also affected our efficacy estimates, though the direction of the effect depended upon
311 the chosen efficacy metric. For efficacy estimates based upon incidence rates or the proportion at
312 risk, shifting from a more sensitive assay (i.e., PCR-based detection) to a less sensitive assay
313 (i.e., monitoring for clinical infections) reduced the downward bias due to reinfection,
314 particularly at higher transmission intensities. This was due to the fact that the fraction of
315 recurrent infections ascertained increased with a more sensitive assay, causing us to detect more
316 recurrent infections during follow-up and making the incidence rates and proportions at risk
317 appear more similar across the treatment and control arms. By contrast, for efficacy estimates
318 based upon the Cox proportional hazards model, efficacy estimates were lower for a less
319 sensitive assay at EIRs of 0.1, 1, and 10. A less sensitive assay resulted in the time to the first
320 detected recurrent infection occurring later during follow-up for participants in both the
321 treatment and control arm, causing the observed downward bias. At a very high transmission
322 intensity (i.e., EIR of 100), a less sensitive assay minimized bias, because effectively all trial
323 participants were expected to be reinfected during follow-up (Fig. 2C).

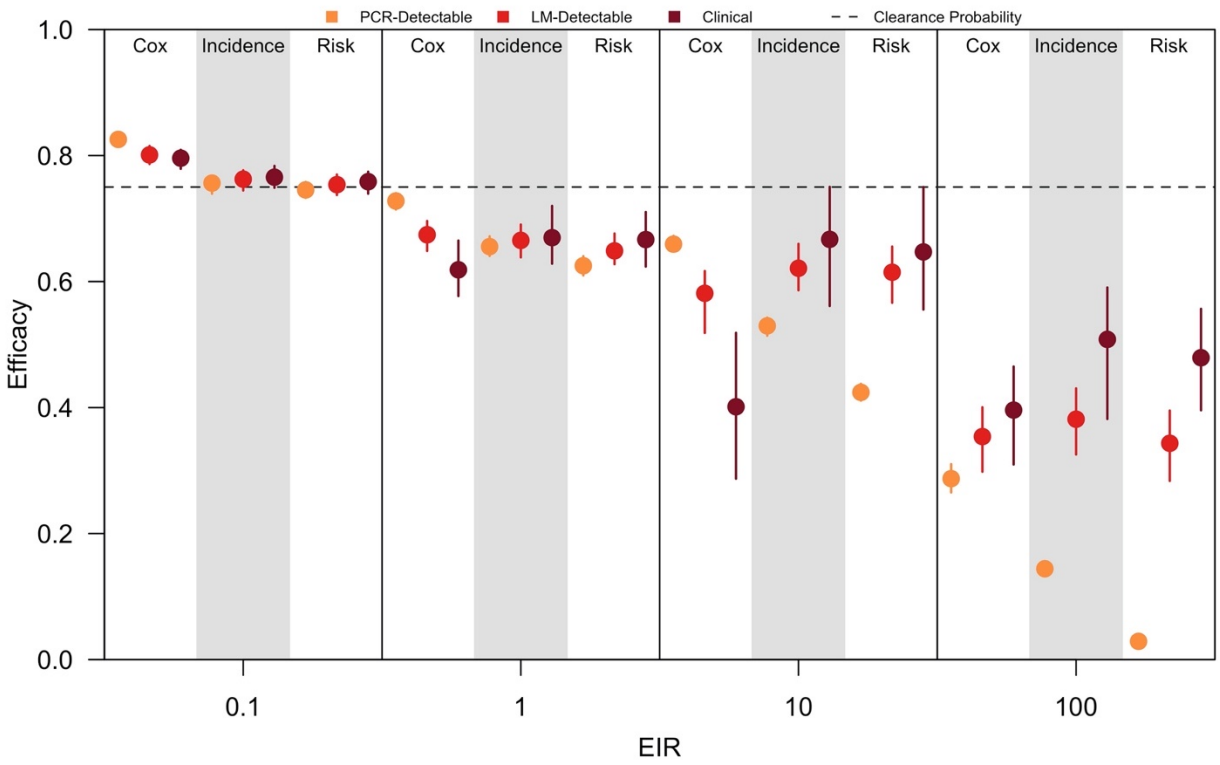


Figure 7. Effect of efficacy metric and infection endpoint on efficacy estimates. *Efficacy estimates obtained from simulated clinical trials is shown when calculated using the Cox proportional hazards model, incidence rates, or the proportion at risk. Efficacy calculations were performed using clinical (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

341 and genotyping methods are two approaches that can reduce and, in some cases, even correct
342 these site-specific biases and yield more standardized estimates of efficacy against relapse.
343 In the recent GATHER, DETECTIVE, and IMPROV trials (6–9), efficacy estimates
344 varied across trial sites. Interpreting site-specific differences in efficacy is important for past and
345 future clinical trials (19), and our simulation results suggest that the differences in efficacy
346 estimates could be caused by site-specific biases that arise when the infection endpoint used to
347 measure efficacy does not directly reflect the action of the therapeutic being trialed. If the trial
348 sites vary in transmission intensity, our results suggest that efficacy estimates could be lower in
349 high-transmission settings, because trial participants are more frequently reinfected during
350 follow-up. Greater heterogeneity in mosquito biting reduces this downward bias by reducing the
351 number of trial participants who are bitten and possibly reinfected during follow-up, though this
352 effect is non-monotonic at very high transmission intensities due to the predicted interaction with
353 anti-parasite immunity. Trial sites may also differ in the rates at which *P. vivax* hypnozoites
354 relapse (17). A longer time to relapse implies that an appreciable number of trial participants will
355 not yet have relapsed during follow-up, thereby biasing efficacy estimates downward (20).
356 Increasing the duration of follow-up further compounds the bias rather than corrects it, because a
357 long duration of follow-up causes more participants in both the treatment and control arms to
358 become reinfected by mosquito biting, particularly at higher transmission intensities. Finally,
359 frequent reinfection at higher transmission intensities makes the comparison of treatment
360 regimens with different durations of prophylaxis challenging, because a longer duration of
361 prophylaxis prevents more individuals in the treatment arm from reinfection. This results in an
362 upward bias that may cause us to overestimate the effect against hypnozoites of treatment

363 regimens that provide a longer period of prophylaxis, such as tafenoquine co-administered with
364 chloroquine.

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

378 Our results demonstrate that parasite genotyping could be used to generate less biased
379 efficacy estimates that better reflect the effect of the 8-aminoquinoline against hypnozoite
380 batches. Although in practice the sensitivity of a genotyping method may be reduced by not
381 observing infections that occurred prior to trial enrollment, the extent to which genotyping
382 reduced bias in our analysis was robust to changes in sensitivity and depended most upon the
383 specificity. High specificity can likely be achieved given the high expected heterozygosity of
384 microsatellite and amplicon deep sequencing panels, as well as the ability to detect minority
385 clones in multiclonal *P. vivax* infections (29–33). Statistical frameworks leveraging genotyping

386 data to distinguish relapses from reinfections have been successfully developed and could be
387 readily integrated into the analysis of future clinical trial data (12,33).

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

396 There are a number of limitations of this analysis. First, we did not calibrate our model to
397 clinical trial data, so our results are not representative of any specific trial settings. However, our
398 simulations encompassed a wide range of epidemiological settings, so the site-specific biases
399 identified in our analysis should reflect variation possible across trial sites. Second, there remains
400 much about hypnozoite biology that is not well understood (34), so our simplified representation
401 of hypnozoite activation and death in our transmission model may not fully capture reality.
402 Finally, we have an imperfect understanding of the acquisition and loss of anti-parasite and
403 clinical immunity (35). Underestimating the level of anti-parasite and clinical immunity among
404 trial participants may lead to a greater fraction of recurrent infections detected during follow-up
405 than would occur in an actual trial context. This could have led us to overestimate site-specific
406 biases, in which case more refined mathematical representations of immunity could be beneficial
407 for future studies building on ours (10,36).

408 Our analysis predicted that site-specific biases are likely to occur in phase-III trials for
409 radical cure, and these results suggest that care should be taken in the planning of future trials
410 and the interpretation of trial data. Mathematical modeling that accounts for site-specific biases
411 can aid in the interpretation of clinical trial data and may be useful for informing the choice of
412 efficacy metric and identifying future trial sites where biases may be less severe (37). As always,
413 the utility of the insights from mathematical models is improved with additional data, so
414 modeling should be integrated into clinical trial design to identify data needs and inform data
415 collection in order to reduce these biases and improve understanding of radical cure's potential to
416 control *P. vivax* malaria.
417

418 **METHODS**

419 ***Transmission Model***

420 To simulate *Plasmodium vivax* transmission, we used a stochastic, individual-based model
421 developed by White *et al*. (10) and Nekkab *et al*. (18). This model extends the Ross-MacDonald
422 framework of *P. falciparum* transmission to incorporate relapses (38), a characteristic feature of
423 *P. vivax* transmission (39). Calibrated to epidemiological surveys conducted in Papua New
424 Guinea, the Solomon Islands, and Brazil, the model by White *et al.* (10) and Nekkab *et al*. (18)
425 reproduces *P. vivax* transmission dynamics across a range of epidemiological settings.
426 The rates at which individuals in the population are reinfected and relapse depend upon
427 the respective number of infectious mosquito bites received (i.e., the entomological inoculation
428 rate, or EIR) and the number of hypnozoite batches present in each individual's liver. The
429 average EIR within the population is determined by the mosquito population dynamics, and
430 individual variation therein is due to heterogeneity in exposure and age-dependent differences in

431 biting. Following successful inoculation with *P. vivax* sporozoites from an infectious mosquito
432 bite, an individual experiences a primary bloodstream infection and accumulates an additional
433 batch of hypnozoites within the liver. In the absence of treatment, each batch of hypnozoites can
434 activate to cause relapses or be cleared naturally from the liver according to fixed rates (Table
435 S1).

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

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

450 Simulation of the *P. vivax* transmission model occurs in two steps. First, the population is
451 initialized at equilibrium according to the analogous set of deterministic compartmental
452 differential equations. Then, the stochastic, individual-based model is simulated with the

453 initialized population. For further description of the model and its assumptions, please refer to
454 Tables S1-S3 and the documentation provided in White *et al.* (10) and Nekkab *et al*. (18).
455
456 ***Trial Overview***
457 We constructed phase-III clinical trials for a generic 8-aminoquinoline (e.g., primaquine or
458 tafenoquine). We simulated trials in order to quantify the recurrence-free efficacy as a measure
459 of the therapeutic effect of radical cure against the hypnozoite batches present in each individual
460 presenting with a clinical *P. vivax* infection confirmed by light microscopy.
461 We designed our simulated phase-III clinical trials to be comparable to previous trials of
462 PQ and TFQ (6–8). Participants were randomly assigned to the treatment and control arms, and
463 all participants received chloroquine (CQ) upon enrollment to clear the *P. vivax* blood-stage
464 infection. Consistent with the DETECTIVE trial (6,8), participants were followed for 180 days
465 and assessed for asexual parasitemia by light microscopy at days 8, 15, 22, 29, 60, 90, 120, and
466 180 post-enrollment. For trials simulated with longer duration of follow-up, additional time
467 points were included. See supplement for further details.
468
469 ***Radical Cure Model***
470 We modeled the action of radical cure in each individual receiving treatment as the per-
471 hypnozoite batch probability of clearance. To allow for heterogeneous action of the intervention
472 (40), we assumed that a proportion p_i of the population belonged to each stratum i with per-
473 hypnozoite batch clearance probability, c_i. For a population consisting of two strata, the mean
474 per-hypnozoite batch probability of clearance was
475

476 $$\bar{c} = p_1 c_1 + p_2 c_2, \quad (1)$$

477

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

498 Under the model of heterogeneous action of radical cure, the number of batches of
499 hypnozoites following treatment was distributed as
500
501 $$H_{post} \sim Binomial\left(H_{pre}, 1 - c_i\right), \forall i \in \{1,2\}. \quad (2)$$
502
503 In eq. (2), H_{pre} is the number of batches of hypnozoite present in the liver prior to treatment, and
504 c_i is the per-hypnozoite batch clearance probability for an individual in stratum i.
505
506 ***Trial Design***
507 We constructed our simulated phase-III clinical trials for radical cure to match the trial design
508 used in past trials for PQ and TFQ (6–8). Accordingly, we considered four phases of the
509 randomized control trial design: (1) recruitment, (2) treatment, (3) vector control, and (3) follow-
510 up.
511
512 *Recruitment*
513 We enrolled individuals in the simulated clinical trial if they presented at a health clinic with
514 febrile illness and were diagnosed by light microscopy with a *P. vivax* infection. Consistent with
515 the DETECTIVE trial (6,8), participants were enrolled if they were at least 16 years of age and
516 had a measured glucose-6-phosphate dehydrogenase (G6DP) activity greater than 70% of the
517 median value of the trial location. Following Nekkab *et al*. (18), we assumed that G6PD activity
518 was measured using the SD Biosensor STANDARD G6PD test (44). This test qualitatively
519 classifies G6PD activity as normal (>70% activity), intermediate (30-70% activity), and low
520 (<30% activity).

521 Individuals that met the criteria for enrollment were then randomly assigned to the
522 treatment or control arm of the clinical trial. Allocation of participants to each trial arm occurred
523 with equal probability, provided that the current number of participants in each arm was less than
524 the desired sample size of 1,000.
525
526 *Treatment*
527 Upon enrollment, all trial participants were treated with a 3-day course of CQ to clear the blood-
528 stage, asexual parasites (6,8). We assumed that treatment with CQ was 100% effective and that
529 there were no recrudescences among trial participants. Participants in the treatment arm were
530 also treated with an 8-aminoquinoline to clear the hypnozoite reservoir. For individuals in the
531 treatment arm, the number of hypnozoite batches following radical cure was calculated
532 according to eq. (2). We further assumed that treatment with the 8-aminoquinoline provided both
533 blood-stage and liver-stage prophylaxis. Consistent with White *et al.* (10) and Nekkab *et al*. (18),
534 we assumed that the duration of prophylaxis was 28 days for PQ when co-administered with CQ
535 and 45 days for TFQ when co-administered with CQ.
536
537 *Vector Control*
538 As in the DETECTIVE trial (6,8), all participants were provided with a long-lasting insecticidal
539 net (LLIN) to prevent reinfection from mosquitoes during follow-up and therefore decrease the
540 potential bias that may arise in calculating efficacy based on a recurrent infection endpoint.
541 Following Griffin *et al*. (36) and White *et al*. (10), we assumed that vector control measures
542 decrease the probabilities of mosquito biting and successful feeding and increase the
543 probabilities of repellency and death. The magnitude of the effect depends upon the proportion

544 of mosquito bites that occur while individuals are indoors (Φ_I) and in bed (Φ_B), the duration of
545 usage of the LLIN, and the duration of insecticidal activity. We assumed that the mean duration
546 of usage of LLINs was 3 years and that the mean duration of insecticidal activity was 2.5 years.
547 We also considered whether the use of indoor residual spraying (IRS) in each
548 participant's house administered in isolation or in combination with LLINs decreased the
549 potential bias in our efficacy estimates. As with LLINs, we assumed that the use of IRS
550 decreases the probabilities of mosquito biting and successful feeding and increases the
551 probabilities of repellency and death (10,36). To account for the waning effect of IRS over time,
552 we assumed that the mean duration of insecticidal activity was 6 months.
553
554 *Follow-up*
555 We followed each participant for 180 days following enrollment in the clinical trial and
556 treatment with anti-malarial drugs. Consistent with the DETECTIVE trial (6,8), for each
557 participant, we recorded the date of each clinical *P. vivax* infection that occurred within the
558 duration of follow-up. Furthermore, we tested each participant for *P. vivax* asexual parasites
559 using light microscopy on days 8, 15, 22, 29, 60, 90, 120, 160, and 180 post-enrollment.
560 To examine potential biases that arise from trial surveillance methods, we kept a
561 complete record of all recurrent infections that occurred within the duration of follow-up,
562 including those that would not have been detected under the trial protocol. For each recurrent
563 infection, we recorded the cause (i.e., reinfection or relapse) and the type of blood-stage infection
564 (e.g., sub-microscopic, sub-clinical, or clinical). We further distinguished between relapses
565 caused by hypnozoite batches acquired prior to treatment and relapses caused by hypnozoites

566 batches acquired following treatment. Only relapses caused by hypnozoite batches acquired prior
567 to treatment reflect the action of 8-aminoquinoline treatment against hypnozoites batches.
568
569 ***Statistical Analyses***
570 We calculated the efficacy of the 8-aminoquinoline used in each clinical trial from the output
571 collected in each respective simulation. We calculated efficacy using multiple metrics in order to
572 examine how different data collected during trial follow-up resolved biases in the efficacy
573 estimates.
574 Consistent with previous phase-III clinical trials (6,8), we used freedom from LM-
575 detectable recurrent infection as the default efficacy metric in our simulation studies. We
576 calculated efficacy using the Cox proportional hazards model (45), which computes the hazard
577 ratio between the treatment and control arms based on the times to first LM-detectable recurrent
578 infection. Following the DETECTIVE trial protocol (6,8), recurrent infections that occurred
579 before 32 days post-enrollment were not included. Given that the timing of each recurrent
580 infection was unobserved, all trial participants were interval censored.
581 We calculated efficacy using incidence rates computed from the recurrent infections that
582 occurred within the duration of follow-up. We calculated efficacy based on incidence rates as
583
584 $$E_{IR} = 1 - \frac{e_t/y_t}{e_c/y_c}. \quad (3)$$
585

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

589 The magnitude of the measured efficacy also depends upon the action of the radical cure
590 therapeutic. In absence of other sources of bias, we will overestimate efficacy for 8-
591 aminoquinolines that elicit an "all-or-none" response if we calculate efficacy based on incidence
592 rates. This phenomenon is caused by the subset of individuals in the treatment arm who
593 completely clear the hypnozoites from their livers and increase the number of person-years of
594 follow-up (42). To resolve this bias that arises from the assumption of intervention action, we
595 calculated efficacy based on risk as
596

597 $$E_R = 1 - \frac{p_t}{p_c}, \quad (4)$$

598

599 where p_t and p_c are the proportion of individuals in the treatment and control arms that
600 experience a particular infection event within the duration of follow-up.
601

602 ***Simulation Experiments***

603 We performed simulation experiments to quantify biases in the efficacy estimates that may arise
604 under different transmission settings and other features of the clinical trial design. First, we
605 considered how our efficacy estimates varied with transmission intensity and heterogeneity in
606 mosquito biting patterns. Second, we considered how differences in the rate of relapse of the *P.*
607 *vivax* hypnozoites interact with the duration of the trial follow-up and the transmission intensity
608 to bias our estimates of efficacy. After characterizing these biases, we assessed whether the

609 allocation of vector control measures, such as LLINs and IRS, to trial participants could prevent
610 reinfections and thereby reduce the bias in the efficacy estimates. We next considered whether a
611 genotyping method capable of distinguishing between different types of *P. vivax* recurrent
612 infections at different sensitivities and specificities could reduce or correct the bias in the
613 efficacy estimates. Then, we measured the extent to which the duration of prophylaxis provided
614 by the treatment regimen biases our estimates of efficacy by preventing reinfections in the
615 treatment arm only. Finally, we considered how the efficacy estimates that we obtained varied
616 with the choice of efficacy metric and infection endpoint.

617 For each simulation setting, we simulated a phase-III clinical trial with 1,000 participants
618 in each arm. This sample size was chosen to ensure that the power of our clinical trials exceeded
619 95% for each transmission intensity and efficacy metric considered (Fig. S5). Each trial was
620 simulated in a population of 200,000 individuals, and the maximum duration of each trial was 15
621 years in order to ensure that complete enrollment was attained, even at low transmission
622 intensities. By default, PQ was the 8-aminoquinoline provided to the participants in the treatment
623 arm and when co-administered with CQ provided 28 days of blood-stage and liver-stage
624 prophylaxis during which trial participants could not become reinfected or relapse. The mean
625 per-hypnozoite batch probability of clearance was equal to 0.75. For each simulation setting, we
626 simulated trials in which the therapeutic had an "all-or-none" response. Because the transmission
627 model was stochastic, we ran 200 simulations for each simulation setting and computed the mean
628 and interquartile range for the efficacy estimates across these simulations. By default, we defined
629 efficacy as freedom from LM-detectable recurrent infection calculated using the Cox
630 proportional hazards model.
631

632 ***Table 1. Parameters Varied during the Simulation Analyses.*** *The parameters that were varied*
633 *during each simulation analysis are reported. All other parameters in the transmission model*
634 *are set to default values and are consistent with the values reported in Tables S1 and S2.*

Variance in Mosquito Biting	Time to Relapse (d)	Follow-up (d)	Prophylaxis (d)
Effect of Transmission Intensity and Heterogeneity in Biting			
0, 1, 2, 3	65	180	28
Effect of Rate of Relapse and Duration of Follow-up			
0	30, 60, 90, 180	90, 180, 365, 730	28
Effect of Vector Control / Genotyping / Efficacy Metric and Infection Endpoint			
0	65	180	28
Effect of the 8-aminoquinoline			
0	65	180	28,45

635

636 ***Availability of Code***

637 All code to reproduce the analyses in this study is available on Github at

638 https://github.com/johnhhuber/Radical_Cure_Uncertainty.

639

640 **ACKNOWLEDGMENTS**

641 JHH acknowledges funding from a National Science Foundation Graduate Research Fellowship

642 and a Richard and Peggy Notebaert Premier Fellowship from the University of Notre Dame. CK

643 received funding from NIH R21-AI137891. NN and MTW were supported by a grant from

644 Medicine for Malaria Venture. NN also received funding from a Pasteur Roux-Cantarini

645 Fellowship. This publication was made possible with partial support from Grant Numbers TL1

646 TR002531 and UL1TR002529 (A. Shekhar, PI) from the National Institutes of Health, National

647 Center for Advancing Translational Sciences, Clinical and Translational Sciences Award.

648

649 **COMPETING INTERESTS**

650 The authors have no competing interests to declare.

651 **SUPPORTING INFORMATION**

652 **Methods**

653 ***Transmission Model***

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

Parameter	Value	Reference
Demographics		
Mean Age of Population	31.8 years	Assumed
Max Age of Population	80 years	Assumed
Proportion of Pregnant Women	7.5%	Assumed
G6PD Deficiency Prevalence	0%	Assumed
CYP-2D6 Phenotype Prevalence	0%	Assumed
Exposure to Mosquito Biting		
Degree of Age-Dependent Biting	0.85	(46,47)
Reference Age of Biting	8 years	(46,47)
Mosquito-to-Human Transmission Probability	0.50	(18)
Hypnozoite Dynamics		
Rate of Relapse	1/65 day^{-1}	(18)
Rate of Liver-Stage Hypnozoite Clearance	1/383 day^{-1}	(10)

658

659 ***Table S2. Default Infection Dynamics Parameters for the Human Component of the***
660 ***Transmission Model.*** *A description of the default parameters used in the individual-based model*
661 *to simulate clinical trial data are provided along with the numeric values and the corresponding*
662 *references.*

Parameter	Value	Reference
Human Infection Duration		
Latent Period	10 days	(48)
Max PCR-Detectable Infection (No Immunity)	70 days	(10)
Min PCR-Detectable Infection (Full Immunity)	10 days	(10)
LM-Detectable Infection	10 days	(18)
Clinical Disease (Untreated)	5 days	(10)
Clinical Disease (Treated)	1 day	(10)
Immunity		
Anti-Parasite Immune Boosting Refractory Period	19.77 days	(18)
Duration of Anti-Parasite Immunity	10 years	(10)
Probability of LM-Detectable Infection (No Immunity)	0.8918	(18)
Probability of LM-Detectable Infection (Full Immunity)	0.0043	(18)
Anti-Parasite Immunity for 50% Reduction in Duration of LM-Detectable Infection	27.52	(18)
Shape Parameter for LM-Detectable Infection	2.403	(18)
Anti-Parasite Immunity for 50% Reduction in Duration of PCR-Detectable Infection	9.9	(18)
Shape Parameter for Duration of PCR-Detectable Infection	4.602	(18)
Clinical Immune Boosting Refractory Period	7.85 days	(18)
Duration of Clinical Immunity	30 years	(10)
Probability of Clinical Episode (No Immunity)	0.8605	(18)
Probability of Clinical Episode (Full Immunity)	0.018	(18)
Clinical Immunity for 50% Reduction in Clinical Episode	11.538	(18)
Shape Parameter for Probability of Clinical Episode	2.25	(18)
Newborn Immunity Relative to Mother's	0.421	(18)
Duration of Maternal Immunity	35.148	(18)
Infectiousness to Mosquitoes		
During PCR-Detectable Infection	0.035	(49)
During LM-Detectable Infection	0.1	(49)
During Clinical Disease	0.8	(49)
During Treated Clinical Disease	0.4	(49)

663

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.*

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

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 β and type-I error rate α as

$$N = \frac{\left(z_\beta + z_{1-\frac{\alpha}{2}}\right)^2}{P_t P_c \log^2(E_{Cox})}, \quad (S1)$$

where z_β and $z_{1-\frac{\alpha}{2}}$ 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

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

where z_β and $z_{1-\frac{\alpha}{2}}$ 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

$$n = \left(z_\beta + z_{1-\frac{\alpha}{2}}\right)^2 \frac{2p(1-p)}{(p_c - p_t)^2}. \quad (S3)$$

699 In eq. (S3), z_β and $z_{1-\frac{\alpha}{2}}$ are the z-scores, p_t and p_c are the expected proportion of participants at

700 risk in the treatment and control arms, and p is the average of p_t and p_c (52).

701

702 ***Simulation Scenarios***

703 *Effect of Transmission Intensity and Heterogeneity in Biting*

704 The transmission intensity at the trial location may determine the frequency, cause, and

705 detectability of recurrent infections that occur within the duration of follow-up. The number of

706 reinfection events experienced by each trial participant increases with EIR. Thus, at higher EIR,

707 the proportion of recurrent infections that are relapses caused by hypnozoites acquired prior to

708 treatment may be small. Moreover, the levels of anti-parasite and clinical immunity in each trial

709 participant is a function of exposure and transmission intensity, and, as EIR increases, fewer

710 recurrent infections may be detected by light microscopy. Therefore, the transmission intensity at

711 the trial location may bias efficacy estimates, likely downward, due to the rate at which trial

712 participants are reinfected by mosquitoes.

713 Heterogeneity in mosquito biting contributes to observed heterogeneous patterns of

714 *Plasmodium* transmission (16). For *P. falciparum*, heterogeneity in biting biases efficacy

715 estimates towards the null, because fewer participants in the treatment and control arms are

716 exposed within the duration of follow-up (42). For *P. vivax*, however, heterogeneity in biting

717 may reduce the bias due to transmission intensity, because fewer participants are reinfected by

718 mosquitoes within the duration of follow-up.

719 We varied the transmission intensity and the degree of heterogeneity in mosquito biting

720 to quantify their effect on efficacy estimates. We varied transmission intensity by simulating a

721 population-level EIR of 0.1, 1, 10, or 100 infectious bites per person-year (*ibppy*). To account for

722 heterogeneity in biting, we varied the variance of the distribution of logged biting propensities
723 from zero to three. This distribution determines the variation in exposure to mosquito biting
724 across individuals, and a zero variance implies homogeneous biting.
725
726 *Effect of Rate of Relapse and Duration of Follow-up*
727 The rate of *P. vivax* relapse varies geographically, with the estimated mean time to first relapse
728 ranging from 41-65 days for tropical phenotypes to 299 days for temperate phenotypes (17). If
729 the duration of follow-up is too short relative to the mean time to first relapse, then an
730 appreciable number of trial participants may not have relapsed by the duration of follow-up. A
731 longer duration of follow-up may capture a great proportion of relapses but also increases the
732 temporal window over which trial participants can become reinfected by mosquito biting. The
733 probability of reinfection during the duration of follow-up will vary with the EIR.
734 We examined the interaction between the rate of relapse, the duration of follow-up, and
735 transmission intensity and quantified its effect on the efficacy estimates in simulated clinical
736 trials. We fixed the mean time to relapse at 30, 60, 90, or 180 days and simulated clinical trials in
737 which the duration of follow-up was 90, 180, 365, or 730 days. Trial participants were assessed
738 for asexual parasites using light microscopy according to the following schedule: days 8, 15, 22,
739 29, 60, and 90 for the 90-day follow-up period; days 8, 15, 22, 29, 60, 90, 120, and 180 for the
740 180-day follow-up period; days 8, 15, 22, 29, 60, 90, 120, 180, 240, 300, and 365 for the 365-day
741 follow-up period; and days 8, 15, 22, 29, 60, 90, 120, 180, 240, 300, 360, 420, 480, 540, 600,
742 660, and 730 for the 730-day follow-up period. Moreover, to examine the effect of transmission
743 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.

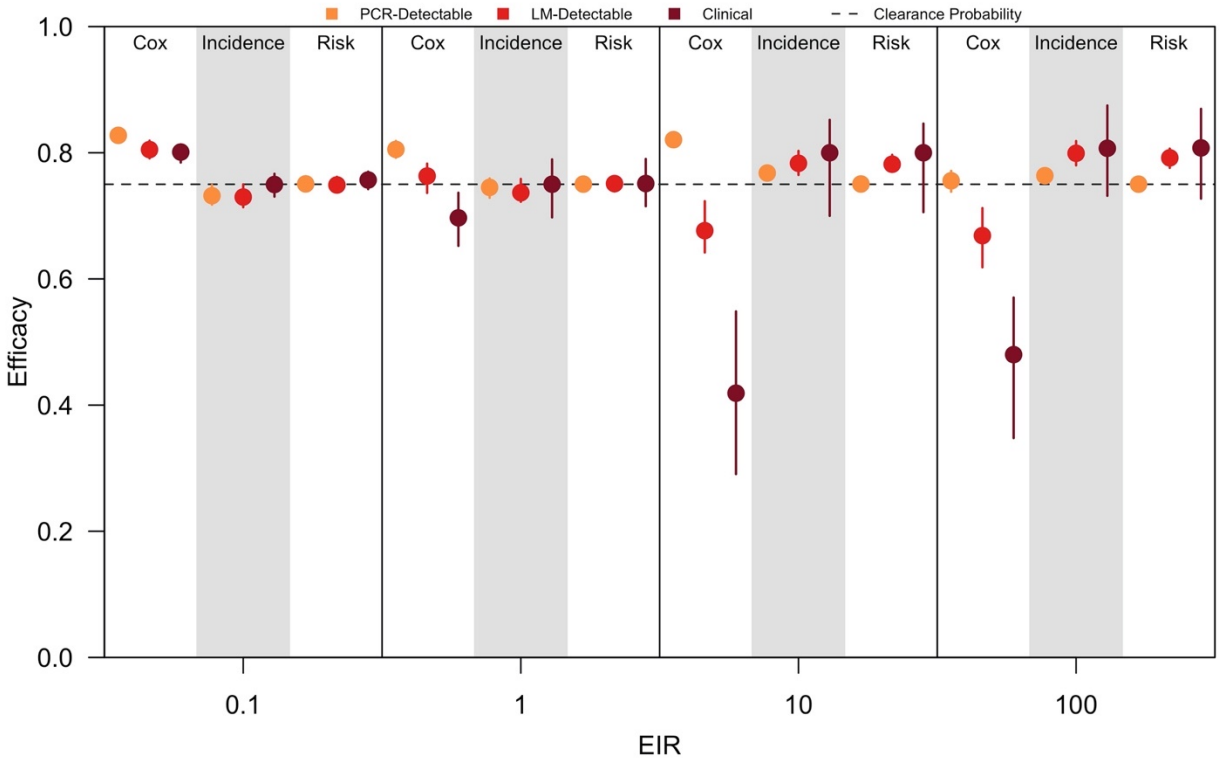
To evaluate whether vector control measures are effective in reducing the bias in the efficacy estimates, we simulated clinical trials under four different vector control scenarios: (1) no vector control; (2) distribution of LLINs; (3) administration of IRS; and (4) distribution of LLINs and administration 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 (Φ_I) and while individuals were in bed (Φ_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

766 Although vector control may decrease the bias by reducing the number of reinfection events, a
767 method capable of distinguishing relapses caused by hypnozoites acquired prior to treatment
768 from all other recurrent infections is needed to completely correct the bias in efficacy estimates.
769 Genotyping methods have been successfully used to determine the cause of each *P. vivax*
770 recurrent infection (12,33) and therefore could correct biases in phase-III clinical trials.
771 We evaluated the potential for a genotyping method at varying sensitivities and
772 specificities to correct the bias in the efficacy estimates. We considered a genotyping method
773 intended to distinguish relapses caused from hypnozoite batches acquired prior to treatment from
774 all other recurrent infections. Therefore, the sensitivity of the genotyping method was the
775 probability of correctly identifying a relapse caused by a hypnozoite batch acquired prior to
776 treatment, and the specificity of the genotyping method was the probability of correctly
777 identifying all other recurrent infections. We simulated clinical trials at population-level EIRs of
778 0.1, 1, 10, and 100, assuming homogeneous biting. We then analyzed the simulated outputs of
779 the clinical trials using a hypothetical genotyping method with sensitivities and specificities of
780 25%, 50%, 75%, or 100%.
781
782 *Effect of the 8-aminoquinoline*
783 By default, we simulated clinical trials for an 8-aminoquinoline with prophylactic effect lasting
784 28 days, consistent with treatment with chloroquine and primaquine. Under this assumption, the
785 duration of prophylaxis is less than the 32-day period of left-censoring, so the suppression of
786 blood-stage parasites during prophylaxis will not bias our efficacy estimates. Treatment with
787 tafenoquine when co-administered with chloroquine, however, provides prophylaxis for 45 days
788 and, therefore, may bias upwards our efficacy estimates (6).

789 To examine whether the duration of prophylaxis affects our estimates of efficacy, we
790 simulated and compared phase-III clinical trials for primaquine and tafenoquine with
791 prophylactic effects lasting 28 and 45 days, respectively. Because the benefit of prophylaxis may
792 vary with transmission intensity, we simulated each trial at population-level EIRs of 0.1, 1, 10,
793 and 100 with homogeneous biting.
794
795 *Effect of the efficacy metric and infection endpoint*
796 The combination of efficacy metric and infection endpoint used to calculate efficacy varied
797 across recent phase-III clinical trials (6–9). Different combinations of efficacy metrics and
798 infection endpoints could make comparison of efficacy estimates across different trials
799 challenging if efficacy estimates are sensitive to this feature of the trial design.
800 To test whether our efficacy estimates are sensitive to the choice of the efficacy metric
801 and the infection endpoint used, we simulated phase-III clinical trials at EIRs of 0.1, 1, 10, and
802 100 and calculated efficacy using different combinations of efficacy metrics and infection
803 endpoints. The efficacy metrics examined were the Cox proportional hazards model, incidence
804 rates, and the proportion at risk. The infection endpoints examined were PCR-detectable
805 recurrent infections, LM-detectable recurrent infections, and clinical recurrent infections.
806 Because we wanted to isolate the effect of the efficacy metric and the infection endpoint, we
807 simulated each trial assuming homogeneous biting.
808
809

810 **Results**



811

812 ***Figure S1. Comparison of efficacy metrics and infection types.*** *Efficacy estimates calculated*
813 *using different methods and using different infection types with perfect genotyping are compared*
814 *across different entomological inoculation rates (EIR). The color of the point denotes the*
815 *recurrent infection endpoint used, where orange is PCR-detectable, red is LM-detectable, and*
816 *maroon is clinical. Because we assumed perfect genotyping, all infections were relapses caused*
817 *by hypnozoite batches acquired prior to 8-aminoquinoline treatment. Each point is the median of*
818 *200 simulations, and the bar is the interquartile range. The dotted line is the true clearance*
819 *probability simulated in each clinical trial.*

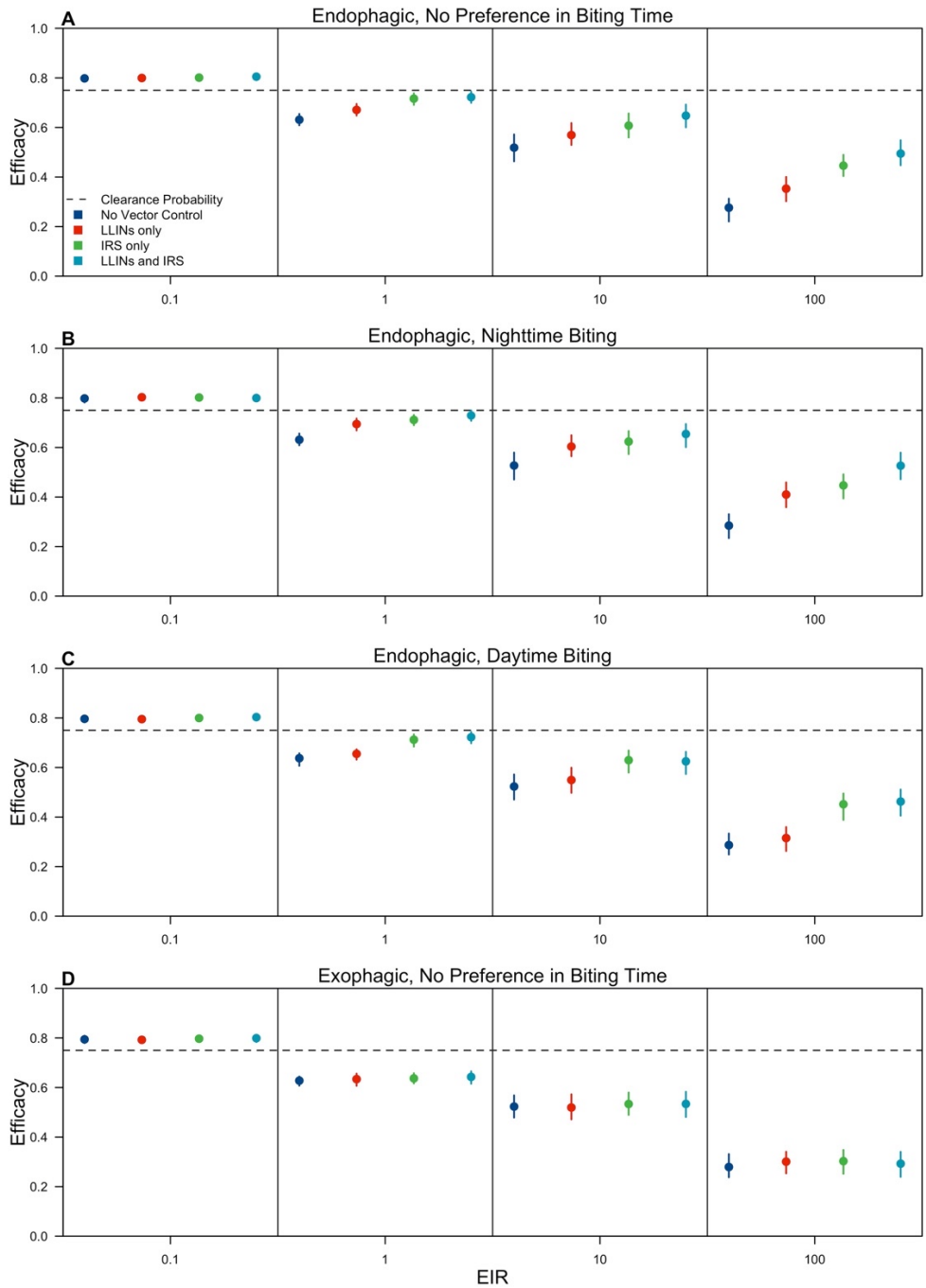


Figure S2. Effect of vector control on efficacy estimates across different vector bionomics. *The impact of LLIN distribution (red), IRS administration (green), and combined LLIN distribution and IRS administration (teal) on resolving biases in LM-detectable recurrence free efficacy estimates is compared 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 (Φ_I) and in bed (Φ_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* $\Phi_B = 0.05$

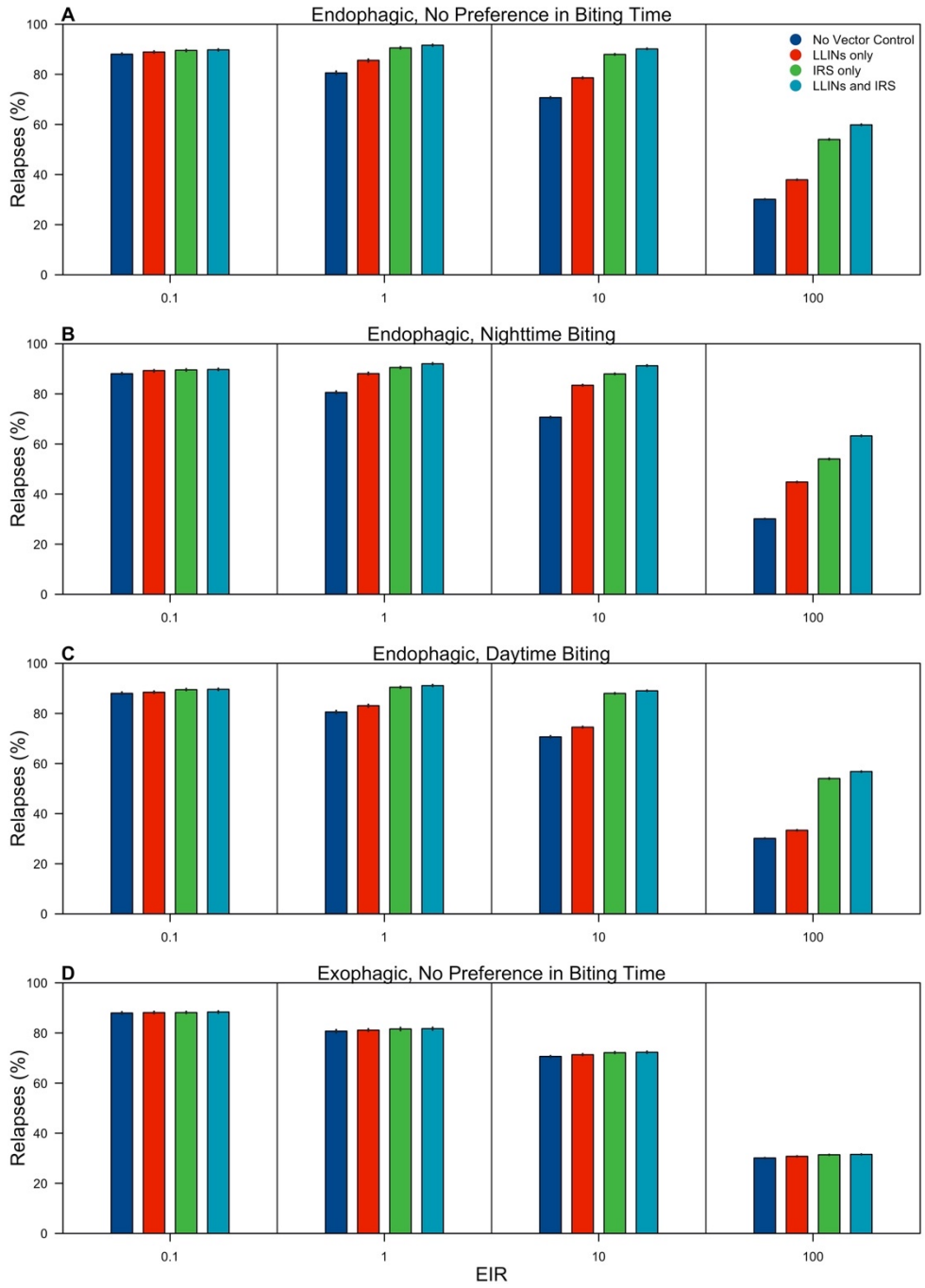
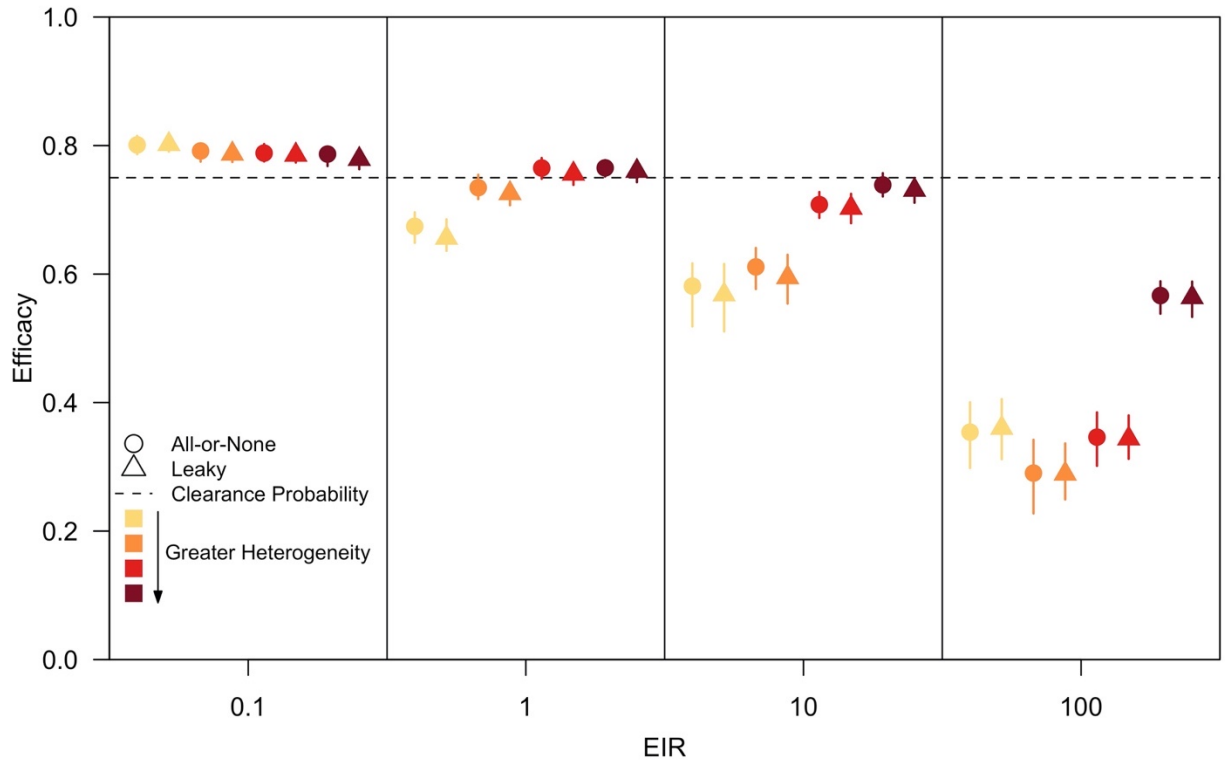
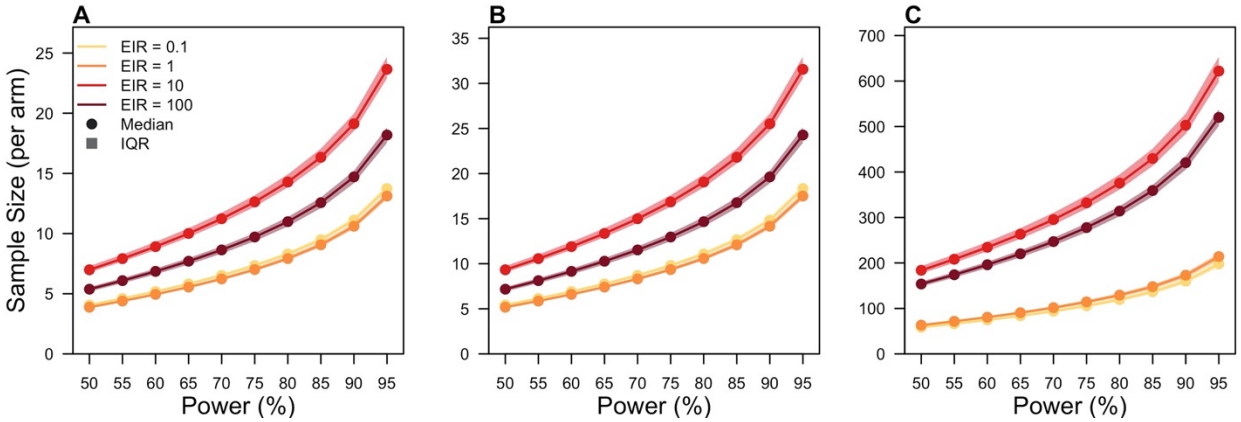


Figure S3. The effect of vector control measures on reducing reinfections across different vector bionomics. *The percentage of recurrent infections in the control arm that are treatment failures (i.e., relapses arising from hypnozoites acquired prior to enrollment) is compared across different entomological 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.*



839

840 ***Figure S4. Comparison of "all-or-none" and "leaky" actions on efficacy estimates.*** *LM-*
841 *detectable recurrence-free efficacy estimates obtained from simulated clinical trials are shown*
842 *for "all-or-none" (circles) and "leaky" (triangles) interventions at different entomological*
843 *inoculation rates (EIR) and levels of heterogeneity in biting. Each point represents the median of*
844 *200 simulations, and each bar is the interquartile range. The color represents the degree of*
845 *heterogeneity in individual-level exposure to mosquito biting with dark colors representing*
846 *greater heterogeneity. The dotted line is the clearance probability simulated in the trials.*



847

848 ***Figure S5. Sample sizes necessary to achieve a specific level of power.*** *For a given EIR, the*
849 *sample size in each arm is reported to achieve a specific level of power when calculating efficacy*
850 *using (A) the Cox proportional hazards model, (B) incidence rates, and (C) the proportion at*
851 *risk. Incidence rates were computed using the number of relapses caused by hypnozoites*
852 *acquired prior to treatment in the control arm. Power calculations were performed for each of*
853 *200 simulations assuming homogeneous biting. The points and line denote the median sample*
854 *size, and the shaded region is the interquartile range. Efficacy of the intervention was assumed*
855 *to be 75%, the type-I error rate was set to 0.05, and the allocation ratio between trial arms was*
856 *one.*

857 **REFERENCES**

858 1. Battle KE, Lucas TCD, Nguyen M, Howes RE, Nandi AK, Twohig KA, et al. Mapping the global
859 endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling
860 study. The Lancet. 2019 Jul;394(10195):332–43.
861 2. Price RN, Commons RJ, Battle KE, Thriemer K, Mendis K. Plasmodium vivax in the Era of the
862 Shrinking P. falciparum Map. Trends in Parasitology. 2020 Jun;36(6):560–70.
863 3. Adekunle AI, Pinkevych M, McGready R, Luxemburger C, White LJ, Nosten F, et al. Modeling the
864 Dynamics of Plasmodium vivax Infection and Hypnozoite Reactivation In Vivo. Sinnis P, editor.
865 PLoS Negl Trop Dis. 2015 Mar 17;9(3):e0003595.
866 4. Commons RJ, Simpson JA, Watson J, White NJ, Price RN. Estimating the Proportion of
867 Plasmodium vivax Recurrences Caused by Relapse: A Systematic Review and Meta-Analysis. The
868 American Journal of Tropical Medicine and Hygiene. 2020 Sep 2;103(3):1094–9.
869 5. World Health Organization, Global Malaria Programme, World Health Organization. Control and
870 elimination of plasmodium vivax malaria: a technical brief [Internet]. 2015 [cited 2020 Nov 17].
871 Available from:
872 http://apps.who.int/iris/bitstream/10665/181162/1/9789241509244_eng.pdf?ua=1&ua=1
873 6. Llanos-Cuentas A, Lacerda MVG, Hien TT, Vélez ID, Namaik-larp C, Chu CS, et al. Tafenoquine
874 versus Primaquine to Prevent Relapse of *Plasmodium vivax* Malaria. N Engl J Med. 2019 Jan
875 17;380(3):229–41.
876 7. Llanos-Cuentas A, Lacerda MV, Rueangweerayut R, Krudsood S, Gupta SK, Kochar SK, et al.
877 Tafenoquine plus chloroquine for the treatment and relapse prevention of Plasmodium vivax
878 malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study.
879 The Lancet. 2014 Mar;383(9922):1049–58.
880 8. Lacerda MVG, Llanos-Cuentas A, Krudsood S, Lon C, Saunders DL, Mohammed R, et al. Single-
881 Dose Tafenoquine to Prevent Relapse of *Plasmodium vivax* Malaria. N Engl J Med. 2019 Jan
882 17;380(3):215–28.
883 9. Taylor WRJ, Thriemer K, von Seidlein L, Yuentrakul P, Assawariyathipat T, Assefa A, et al. Short-
884 course primaquine for the radical cure of Plasmodium vivax malaria: a multicentre, randomised,
885 placebo-controlled non-inferiority trial. The Lancet. 2019 Sep;394(10202):929–38.
886 10. White MT, Walker P, Karl S, Hetzel MW, Freeman T, Waltmann A, et al. Mathematical modelling
887 of the impact of expanding levels of malaria control interventions on Plasmodium vivax. Nat
888 Commun. 2018 Dec;9(1):3300.
889 11. Bennett JW, Pybus BS, Yadava A, Tosh D, Sousa JC, McCarthy WF, et al. Primaquine Failure and
890 Cytochrome P-450 2D6 in *Plasmodium vivax* Malaria. N Engl J Med. 2013 Oct 3;369(14):1381–2.
891 12. Taylor AR, Watson JA, Chu CS, Puaprasert K, Duanguppama J, Day NPJ, et al. Resolving the
892 cause of recurrent Plasmodium vivax malaria probabilistically. Nat Commun. 2019 Dec;10(1):5595.
893 13. Stepniewska K, White NJ. Some considerations in the design and interpretation of antimalarial drug
894 trials in uncomplicated falciparum malaria. Malar J. 2006 Dec;5(1):127.
895 14. Vynnycky E, Sumner T, Fielding KL, Lewis JJ, Cox AP, Hayes RJ, et al. Tuberculosis Control in
896 South African Gold Mines: Mathematical Modeling of a Trial of Community-Wide Isoniazid
897 Preventive Therapy. Am J Epidemiol. 2015 Apr 15;181(8):619–32.
898 15. WorldWide Antimalarial Resistance Network Methodology Study Group. Competing risk events in
899 antimalarial drug trials in uncomplicated Plasmodium falciparum malaria: a WorldWide
900 Antimalarial Resistance Network individual participant data meta-analysis. Malar J. 2019 Jul
901 5;18(1):225.
902 16. Cooper L, Kang SY, Bisanzio D, Maxwell K, Rodriguez-Barraquer I, Greenhouse B, et al. Pareto
903 rules for malaria super-spreaders and super-spreading. Nat Commun. 2019 Dec;10(1):3939.
904 17. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical
905 variation in Plasmodium vivax relapse. Malar J. 2014 Dec;13(1):144.
906 18. Nekkab N, Lana R, Lacerda M, Obadia T, Siqueira AM, Monteiro WM, et al. Estimated impact of

907 tafenoquine for Plasmodium vivax control and elimination in Brazil: A modelling study. PLoS
908 Medicine. 2021 Jan 6;
909 19. Dahal P, Simpson JA, Dorsey G, Guérin PJ, Price RN, Stepniewska K. Statistical methods to derive
910 efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and
911 challenges. Malar J. 2017 Dec;16(1):430.
912 20. Commons RJ, Simpson JA, Thriemer K, Humphreys GS, Abreha T, Alemu SG, et al. The effect of
913 chloroquine dose and primaquine on Plasmodium vivax recurrence: a WorldWide Antimalarial
914 Resistance Network systematic review and individual patient pooled meta-analysis. The Lancet
915 Infectious Diseases. 2018 Sep;18(9):1025–34.
916 21. Pollard EJM, MacLaren D, Russell TL, Burkot TR. Protecting the peri-domestic environment: the
917 challenge for eliminating residual malaria. Sci Rep. 2020 Dec;10(1):7018.
918 22. Monroe A, Moore S, Okumu F, Kiware S, Lobo NF, Koenker H, et al. Methods and indicators for
919 measuring patterns of human exposure to malaria vectors. Malar J. 2020 Dec;19(1):207.
920 23. Moreno JE, Rubio-Palis Y, Páez E, Pérez E, Sánchez V. Abundance, biting behaviour and parous
921 rate of anopheline mosquito species in relation to malaria incidence in gold-mining areas of
922 southern Venezuela. Medical and Veterinary Entomology. 2007 Dec 18;21(4):339–49.
923 24. Paul MR, Grenfell BT, Hoti SL, Ramaiah KD, Bundy DA, Das PK, et al. Quantifying mosquito
924 biting patterns on humans by DNA fingerprinting of bloodmeals. The American Journal of Tropical
925 Medicine and Hygiene. 2001 Dec 1;65(6):722–8.
926 25. Guelbéogo WM, Gonçalves BP, Grignard L, Bradley J, Serme SS, Hellewell J, et al. Variation in
927 natural exposure to anopheles mosquitoes and its effects on malaria transmission. eLife. 2018 Jan
928 23;7:e32625.
929 26. Liebman KA, Stoddard ST, Reiner RC, Perkins TA, Astete H, Sihuincha M, et al. Determinants of
930 Heterogeneous Blood Feeding Patterns by Aedes aegypti in Iquitos, Peru. Barrera R, editor. PLoS
931 Negl Trop Dis. 2014 Feb 13;8(2):e2702.
932 27. Sagna AB, Kibria MG, Naher S, Islam S, Aktaruzzaman MM, Alam MS, et al. Stratifying malaria
933 receptivity in Bangladesh using archived rapid diagnostic tests. Malar J. 2020 Sep 23;19(1):345.
934 28. Poinsignon A, Cornelie S, Ba F, Boulanger D, Sow C, Rossignol M, et al. Human IgG response to a
935 salivary peptide, gSG6-P1, as a new immuno-epidemiological tool for evaluating low-level
936 exposure to Anopheles bites. Malar J. 2009 Aug 13;8:198.
937 29. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, et al. Relapses of
938 *Plasmodium vivax* Infection Usually Result from Activation of Heterologous Hypnozoites. J
939 INFECT DIS. 2007 Apr;195(7):927–33.
940 30. Imwong M, Boel ME, Pagornrat W, Pimanpanarak M, McGready R, Day NPJ, et al. The First
941 Plasmodium vivax Relapses of Life Are Usually Genetically Homologous. The Journal of Infectious
942 Diseases. 2012 Feb 15;205(4):680–3.
943 31. Koepfli C, Mueller I, Marfurt J, Goroti M, Sie A, Oa O, et al. Evaluation of *Plasmodium vivax*
944 Genotyping Markers for Molecular Monitoring in Clinical Trials. J INFECT DIS. 2009
945 Apr;199(7):1074–80.
946 32. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of
947 amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria
948 infections. BMC Genomics. 2017 Dec;18(1):864.
949 33. Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabora O, et al. Using
950 Amplicon Deep Sequencing to Detect Genetic Signatures of *Plasmodium vivax* Relapse. J Infect
951 Dis. 2015 Sep 15;212(6):999–1008.
952 34. Shanks GD, White NJ. The activation of vivax malaria hypnozoites by infectious diseases. The
953 Lancet Infectious Diseases. 2013 Oct;13(10):900–6.
954 35. Mandal S, Sarkar R, Sinha S. Mathematical models of malaria - a review. Malar J. 2011;10(1):202.
955 36. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, et al. Reducing
956 Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention
957 Strategies. Krishna S, editor. PLoS Med. 2010 Aug 10;7(8):e1000324.

37. Halloran ME, Auranen K, Baird S, Basta NE, Bellan SE, Brookmeyer R, et al. Simulations for designing and interpreting intervention trials in infectious diseases. BMC Med. 2017 Dec;15(1):223.
38. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE. Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. Chitnis CE, editor. PLoS Pathog. 2012 Apr 5;8(4):e1002588.
39. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. The Lancet Infectious Diseases. 2009 Sep;9(9):555–66.
40. Halloran ME, Haber M, Longini IM. Interpretation and Estimation of Vaccine Efficacy under Heterogeneity. American Journal of Epidemiology. 1992 Aug 1;136(3):328–43.
41. Smith PG, Rodrigues LC, Fine PEM. Assessment of the Protective Efficacy of Vaccines against Common Diseases Using Case-Control and Cohort Studies. Int J Epidemiol. 1984;13(1):87–93.
42. White MT, Griffin JT, Drakeley CJ, Ghani AC. Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. Malar J. 2010 Dec;9(1):82.
43. Gaedigk A, Dinh JC, Jeong H, Prasad B, Leeder JS. Ten Years' Experience with the CYP2D6 Activity Score: A Perspective on Future Investigations to Improve Clinical Predictions for Precision Therapeutics. J Pers Med. 2018 Apr 17;8(2).
44. Pal S, Bansil P, Bancone G, Hrutkay S, Kahn M, Gornsawun G, et al. Evaluation of a Novel Quantitative Test for Glucose-6-Phosphate Dehydrogenase Deficiency: Bringing Quantitative Testing for Glucose-6-Phosphate Dehydrogenase Deficiency Closer to the Patient. Am J Trop Med Hyg. 2019;100(1):213–21.
45. Halloran ME, Longini IM, Struchiner CJ. Design and Analysis of Vaccine Studies [Internet]. New York, NY: Springer New York; 2010 [cited 2020 Oct 25]. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-0-387-68636-3
46. Carnevale P, Frézil JL, Bosseno MF, Le Pont F, Lancien J. [The aggressiveness of Anopheles gambiae A in relation to the age and sex of the human subjects]. Bull World Health Organ. 1978;56(1):147–54.
47. Port GR, Boreham PFL, Bryan JH. The relationship of host size to feeding by mosquitoes of the *Anopheles gambiae* Giles complex (Diptera: Culicidae). Bull Entomol Res. 1980 Mar;70(1):133–44.
48. Herrera S, Solarte Y, Jordán-Villegas A, Echavarría JF, Rocha L, Palacios R, et al. Consistent safety and infectivity in sporozoite challenge model of Plasmodium vivax in malaria-naive human volunteers. Am J Trop Med Hyg. 2011 Feb;84(2 Suppl):4–11.
49. Kiattibutr K, Roobsoong W, Sriwichai P, Saeseu T, Rachaphaew N, Suansomjit C, et al. Infectivity of symptomatic and asymptomatic Plasmodium vivax infections to a Southeast Asian vector, Anopheles dirus. Int J Parasitol. 2017;47(2–3):163–70.
50. White MT, Griffin JT, Churcher TS, Ferguson NM, Basáñez M-G, Ghani AC. Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. Parasites Vectors. 2011 Dec;4(1):153.
51. Schoenfeld DA. Sample-Size Formula for the Proportional-Hazards Regression Model. Biometrics. 1983 Jun;39(2):499.
52. Smith PG, Morrow RH, Ross DA, International Epidemiological Association, Wellcome Trust (London, England), editors. Field trials of health interventions: a toolbox. 3rd edition. Oxford: Oxford University Press; 2015. 444 p.