How radical is radical cure? 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)

Short Title: Site-specific biases in phase-III clinical trials of radical cure

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 12-78%, despite simulating trial data under the uniform assumption that treatment had a 75% chance of clearing hypnozoites. A longer duration of prophylaxis leads to a greater measured efficacy, particularly at higher transmission intensities, making the comparison of the protection of different radical cure treatment regimens against relapse more challenging. We show that vector control and parasite genotyping offer two potential means to yield more standardized efficacy estimates that better reflect protection against relapse. We predict that site-specific biases are likely to contribute to variation in efficacy estimates both within and across phase-III clinical trials. Future clinical trials can reduce site-specific biases by conducting trials in low-transmission settings where reinfections from mosquito biting are less common, by preventing reinfections using vector control measures, or by identifying and excluding likely reinfections that occur during follow-up using parasite genotyping methods.

AUTHOR SUMMARY

Radical cure holds promise as a strategy for *Plasmodium vivax* malaria control by clearing the parasites known as hypnozoites that latently infect the liver and cause relapsing infections. The efficacy of radical cure treatment regimens is evaluated in phase-III clinical trials. Recent trial results have noted substantial variation in efficacy estimates across trial sites, complicating the interpretation of the benefit of radical cure. However, *P. vivax* infections identified during the course of the clinical trial could include reinfections from mosquito biting that do not directly reflect the effect of the therapeutic being trialed, potentially biasing efficacy estimates. In this study, we simulated clinical trials to identify the causes and solutions of these site-specific biases. We found that features of both the trial location, such as the transmission intensity, and the trial design, such as the duration of follow-up, lead to an underestimate of the effect of radical cure against hypnozoites. We then demonstrated that vector control and parasite genotyping are two possible strategies to reduce these biases. These insights can be leveraged to aid in the interpretation of past trial results and to help design future clinical trials that minimize site-specific biases.

51 **INTRODUCTION**

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

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

70 The primary endpoint used in recent phase-III clinical trials is freedom from *P. vivax*
71 recurrent infection (6–8), which encompasses relapses caused by hypnozoite activation,
72 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 various epidemiological features, including 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 the ability to measure the clearance probability—has not been explored.

Here, we extend an existing and validated individual-based model of *P. vivax* transmission (10,18) to simulate phase-III clinical trials for radical cure. Individual-based models have been used to design and evaluate clinical trials, both for malaria (19,20) and other infectious diseases (21–24). 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.

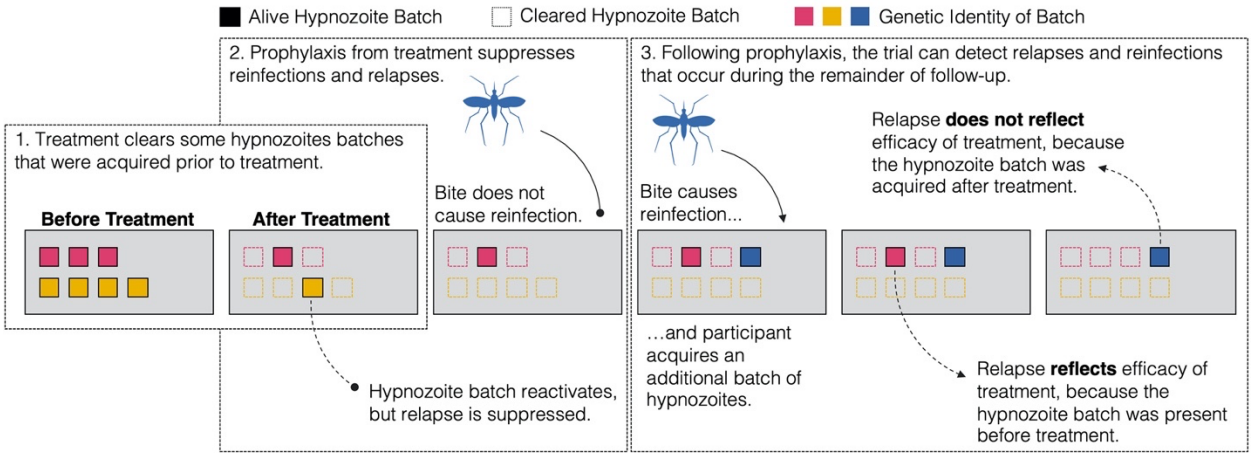

Fig 1. Schematic of potential infection outcomes during trial follow-up. *Participants are enrolled in the trial and initially harbor hypnozoite batches (colored squares). The colors represent the genetic identity of each hypnozoite batch. Hypnozoite batches are cleared (unfilled dotted squares) with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

In this analysis, we simulated phase-III clinical trials of an 8-aminoquinoline with an "all-or-none" intervention action and a mean clearance probability of 0.75 across the population. Under this assumption, treatment with the 8-aminoquinoline on average completely cleared hypnozoite batches in 75% of individuals and had no effect on hypnozoite batches in the remaining 25% of individuals. By default, each simulated clinical trial was modeled after the DETECTIVE trials (6–8). Unless otherwise specified, trial participants were followed for 180 days and assessed for clinical and light microscopy (LM)-detectable recurrent infections, and the mean time to first relapse of each *P. vivax* hypnozoite batch was 65 days. Consistent with the DETECTIVE trial (6–8), recurrent infections were left-censored by 32 days to exclude recrudescences caused by blood-stage therapeutic failure, though these infection types were not simulated in the model. A full description of the trial simulations can be found in the Methods and Supplement.

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) to 1, 10, or 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 0.70 (IQR: 0.68 – 0.72) at an

EIR of 1 to 0.38 (0.32 – 0.43) at an EIR of 100. This downward bias in efficacy occurred due to both frequent reinfection–an infection event that does not reflect the clearance probability–and the reduced detectability of recurrent infections due to higher levels of anti-parasite immunity at higher transmission intensities. As transmission intensity increases, more trial participants were reinfected by mosquitoes, yet parasite densities were reduced and fewer recurrent infections were LM-detectable (Fig. S1, left column), causing the infection profiles of the treatment and control arms to appear more similar and leading to a lower measured efficacy.

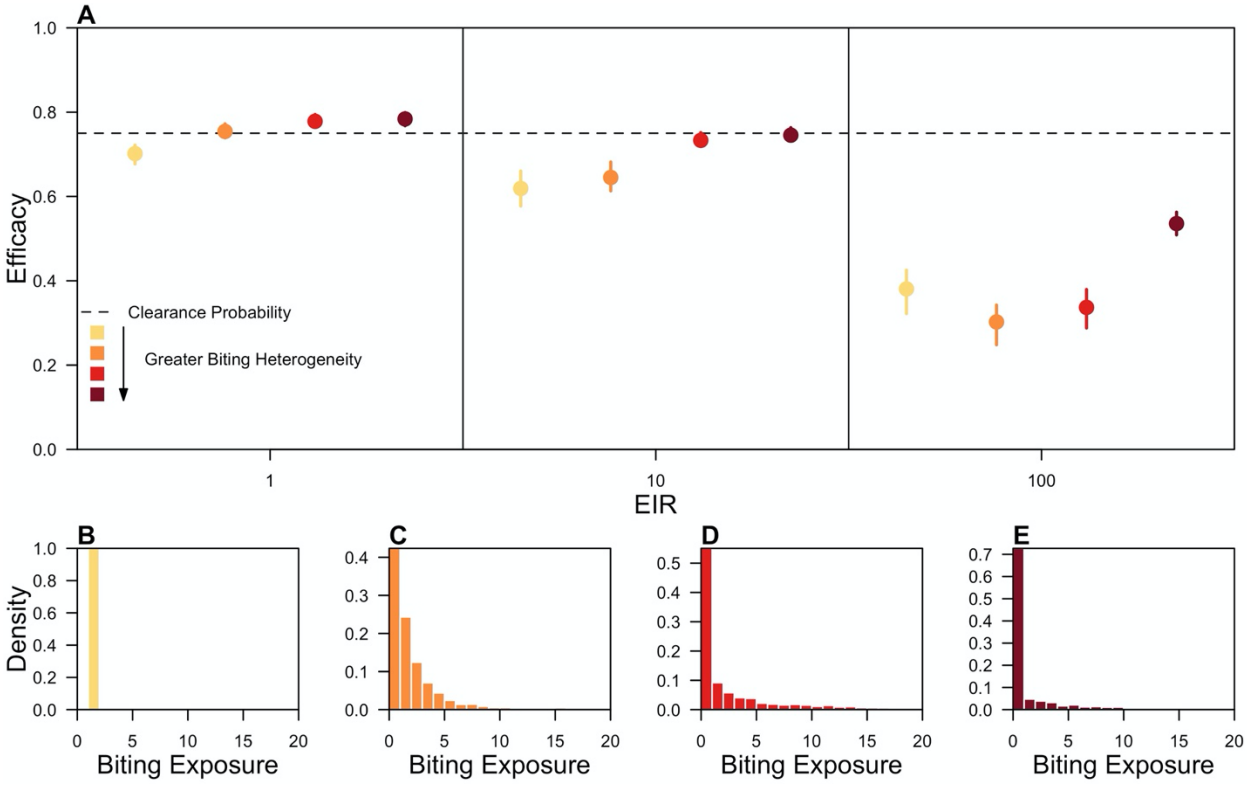

Fig 2. Effect of transmission intensity and heterogeneity in biting on efficacy estimates. *(A) LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIR) and levels of heterogeneity in biting. Each point represents the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

157

158 At low transmission intensities when reinfection rates were low and heterogeneity in
159 biting was high, our efficacy estimates exceeded the clearance probability, revealing a positive
160 bias. This positive bias occurred because we assumed that radical cure completely cleared
161 hypnozoite batches in a subset of trial participants and did not clear hypnozoite batches in the
162 remainder of trial participants. In trials measuring time to first recurrent infection, those
163 participants that completely clear hypnozoite batches increase the total person-time at risk in the
164 treatment arm, reducing the estimated hazard of recurrent infection in the treatment arm and
165 leading to a competing positive bias that was present in all trial simulations. The magnitude of
166 this bias increases with follow-up time, provided that participants are not reinfected by mosquito
167 biting, and was eliminated if efficacy was instead calculated based upon the proportion at risk
168 using the complete record of recurrent infections (Fig. S2).

169 In addition to the effect of transmission intensity itself, the effect of heterogeneous biting
170 on efficacy estimates was modulated by the transmission intensity of the trial location. At EIRs
171 of 1 and 10, the downward bias was reduced with greater heterogeneity in biting. Efficacy
172 measured at an EIR of 10 ranged from 0.62 (0.58 – 0.66) under homogeneous biting to 0.75
173 (0.73 – 0.76) under the highest level of heterogeneity in biting simulated. Heterogeneous biting
174 reduced the downward bias caused by frequent reinfection, because fewer reinfections from
175 mosquito biting occurred during follow-up. As exposure decreased on average with greater
176 heterogeneity in biting, anti-parasite immunity levels were lower, and a greater proportion of
177 recurrent infections were LM-detectable (Fig. S1). The improved detectability of recurrent
178 revealed the differences in the infection profiles across the trial arms, reducing the downward
179 bias due to transmission intensity. At an EIR of 100, the effect of heterogeneous biting on

180 efficacy estimates was non-monotonic (Fig. 2A). Although the percentage of participants
181 reinfected during follow-up decreased monotonically with greater heterogeneity in mosquito
182 biting, the percentage of participants with a detectable reinfection (e.g., clinical or
183 subclinical/LM-detectable) varied non-monotonically and was 36% (35 – 36%), 39% (38 –
184 40%), 35% (35 – 36%), and 25% (24 – 25%) when the respective variance in exposure to biting
185 was 0, 1, 2, and 3 (Fig. 2B-E).
186
187 *Effect of the Rate of Relapse and Duration of Follow-up*
188 We next examined how the rate of relapse of *P. vivax* hypnozoites and the duration of follow-up
189 of trial participants interact to affect efficacy estimates. We considered mean times to relapse of
190 30, 60, 90, and 180 days and simulated phase-III clinical trials in which participants were
191 followed for 90, 180, 365, or 730 days. To assess the effect of transmission intensity, all trials
192 were simulated at EIRs of 1, 10, and 100, assuming homogeneous biting.
193 For a given duration of follow-up, efficacy estimates decreased with a longer mean time
194 to first relapse (Fig. 3A). In simulated trials in which participants were followed for 180 days at
195 an EIR of 1, we estimated that the efficacy of an 8-aminoquinoline with a clearance probability
196 of 0.75 decreased from 0.74 (IQR: 0.72 – 0.75) to 0.63 (0.60 – 0.66) as the mean time to first
197 relapse increased from 30 to 180 days. This downward bias occurred because, as the mean time
198 to first relapse increased, fewer trial participants were expected to relapse within the duration of
199 follow-up (Fig. 3B), causing the distribution of times to first recurrent infection to appear more
200 similar across trial arms (Fig. S3, top row). Only 63.2% of participants were expected to relapse
201 by 180 days when the mean time to first relapse was 180 days, compared to 99.8% of
202 participants when the time to first relapse was 30 days.

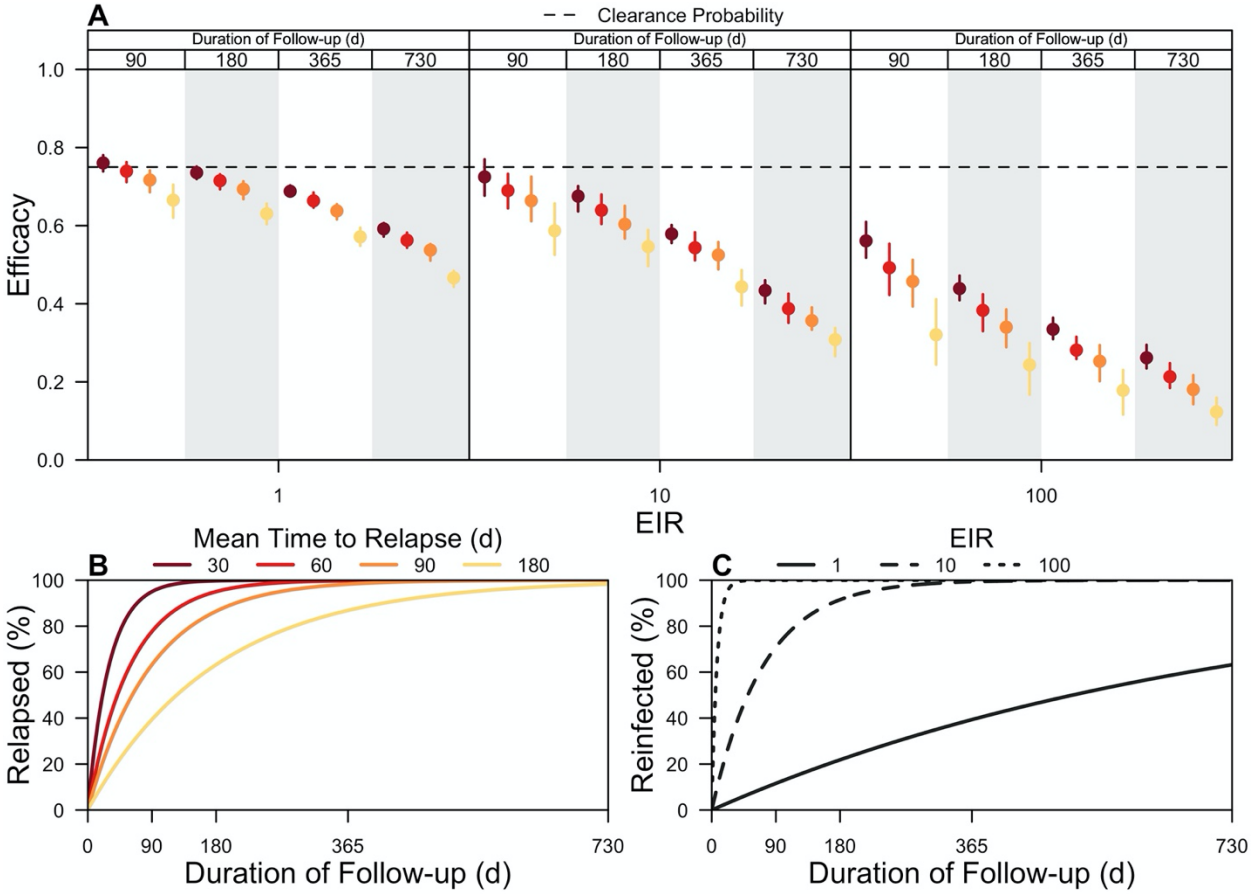

Fig 3. Effect of rate of relapse and duration of follow-up on efficacy estimates. *(A) LM-detectable recurrence-free efficacy estimated from simulated clinical trials is shown at different entomological inoculation rates (EIRs), mean times to relapse, and durations of follow-up. Each point is the median of 200 simulations, and each bar is the interquartile range. The color represents the mean time to relapse (30, 60, 90, or 180 days), 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.*

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 0.76 (0.74 – 0.78) when trial participants were followed for 90 days, compared to 0.59 (0.57 – 0.61) when trial participants

220 were followed for 730 days. Under an alternative scenario in which the EIR was 10, the
221 estimated efficacies were 0.72 (0.68 – 0.77) and 0.43 (0.40 – 0.46) at the respective durations of
222 follow-up. The downward bias occurred because a longer duration of follow-up ensured that
223 more trial participants were reinfected by mosquito biting (Fig. 3C), similarly causing the
224 distribution of times to first recurrent infection to appear more similar across the trial arms. By
225 90 days of follow-up, 12% and 71% of trial participants were expected to have been reinfected
226 when the respective EIRs were 1 and 10. By 730 days of follow-up, 63% and 100% of trial
227 participants were now expected to have been reinfected at respective EIRs of 1 and 10.
228
229 *Effect of Vector Control*
230 Higher transmission intensity biased our efficacy estimates downward, because trial participants
231 in both the treatment and control arms were frequently reinfected (Fig. 2A). We assessed the
232 potential of long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) to protect
233 trial participants from reinfection and therefore reduce the downward bias due to transmission
234 intensity. Because the effects of LLINs and IRS depend upon vector bionomics, we varied the
235 absolute proportion of bites that occurred indoors and the absolute proportion of bites that
236 occurred while trial participants were in bed.
237 Vector control interventions were effective in reducing the downward bias due to
238 transmission intensity in trial locations where the vector was predominantly endophagic (i.e.,
239 feeds indoors) (Fig. 4). With an endophagic vector with no preference in biting time and at an
240 EIR of 10, the estimated efficacies improved from 0.58 (0.53 – 0.62) in the absence of vector
241 control to 0.63 (0.58 – 0.67) with LLINs, 0.69 (0.64 – 0.73) with IRS, and 0.70 (0.66 – 0.75)
242 with the combination of LLINs and IRS. By comparison, with an exophagic (i.e., feeds outdoors)

vector with no preference in biting time and at an EIR of 10, efficacy estimates improved only slightly from 0.57 (0.52 – 0.61) in the absence of vector control, 0.57 (0.52 – 0.61) with LLINs, and 0.57 (0.52 – 0.61) with IRS to 0.58 (0.55 – 0.63) with the combination of LLINs and IRS (Fig. S4D).

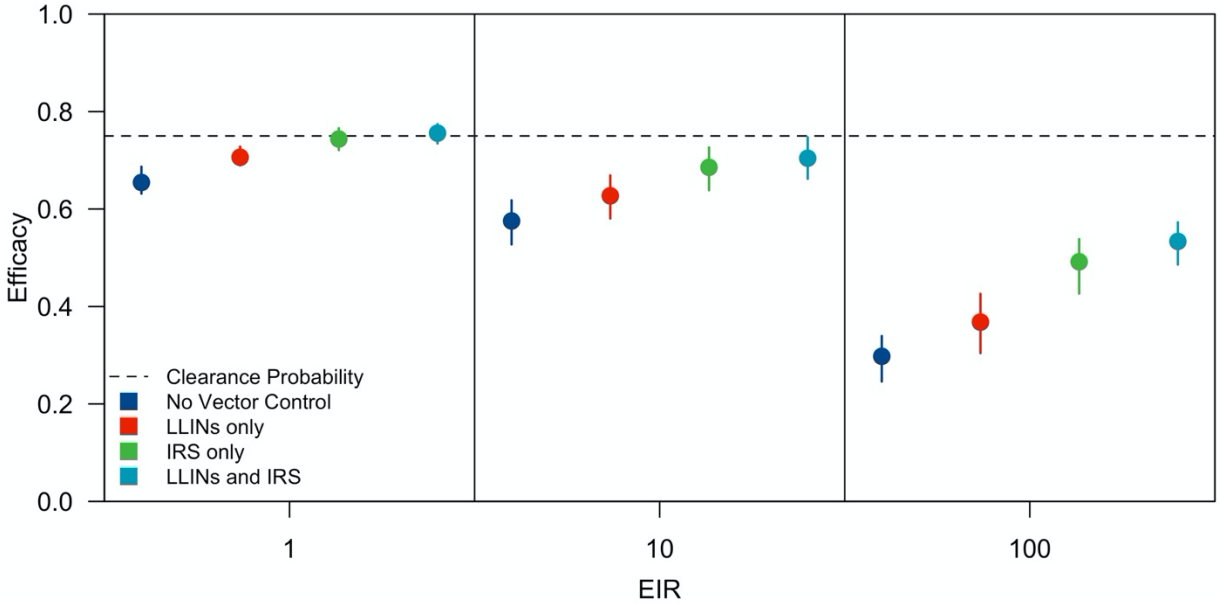

Fig 4. Effect of vector control on efficacy estimates. *The impact of LLIN distribution (red), IRS administration (green), and combined LLIN distribution and IRS administration (teal) on LM-detectable recurrence free efficacy estimates is compared to a no-intervention scenario (dark blue) across a range 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.*

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. S4). When considering interventions in isolation, IRS was more effective than LLINs in reducing the downward bias, and the distribution of LLINs most improved

261 efficacy estimates under scenarios in which an appreciable number of mosquito bites were taken
262 at night while trial participants were in bed.
263
264 *Effect of Parasite Genotyping*
265 Unlike vector control interventions that reduce the downward bias in efficacy by reducing the
266 number of reinfection events among trial participants (Fig. S5), parasite genotyping used in
267 combination with time-to-event data could correct or reduce the downward bias by
268 distinguishing recurrent infection events that directly reflect the clearance probability (i.e.,
269 relapses arising from hypnozoite batches acquired prior to treatment) from all other recurrent
270 infections. We assessed the potential of a generic method leveraging genotyping and time-to-
271 event data to correct the bias in efficacy estimates at different transmission intensities and
272 considered how performance characteristics of this method affected the extent to which the bias
273 was corrected by varying its sensitivity and specificity. For the purpose of this simulation study,
274 we defined sensitivity as the probability of correctly identifying relapses caused by hypnozoite
275 batches acquired prior to treatment and specificity as the probability of correctly identifying all
276 other recurrent infections (e.g., reinfections and relapses caused by hypnozoite batches acquired
277 after treatment).
278 At lower and intermediate transmission intensities (i.e., EIRs of 1 and 10), efficacy
279 estimates changed more with improved sensitivity than improved specificity of the method (Fig.
280 5). By contrast, at higher transmission intensities, high specificity was needed to reduce the
281 downward bias in efficacy. At an EIR of 100 and assuming 100% sensitivity of the method, the
282 estimated efficacy of an 8-aminoquinoline with clearance probability equal to 0.75 improved
283 from 0.41 (0.37 – 0.47) at 25% specificity to 0.78 (0.73 – 0.81) at 100% specificity. Under the

alternative scenario in which specificity was 100%, the estimated efficacy at the respective EIR did not significantly change with improved sensitivity and was 0.78 (0.69 – 0.85) at 25% sensitivity and 0.78 (0.73 – 0.81) at 100% sensitivity. The specificity 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) led 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.

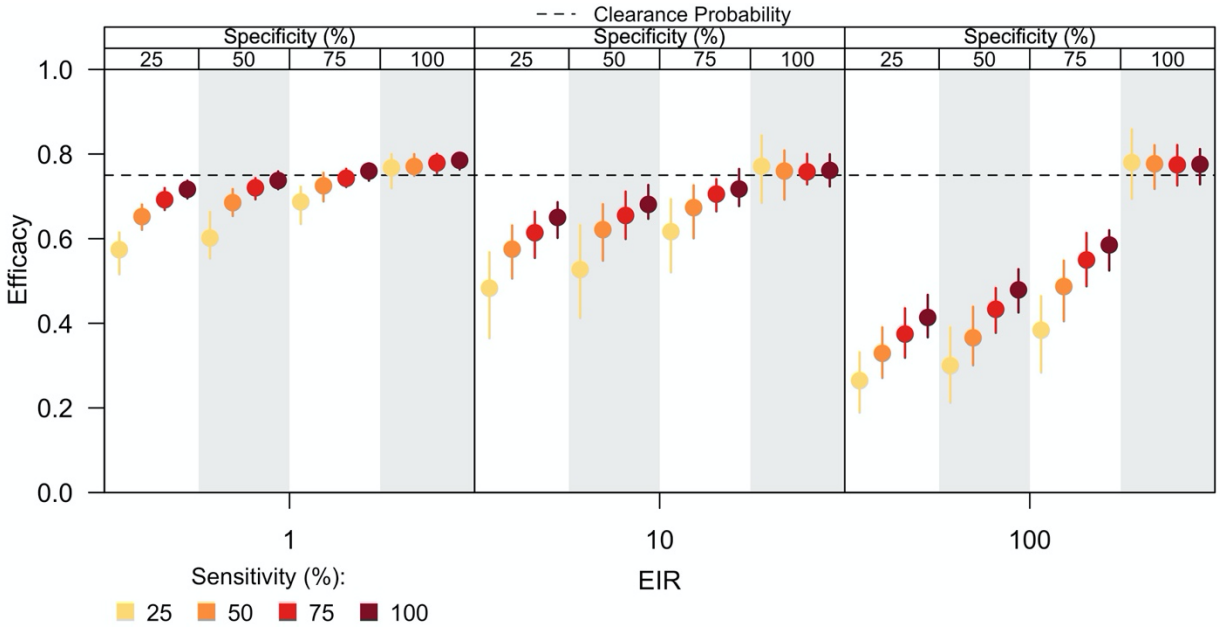

Fig 5. Effect of parasite genotyping on efficacy estimates. *The impact of genotyping recurrent infections to estimate the efficacy of radical cure is shown for different sensitivities and specificities of the genotyping method across a range of entomological inoculation rates (EIR). Sensitivity of the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

305

306 *Effect of the 8-aminoquinoline*

307 The previous analyses considered phase-III trials of an 8-aminoquinoline that provided
308 prophylaxis for 28 days, a duration of time less than the 32-day period of left-censoring used to
309 calculate efficacy. Consequently, prophylaxis did not impact our efficacy estimates. To examine
310 whether a longer duration of prophylaxis biased our efficacy estimates, we simulated phase-III
311 trials for two different 8-aminoquinolines, PQ and TFQ, that provide prophylactic effects for
312 fixed periods of 28 and 45 days, respectively. Because the benefit of prophylaxis increases with
313 transmission intensity, we simulated trials at EIRs equal to 1, 10, and 100, assuming
314 homogeneous biting.

315 A longer duration of prophylaxis biased our efficacy estimates upward, though the
316 magnitude of the bias depended upon transmission intensity (Fig. 6). In a low-transmission
317 setting (i.e., EIR of 1), our efficacy estimates did not change much with an increased duration of
318 prophylaxis, because a median of 1.3% (1.1 – 1.6%) of participants in the control arm were
319 reinfected between days 32 and 45 post-enrollment—the time period over which TFQ provided
320 prophylactic effects not accounted for by left-censoring. Consequently, at an EIR of 1, we
321 estimated that the efficacies of TFQ and PQ were 0.72 (0.69 – 0.74) and 0.70 (0.68 – 0.72),
322 respectively. At higher EIRs, we estimated a greater efficacy for TFQ than PQ, because
323 participants in the control arm were frequently reinfected whereas participants in the treatment
324 arm were protected from reinfection. On average, the percentage of participants in the control
325 arm who were reinfected between 32- and 45-days post-enrollment was 10% (9.5 – 11%) and
326 54% (52 – 55%) at EIRs of 10 and 100, respectively. Consequently, we estimated that the

respective efficacies of TFQ and PQ were 0.65 (0.62 – 0.69) and 0.63 (0.57 – 0.67) at an EIR of 10 and 0.46 (0.41 – 0.49) and 0.36 (0.32 – 0.41) at an EIR of 100.

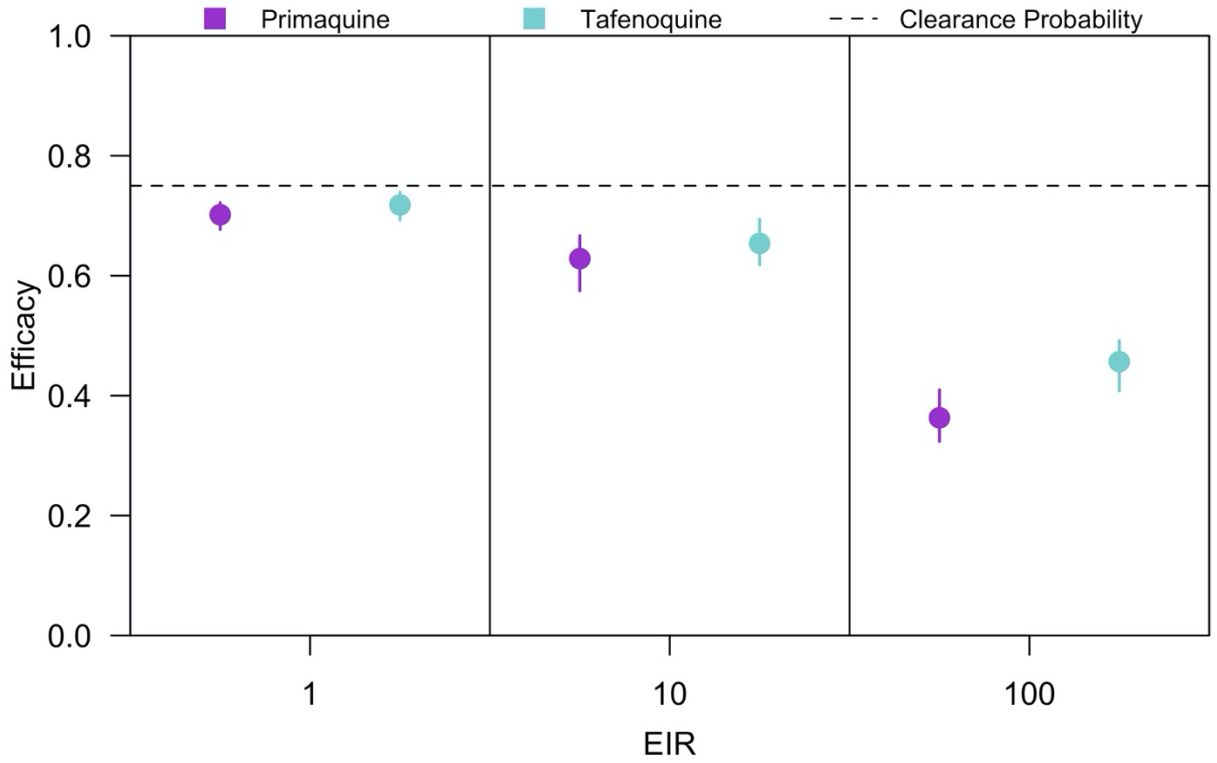

Fig 6. Effect of 8-aminoquinoline on efficacy estimates. *LM-detectable recurrence free efficacy estimated from simulated clinical trials is shown for primaquine (purple) and tafenoquine (blue) at different entomological inoculation rates (EIR). Each point is the median of 200 simulations, and each bar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

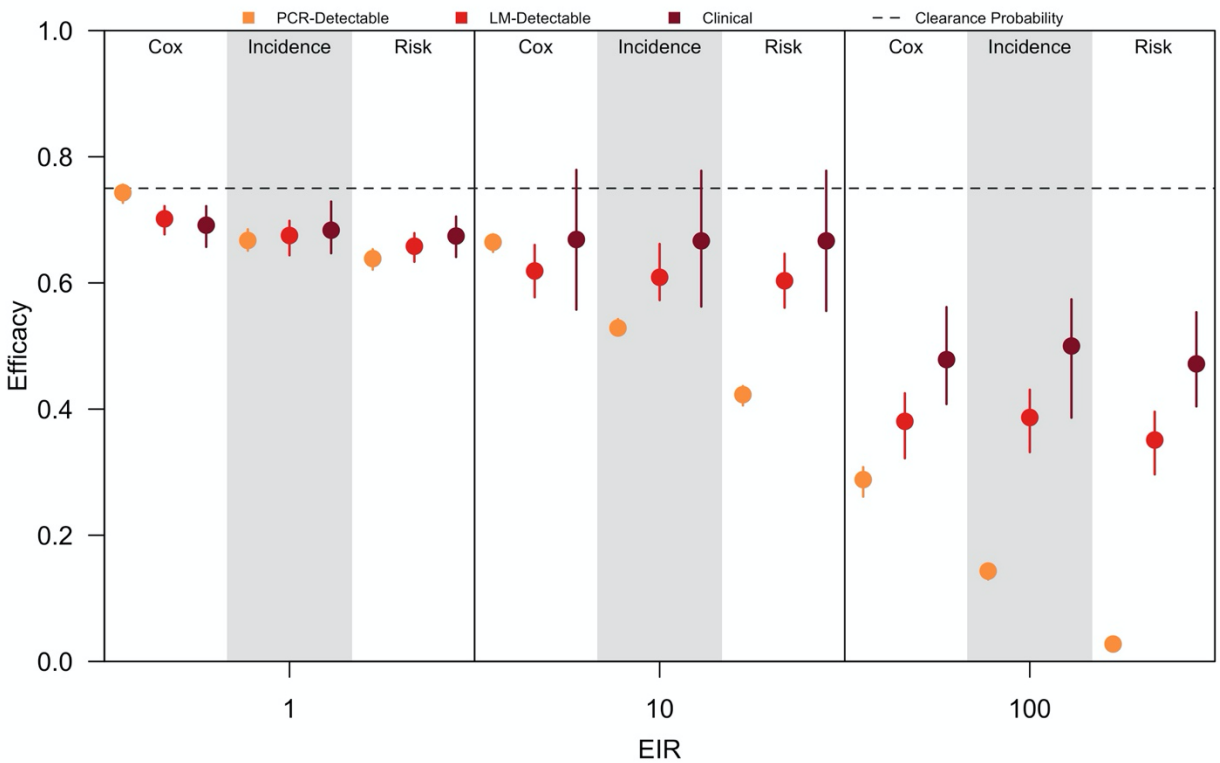

Fig 7. Effect of efficacy metric and infection endpoint on efficacy estimates. *Efficacy estimates obtained from simulated clinical trials at different entomological inoculation rates (EIR) is shown when calculated using the Cox proportional hazards model, incidence rates, or the proportion at risk. The infection endpoint was clinical (maroon), LM-detectable (red), or PCR-detectable (orange) recurrent infections identified during follow-up. 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.*

361

362 The sensitivity of the assay by which trial participants were assessed for recurrent
363 infections also affected our efficacy estimates, though the direction of the effect depended upon
364 the chosen efficacy metric. For efficacy estimates based upon incidence rates or the proportion at
365 risk, shifting from a more sensitive assay (i.e., PCR-based detection) to a less sensitive assay
366 (i.e., monitoring for clinical infections) reduced the downward bias due to reinfection,
367 particularly at higher transmission intensities. This was due to the fact that the fraction of
368 recurrent infections ascertained increased with a more sensitive assay, causing us to detect more
369 recurrent infections during follow-up and making the incidence rates and proportions at risk
370 appear more similar across the treatment and control arms. By contrast, using the Cox
371 proportional hazards model, the relationship between efficacy estimates and the assay sensitivity
372 depended upon transmission intensity. Across all transmission intensities, clinical infections both
373 occurred less frequently and were detected later than PCR- and LM-detectable infections.
374 Therefore, at an EIR of 1 where reinfection from mosquito biting was less frequent, a more
375 sensitive assay more accurately captured differences in the timing of recurrent infections across
376 trial arms that was attributable to the effect of radical cure. By contrast, at an EIR of 100, nearly
377 all trial participants were reinfected during follow-up. At this transmission intensity, a more
378 sensitive assay instead made the timing of recurrent infections appear more similar across trial
379 arms. Thus, using a less sensitive assay improved efficacy estimates. The non-monotonic
380 relationship observed at the intermediate transmission intensity (i.e., EIR of 10) reflects a change
381 between these two extremes.

382

383 **Table 1. Summary of identified biases.** A complete description, including the direction and
384 cause, of the biases identified in this analysis are provided.

Feature of Trial Setting	Direction of Bias	Cause of Bias
Transmission Intensity (EIR)	Downwards	At higher transmission intensities, there is more frequent reinfection and reduced detectability of recurrent infections.
Relapse Rate	Downwards	With a longer time to first relapse, fewer trial participants relapse during follow-up
Duration of Follow-up	Downwards	With a longer duration of follow-up, more trial participants are reinfected by mosquito biting.
Duration of Prophylaxis	Upwards	With a longer duration of prophylaxis, fewer trial participants are reinfected by mosquito biting.
Efficacy Metric	Downwards	Measured efficacy is lower when based upon incidence rates or proportion at risk than upon proportional hazards.
Assay Sensitivity	Downward	For incidence rates and the proportion at risk, a more sensitive assay detects more recurrent infections.
Intervention Action	Upwards	Efficacy based upon proportional hazards is biased upwards for an "all-or-none" intervention, because the "all" group increases the person-time at risk.

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 estimates of efficacy and the utility of clinical trial data to assess the extent to which radical cure prevents relapse (Table 1). 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.

397 In the recent GATHER, DETECTIVE, and IMPROV trials (6–9), efficacy estimates
398 varied across trial sites. Interpreting site-specific differences in efficacy is important for past and
399 future clinical trials (25), and our simulation results suggest that the differences in efficacy
400 estimates could be caused by site-specific biases that arise when the infection endpoint used to
401 measure efficacy does not directly reflect the action of the therapeutic being trialed. If the trial
402 sites vary in transmission intensity, our results suggest that efficacy estimates could be lower in
403 high-transmission settings, because trial participants are more frequently reinfected during
404 follow-up, and the detectability of their recurrent infections is reduced. Greater heterogeneity in
405 mosquito biting reduces this downward bias by reducing the number of trial participants who are
406 bitten and possibly reinfected during follow-up, though this effect is non-monotonic at very high
407 transmission intensities due to the predicted interaction with anti-parasite immunity. Trial sites
408 may also differ in the rates at which *P. vivax* hypnozoites relapse (17). A slower rate of relapse
409 (i.e., a longer time to relapse) implies that an appreciable number of trial participants will not yet
410 have relapsed during follow-up, thereby biasing efficacy estimates downward (26). Increasing
411 the duration of follow-up further compounds the bias rather than corrects it, because a long
412 duration of follow-up causes more participants in both the treatment and control arms to become
413 reinfected by mosquito biting, particularly at higher transmission intensities. Finally, frequent
414 reinfection at higher transmission intensities makes the comparison of treatment regimens with
415 different durations of prophylaxis challenging, because a longer duration of prophylaxis prevents
416 more individuals in the treatment arm from reinfection. Ignoring differences in prophylaxis may
417 lead to an overestimate of the effect against hypnozoites for treatment regimens that provide a
418 longer period of prophylaxis, such as tafenoquine co-administered with chloroquine, relative to

419 other treatment regimens that provide a shorter period of prophylaxis, such as primaquine co-
420 administered with chloroquine.

421 In general, there was greater bias in high-transmission settings, because of the higher risk
422 of recurrent infections being due to reinfections from new mosquito bites. In low-transmission
423 settings, this bias is reduced, although there is the concomitant challenge of recruiting sufficient
424 trial participants. Our simulation results suggest that trial investigators can reduce bias by
425 prospectively preventing reinfections with vector control or by retrospectively accounting for
426 reinfections by leveraging time-to-event and parasite genotyping data.

427 LLINs and IRS were predicted to be most effective in trial settings with an endophagic
428 vector that takes most bites indoors, so novel vector control interventions, such as spatial
429 repellents, may be needed in settings with an exophagic vector to reduce peri-domestic biting
430 (27). Distribution of LLINs occurred as part of the DETECTIVE trial (6–8), suggesting the
431 feasibility of implementing vector control in a trial context. Entomological data collected prior to
432 trial enrollment could characterize the local vector bionomics and inform the selection of
433 interventions (28,29), and blood meal analysis (30–32) or serological assays (33,34) performed
434 as part of the trial could quantify heterogeneous feeding, a factor that directly affects the
435 magnitude of the site-specific bias.

436 Our results demonstrate that parasite genotyping could be used in concert with time-to-
437 event data to generate less biased efficacy estimates that better reflect the effect of the 8-
438 aminoquinoline against hypnozoite batches. Although in practice the sensitivity of a genotyping
439 method may be reduced by not observing infections that occurred prior to trial enrollment, the
440 extent to which bias was reduced in our analysis was robust to changes in sensitivity of the
441 method and depended most upon its specificity. High specificity can likely be achieved given the

442 high expected heterozygosity of microsatellite and amplicon deep sequencing panels, as well as
443 the ability to detect minority clones in multiclonal *P. vivax* infections (35–39). Parasites sampled
444 during follow-up were genotyped in the DETECTIVE and IMPROV trials (6–9), and statistical
445 frameworks leveraging genotyping and time-to-event data to distinguish relapses from
446 reinfections have been successfully developed and could be readily integrated into the analysis of
447 past and future clinical trial data (12,39). Nevertheless, this approach may be insufficient to
448 overcome biases introduced by improper trial design, such as a short duration of follow-up
449 relative to the relapse rate in the trial location. Furthermore, higher performance characteristics
450 may be unachievable at higher transmission intensities, where the detectability of recurrent
451 infections is reduced and higher hypnozoite burdens and frequent reinfections limit the
452 information content of genotyping data.

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

461 There are a number of limitations of this analysis. First, we did not calibrate our model to
462 clinical trial data, so our results are not representative of any specific trial settings. However, our
463 simulations encompassed a wide range of epidemiological settings, so the site-specific biases
464 identified in our analysis should reflect variation possible across trial sites. Second, there remains

465 much about hypnozoite biology that is not well understood (40), so our simplified representation
466 of hypnozoite activation and death in our transmission model may not fully capture reality.
467 Finally, we have an imperfect understanding of the acquisition and loss of anti-parasite and
468 clinical immunity (41). Underestimating the level of anti-parasite and clinical immunity among
469 trial participants may lead to a greater fraction of recurrent infections detected during follow-up
470 than would occur in an actual trial context. This could have led us to overestimate site-specific
471 biases, in which case more refined mathematical representations of immunity could be beneficial
472 for future studies building on ours (10,42).

473 Our analysis predicted that site-specific biases are likely to occur in phase-III trials for
474 radical cure, and these results suggest that care should be taken in the planning of future trials
475 and the interpretation of trial data. Mathematical modeling that accounts for site-specific biases
476 can aid in the interpretation of clinical trial data and may be useful for identifying future trial
477 sites where biases may be less severe (43). As always, the utility of the insights from
478 mathematical models is improved with additional data, so modeling should be integrated into
479 clinical trial design to identify data needs and inform data collection in order to reduce these
480 biases and improve understanding of radical cure's potential to control *P. vivax* malaria.

481

482 **METHODS**

483 ***Ethics Statement***

484 No human subjects were used in this study, so IRB approval was not required.

485

486 ***Transmission Model***

487 To simulate *Plasmodium vivax* transmission, we used a stochastic, individual-based model
488 developed by White *et al*. (10) and Nekkab *et al*. (18). This model extends the Ross-MacDonald
489 framework of *P. falciparum* transmission to incorporate relapses (44), a characteristic feature of
490 *P. vivax* transmission (45). Calibrated to epidemiological surveys conducted in Papua New
491 Guinea, the Solomon Islands, and Brazil, the model by White *et al.* (10) and Nekkab *et al*. (18)
492 reproduces *P. vivax* transmission dynamics across a range of epidemiological settings.
493 The rates at which individuals in the population are reinfected and relapse depend upon
494 the respective number of infectious mosquito bites received (i.e., the entomological inoculation
495 rate, or EIR) and the number of hypnozoite batches present in each individual's liver. The
496 average EIR within the population is determined by the mosquito population dynamics, and
497 individual variation therein is due to heterogeneity in exposure and age-dependent differences in
498 biting. Following successful inoculation with *P. vivax* sporozoites from an infectious mosquito
499 bite, an individual accumulates an additional batch of hypnozoites within the liver and
500 experiences a primary bloodstream infection. In the absence of treatment, each batch of
501 hypnozoites can activate to cause relapses or be cleared naturally from the liver according to
502 fixed rates (Table S1).
503 The nature of each blood-stage infection (e.g., reinfection or relapse) is determined by the
504 level of anti-parasite and clinical immunity in the infected individual. In the model, the level of
505 anti-parasite immunity determines the probability that a blood-stage infection is detectable by
506 light microscopy (LM) and increases the rate at which low-density infections are cleared. The
507 level of clinical immunity determines the probability that an individual with a blood-stage
508 infection exhibits symptoms.

509 The levels of anti-parasite and clinical immunity interact to determine the detectability of
510 each blood-stage *P. vivax* infection. The model by White *et al.* (10) and Nekkab *et al.* (18)
511 considers three types of blood-stage infections: (1) sub-microscopic infections detectable by
512 PCR; (2) subclinical infections detectable by LM and PCR; and (3) clinical infections detectable
513 by LM and PCR. Following White *et al.* (10) and Nekkab *et al*. (18), we assumed that each
514 clinical infection was characterized by a fever exceeding 38°C within the last 48 hours and a
515 parasite density greater than 500/μL. Individuals with clinical *P. vivax* infections then seek
516 treatment with anti-malarial drugs according to a probability of treatment-seeking behavior.
517 Simulation of the *P. vivax* transmission model occurs in two steps. First, the population is
518 initialized at equilibrium according to the analogous set of deterministic compartmental
519 differential equations. Then, the stochastic, individual-based model is simulated with the
520 initialized population. For further description of the model and its assumptions, please refer to
521 the Supplement and the documentation provided in White *et al.* (10) and Nekkab *et al*. (18).
522
523 ***Trial Overview***
524 We constructed phase-III clinical trials for a generic 8-aminoquinoline (e.g., primaquine or
525 tafenoquine). We simulated trials in order to quantify the recurrence-free efficacy as a measure
526 of the therapeutic effect of radical cure against the hypnozoite batches present in each individual
527 presenting with a clinical *P. vivax* infection confirmed by light microscopy.
528 We designed our simulated phase-III clinical trials to be comparable to previous trials of
529 PQ and TFQ (6–8). Participants were randomly assigned to the treatment and control arms, and
530 all participants received chloroquine (CQ) upon enrollment to clear the *P. vivax* blood-stage
531 infection. Consistent with the DETECTIVE trial (6,8), participants were followed for 180 days

532 and assessed for asexual parasitemia by light microscopy at days 8, 15, 22, 29, 60, 90, 120, and
533 180 post-enrollment. For trials simulated with longer duration of follow-up, additional time
534 points were included. See supplement for further details.
535
536 ***Radical Cure Model***
537 We modeled the action of radical cure in each individual receiving treatment as the per-
538 hypnozoite batch probability of clearance. To allow for heterogeneous action of the intervention
539 (46), we assumed that a proportion p_i of the population belonged to each stratum *i* with per-
540 hypnozoite batch clearance probability, c_i. For a population consisting of two strata, the mean
541 per-hypnozoite batch probability of clearance was
542
543 $$\bar{c} = p_1 c_1 + p_2 c_2, \quad (1)$$
544
545 where $p_2 = 1 - p_1$. This model of heterogeneous action of radical cure generalized the
546 intervention actions commonly considered in the evaluation of clinical trials. Under the "leaky"
547 model (46–48), the action of radical cure is the same across all individuals in the population
548 ($p_1 = 1$) with $0 \leq c_1 \leq 1$. Under the "all-or-none" model (46–48), radical cure completely
549 clears hypnozoites in a subset of the population ($0 < p_1 < 1; c_1 = 1$) and has no effect on the
550 hypnozoites in the remainder of the population ($p_2 = 1 - p_1; c_2 = 0$). These two models of
551 intervention action could manifest due to host-specific factors, most notably the cytochrome P-
552 450 isoenzyme 2D6 (CYP2D6) genotype, an enzyme involved in metabolizing PQ. For
553 individuals with a low CYP2D6 metabolization phenotype, evidence suggests that PQ may not
554 prevent *P. vivax* relapses (11), though its precise effect on 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. S6). Therefore, we assumed by default that each 8-aminoquinoline had an "all-or-none" action with $p_1 = 0.75$ and $c_1 = 1$.

Under the model of heterogeneous action of radical cure, the number of batches of hypnozoites following treatment was distributed as

$$H_{post} \sim Binomial\left(H_{pre}, 1 - c_i\right), \forall i \in \{1,2\}. \quad (2)$$

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

577 We enrolled individuals in the simulated clinical trial if they presented at a health clinic with
578 febrile illness and were diagnosed by light microscopy with a *P. vivax* infection. Consistent with
579 the DETECTIVE trial (6,8), participants were enrolled if they were at least 16 years of age and
580 had a measured glucose-6-phosphate dehydrogenase (G6DP) activity greater than 70% of the
581 median value of the trial location. Following Nekkab *et al*. (18), we assumed that G6PD activity
582 was measured using the SD Biosensor STANDARD G6PD test (49). This test qualitatively
583 classifies G6PD activity as normal (>70% activity), intermediate (30-70% activity), and low
584 (<30% activity).

585 Individuals that met the criteria for enrollment were then randomly assigned to the
586 treatment or control arm of the clinical trial. Allocation of participants to each trial arm occurred
587 with equal probability, provided that the current number of participants in each arm was less than
588 the desired sample size of 1,000.

589

590 *Treatment*

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

600

601 *Vector Control*

602 As in the DETECTIVE trial (6,8), all participants were provided with a long-lasting insecticidal
603 net (LLIN) to prevent reinfection from mosquitoes during follow-up and therefore decrease the
604 potential bias that may arise in calculating efficacy based on a recurrent infection endpoint.
605 Following Griffin *et al*. (42) and White *et al*. (10), we assumed that vector control measures
606 decrease the probabilities of mosquito biting and successful feeding and increase the
607 probabilities of repellency and death. The magnitude of the effect depends upon the proportion
608 of mosquito bites that occur while individuals are indoors (Φ_I) and in bed (Φ_B), the duration of
609 usage of the LLIN, and the duration of insecticidal activity. We assumed that the probability of
610 usage decayed exponentially over time with a half-life of 3 years. We further assumed that
611 insecticidal activity decayed exponentially over time with a half-life of 2.5 years (18).

612 We also considered whether the use of indoor residual spraying (IRS) in each
613 participant's house administered in isolation or in combination with LLINs decreased the
614 potential bias in our efficacy estimates. As with LLINs, we assumed that the use of IRS
615 decreases the probabilities of mosquito biting and successful feeding and increases the
616 probabilities of repellency and death (10,42). To account for the waning effect of IRS over time,
617 we assumed that the insecticidal activity decayed exponentially with a half-life of 6 months (18).

618

619 *Follow-up*

620 We followed each participant for 180 days following enrollment in the clinical trial and
621 treatment with anti-malarial drugs. Consistent with the DETECTIVE trial (6,8), for each
622 participant, we recorded the date of each clinical *P. vivax* infection that occurred within the

623 duration of follow-up. Furthermore, we tested each participant for *P. vivax* asexual parasites
624 using light microscopy on days 8, 15, 22, 29, 60, 90, 120, 160, and 180 post-enrollment. We
625 assumed no mortality among trial participants during the follow-up period.
626 To examine potential biases that arise from trial surveillance methods, we kept a
627 complete record of all recurrent infections that occurred within the duration of follow-up,
628 including those that would not have been detected under the trial protocol. For each recurrent
629 infection, we recorded the cause (i.e., reinfection or relapse) and the type of blood-stage infection
630 (e.g., sub-microscopic, sub-clinical, or clinical). We further distinguished between relapses
631 caused by hypnozoite batches acquired prior to treatment and relapses caused by hypnozoites
632 batches acquired following treatment. Only relapses caused by hypnozoite batches acquired prior
633 to treatment reflect the action of 8-aminoquinoline treatment against hypnozoites batches.
634
635 ***Statistical Analyses***
636 We calculated the efficacy of the 8-aminoquinoline used in each clinical trial from the output
637 collected in each respective simulation. We calculated efficacy using multiple metrics in order to
638 examine how different data collected during trial follow-up resolved biases in the efficacy
639 estimates.
640 Consistent with previous clinical trials (6,8), we used freedom from LM-detectable
641 recurrent infection as the default efficacy metric in our simulation studies. We calculated
642 efficacy using the Cox proportional hazards model (50), which computes the hazard ratio
643 between the treatment and control arms based on the times to first LM-detectable recurrent
644 infection. Following the DETECTIVE trial protocol (6,8), recurrent infections that occurred

645 before 32 days post-enrollment were not included. Given that the timing of each recurrent
646 infection was unobserved, all trial participants were interval censored.
647 We calculated efficacy using incidence rates computed from the recurrent infections that
648 occurred within the duration of follow-up. We calculated efficacy based on incidence rates as
649

650 $$E_{IR} = 1 - \frac{e_t/y_t}{e_c/y_c}. \quad (3)$$

651

652 In eq. (3), e_t and e_c are the respective number of infection events in the treatment and control
653 arms, and y_t and y_c are the respective number of person-years of follow-up in the treatment and
654 control arms (48,50).
655 The magnitude of the measured efficacy also depends upon the action of the radical cure
656 therapeutic. In absence of other sources of bias, we will overestimate efficacy for 8-
657 aminoquinolines that elicit an "all-or-none" response if we calculate efficacy based on incidence
658 rates. This phenomenon is caused by the subset of individuals in the treatment arm who
659 completely clear the hypnozoites from their livers and increase the number of person-years of
660 follow-up (48). To resolve this bias that arises from the assumption of intervention action, we
661 calculated efficacy based on risk as
662

663 $$E_R = 1 - \frac{p_t}{p_c}, \quad (4)$$

664

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

667

668 ***Simulation Experiments***

669 We performed simulation experiments to quantify biases in the efficacy estimates that may arise
670 under different transmission settings and other features of the clinical trial design (Table 2). First,
671 we considered how our efficacy estimates varied with transmission intensity and heterogeneity in
672 mosquito biting patterns. Second, we considered how differences in the rate of relapse of the *P.*
673 *vivax* hypnozoites interact with the duration of the trial follow-up and the transmission intensity
674 to bias our estimates of efficacy. After characterizing these biases, we assessed whether the
675 allocation of vector control measures, such as LLINs and IRS, to trial participants could prevent
676 reinfections and thereby reduce the bias in the efficacy estimates. We next considered whether a
677 method using time-to-event and genotyping data that is capable of distinguishing between
678 different types of *P. vivax* recurrent infections at different sensitivities and specificities could
679 reduce or correct the bias in the efficacy estimates. Then, we measured the extent to which the
680 duration of prophylaxis provided by the treatment regimen biases our estimates of efficacy by
681 preventing reinfections in the treatment arm only. Finally, we considered how the efficacy
682 estimates that we obtained varied with the choice of efficacy metric and infection endpoint.

683 For each simulation setting, we simulated a phase-III clinical trial with 1,000 participants
684 in each arm. This sample size was chosen to ensure that the power of our clinical trials exceeded
685 95% for each transmission intensity and efficacy metric considered (Fig. S7). Each population
686 was simulated for 30 years prior to trial enrollment to ensure that transmission stabilized. Each
687 trial was simulated in a population of 200,000 individuals, and the maximum duration of each
688 trial was 15 years in order to ensure that complete enrollment was attained, even at low
689 transmission intensities. By default, PQ was the 8-aminoquinoline provided to the participants in

690 the treatment arm and when co-administered with CQ provided 28 days of blood-stage and liver-
691 stage prophylaxis during which trial participants could not become reinfected or relapse. The
692 mean per-hypnozoite batch probability of clearance was equal to 0.75. For each simulation
693 setting, we simulated trials in which the therapeutic had an "all-or-none" response. Because the
694 transmission model was stochastic, we ran 200 simulations for each simulation setting and
695 computed the mean and interquartile range for the efficacy estimates across these simulations.
696 By default, we defined efficacy as freedom from LM-detectable recurrent infection calculated
697 using the Cox proportional hazards model.
698

699 ***Table 2. Parameters Varied during the Simulation Analyses.*** *The parameters that were varied*
700 *during each simulation analysis are reported. All other parameters in the transmission model*
701 *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

702

703 ***Availability of Code***

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

REFERENCES

1. Battle KE, Lucas TCD, Nguyen M, Howes RE, Nandi AK, Twohig KA, et al. Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. The Lancet. 2019 Jul;394(10195):332–43.

2. Price RN, Commons RJ, Battle KE, Thriemer K, Mendis K. Plasmodium vivax in the Era of the Shrinking P. falciparum Map. Trends in Parasitology. 2020 Jun;36(6):560–70.

3. Adekunle AI, Pinkevych M, McGready R, Luxemburger C, White LJ, Nosten F, et al. Modeling the Dynamics of Plasmodium vivax Infection and Hypnozoite Reactivation In Vivo. Sinnis P, editor. PLoS Negl Trop Dis. 2015 Mar 17;9(3):e0003595.

4. Commons RJ, Simpson JA, Watson J, White NJ, Price RN. Estimating the Proportion of Plasmodium vivax Recurrences Caused by Relapse: A Systematic Review and Meta-Analysis. The American Journal of Tropical Medicine and Hygiene. 2020 Sep 2;103(3):1094–9.

5. World Health Organization, Global Malaria Programme, World Health Organization. Control and elimination of plasmodium vivax malaria: a technical brief [Internet]. 2015 [cited 2020 Nov 17]. Available from: http://apps.who.int/iris/bitstream/10665/181162/1/9789241509244_eng.pdf?ua=1&ua=1

6. Llanos-Cuentas A, Lacerda MVG, Hien TT, Vélez ID, Namaik-larp C, Chu CS, et al. Tafenoquine versus Primaquine to Prevent Relapse of *Plasmodium vivax* Malaria. N Engl J Med. 2019 Jan 17;380(3):229–41.

7. Llanos-Cuentas A, Lacerda MV, Rueangweerayut R, Krudsood S, Gupta SK, Kochar SK, et al. Tafenoquine plus chloroquine for the treatment and relapse prevention of Plasmodium vivax malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study. The Lancet. 2014 Mar;383(9922):1049–58.

8. Lacerda MVG, Llanos-Cuentas A, Krudsood S, Lon C, Saunders DL, Mohammed R, et al. Single-Dose Tafenoquine to Prevent Relapse of *Plasmodium vivax* Malaria. N Engl J Med. 2019 Jan 17;380(3):215–28.

9. Taylor WRJ, Thriemer K, von Seidlein L, Yuentrakul P, Assawariyathipat T, Assefa A, et al. Short-course primaquine for the radical cure of Plasmodium vivax malaria: a multicentre, randomised, placebo-controlled non-inferiority trial. The Lancet. 2019 Sep;394(10202):929–38.

10. White MT, Walker P, Karl S, Hetzel MW, Freeman T, Waltmann A, et al. Mathematical modelling of the impact of expanding levels of malaria control interventions on Plasmodium vivax. Nat Commun. 2018 Dec;9(1):3300.

11. Bennett JW, Pybus BS, Yadava A, Tosh D, Sousa JC, McCarthy WF, et al. Primaquine Failure and Cytochrome P-450 2D6 in *Plasmodium vivax* Malaria. N Engl J Med. 2013 Oct 3;369(14):1381–2.

12. Taylor AR, Watson JA, Chu CS, Puaprasert K, Duanguppama J, Day NPJ, et al. Resolving the cause of recurrent Plasmodium vivax malaria probabilistically. Nat Commun. 2019 Dec;10(1):5595.

13. Stepniewska K, White NJ. Some considerations in the design and interpretation of antimalarial drug trials in uncomplicated falciparum malaria. Malar J. 2006 Dec;5(1):127.

14. Vynnycky E, Sumner T, Fielding KL, Lewis JJ, Cox AP, Hayes RJ, et al. Tuberculosis Control in South African Gold Mines: Mathematical Modeling of a Trial of Community-Wide Isoniazid Preventive Therapy. Am J Epidemiol. 2015 Apr 15;181(8):619–32.

15. WorldWide Antimalarial Resistance Network Methodology Study Group. Competing risk events in antimalarial drug trials in uncomplicated Plasmodium falciparum malaria: a WorldWide Antimalarial Resistance Network individual participant data meta-analysis. Malar J. 2019 Jul 5;18(1):225.

16. Cooper L, Kang SY, Bisanzio D, Maxwell K, Rodriguez-Barraquer I, Greenhouse B, et al. Pareto rules for malaria super-spreaders and super-spreading. Nat Commun. 2019 Dec;10(1):3939.

17. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical variation in Plasmodium vivax relapse. Malar J. 2014 Dec;13(1):144.

18. Nekkab N, Lana R, Lacerda M, Obadia T, Siqueira AM, Monteiro WM, et al. Estimated impact of tafenoquine for Plasmodium vivax control and elimination in Brazil: A modelling study. PLoS Medicine. 2021 Jan 6;

19. Maire N, Aponte JJ, Ross A, Thompson R, Alonso P, Utzinger J, et al. Modeling a field trial of the RTS,S/AS02A malaria vaccine. Am J Trop Med Hyg. 2006 Aug;75(2 Suppl):104–10.

20. Penny MA, Galactionova K, Tarantino M, Tanner M, Smith TA. The public health impact of malaria vaccine RTS,S in malaria endemic Africa: country-specific predictions using 18 month follow-up Phase III data and simulation models. BMC Med. 2015 Jul 29;13:170.

21. España G, Hogea C, Guignard A, ten Bosch QA, Morrison AC, Smith DL, et al. Biased efficacy estimates in phase-III dengue vaccine trials due to heterogeneous exposure and differential detectability of primary infections across trial arms. Azman AS, editor. PLoS ONE. 2019 Jan 25;14(1):e0210041.

22. White RG, Orroth KK, Korenromp EL, Bakker R, Wambura M, Sewankambo NK, et al. Can population differences explain the contrasting results of the Mwanza, Rakai, and Masaka HIV/sexually transmitted disease intervention trials?: A modeling study. J Acquir Immune Defic Syndr. 2004 Dec 1;37(4):1500–13.

23. Yang A, Cai F, Lipsitch M. Herd immunity alters the conditions for performing dose schedule comparisons: an individual-based model of pneumococcal carriage. BMC Infect Dis. 2019 Mar 5;19(1):227.

780 24. Werkman M, Toor J, Vegvari C, Wright JE, Truscott JE, Ásbjörnsdóttir KH, et al. Defining
781 stopping criteria for ending randomized clinical trials that investigate the interruption of
782 transmission of soil-transmitted helminths employing mass drug administration. PLoS Negl
783 Trop Dis. 2018 Oct;12(10):e0006864.

784 25. Dahal P, Simpson JA, Dorsey G, Guérin PJ, Price RN, Stepniewska K. Statistical methods
785 to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum
786 malaria: pitfalls and challenges. Malar J. 2017 Dec;16(1):430.

787 26. Commons RJ, Simpson JA, Thriemer K, Humphreys GS, Abreha T, Alemu SG, et al. The
788 effect of chloroquine dose and primaquine on Plasmodium vivax recurrence: a WorldWide
789 Antimalarial Resistance Network systematic review and individual patient pooled meta-
790 analysis. The Lancet Infectious Diseases. 2018 Sep;18(9):1025–34.

791 27. Pollard EJM, MacLaren D, Russell TL, Burkot TR. Protecting the peri-domestic
792 environment: the challenge for eliminating residual malaria. Sci Rep. 2020 Dec;10(1):7018.

793 28. Monroe A, Moore S, Okumu F, Kiware S, Lobo NF, Koenker H, et al. Methods and
794 indicators for measuring patterns of human exposure to malaria vectors. Malar J. 2020
795 Dec;19(1):207.

796 29. Moreno JE, Rubio-Palis Y, Páez E, Pérez E, Sánchez V. Abundance, biting behaviour and
797 parous rate of anopheline mosquito species in relation to malaria incidence in gold-mining
798 areas of southern Venezuela. Medical and Veterinary Entomology. 2007 Dec 18;21(4):339–
799 49.

800 30. Paul MR, Grenfell BT, Hoti SL, Ramaiah KD, Bundy DA, Das PK, et al. Quantifying
801 mosquito biting patterns on humans by DNA fingerprinting of bloodmeals. The American
802 Journal of Tropical Medicine and Hygiene. 2001 Dec 1;65(6):722–8.

803 31. Guelbéogo WM, Gonçalves BP, Grignard L, Bradley J, Serme SS, Hellewell J, et al.
804 Variation in natural exposure to anopheles mosquitoes and its effects on malaria
805 transmission. eLife. 2018 Jan 23;7:e32625.

806 32. Liebman KA, Stoddard ST, Reiner RC, Perkins TA, Astete H, Sihuincha M, et al.
807 Determinants of Heterogeneous Blood Feeding Patterns by Aedes aegypti in Iquitos, Peru.
808 Barrera R, editor. PLoS Negl Trop Dis. 2014 Feb 13;8(2):e2702.

809 33. Sagna AB, Kibria MG, Naher S, Islam S, Aktaruzzaman MM, Alam MS, et al. Stratifying
810 malaria receptivity in Bangladesh using archived rapid diagnostic tests. Malar J. 2020 Sep
811 23;19(1):345.

812 34. Poinsignon A, Cornelie S, Ba F, Boulanger D, Sow C, Rossignol M, et al. Human IgG
813 response to a salivary peptide, gSG6-P1, as a new immuno-epidemiological tool for
814 evaluating low-level exposure to Anopheles bites. Malar J. 2009 Aug 13;8:198.

35. Imwong M, Boel ME, Pagornrat W, Pimanpanarak M, McGready R, Day NPJ, et al. The First Plasmodium vivax Relapses of Life Are Usually Genetically Homologous. The Journal of Infectious Diseases. 2012 Feb 15;205(4):680–3.

36. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, et al. Relapses of *Plasmodium vivax* Infection Usually Result from Activation of Heterologous Hypnozoites. J INFECT DIS. 2007 Apr;195(7):927–33.

37. Koepfli C, Mueller I, Marfurt J, Goroti M, Sie A, Oa O, et al. Evaluation of *Plasmodium vivax* Genotyping Markers for Molecular Monitoring in Clinical Trials. J INFECT DIS. 2009 Apr;199(7):1074–80.

38. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 2017 Dec;18(1):864.

39. Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabora O, et al. Using Amplicon Deep Sequencing to Detect Genetic Signatures of *Plasmodium vivax* Relapse. J Infect Dis. 2015 Sep 15;212(6):999–1008.

40. Shanks GD, White NJ. The activation of vivax malaria hypnozoites by infectious diseases. The Lancet Infectious Diseases. 2013 Oct;13(10):900–6.

41. Mandal S, Sarkar R, Sinha S. Mathematical models of malaria - a review. Malar J. 2011;10(1):202.

42. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, et al. Reducing Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. Krishna S, editor. PLoS Med. 2010 Aug 10;7(8):e1000324.

43. 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.

44. 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.

45. 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.

46. 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.

849 47. Smith PG, Rodrigues LC, Fine PEM. Assessment of the Protective Efficacy of Vaccines
850 against Common Diseases Using Case-Control and Cohort Studies. Int J Epidemiol.
851 1984;13(1):87–93.

852 48. White MT, Griffin JT, Drakeley CJ, Ghani AC. Heterogeneity in malaria exposure and
853 vaccine response: implications for the interpretation of vaccine efficacy trials. Malar J.
854 2010 Dec;9(1):82.

855 49. Pal S, Bansil P, Bancone G, Hrutkay S, Kahn M, Gornsawun G, et al. Evaluation of a Novel
856 Quantitative Test for Glucose-6-Phosphate Dehydrogenase Deficiency: Bringing
857 Quantitative Testing for Glucose-6-Phosphate Dehydrogenase Deficiency Closer to the
858 Patient. Am J Trop Med Hyg. 2019;100(1):213–21.

859 50. Halloran ME, Longini IM, Struchiner CJ. Design and Analysis of Vaccine Studies
860 [Internet]. New York, NY: Springer New York; 2010 [cited 2020 Oct 25]. (Statistics for
861 Biology and Health). Available from: http://link.springer.com/10.1007/978-0-387-68636-3

862