

1 **A systematic review of sample size estimation accuracy on power in**
2 **malaria cluster randomised trials measuring epidemiological**
3 **outcomes**

4

5 **Authors & Affiliations**

6

7 Joseph Biggs^{1#}, Joseph D. Challenger², Joel Hellewell², Thomas S. Churcher² & Jackie Cook¹

8

1. Medical Research Council (MRC) International Statistics and Epidemiology group,
9 Department of Infectious Disease Epidemiology and International Health, London School
10 of Hygiene and Tropical Medicine, London, UK

11

2. Medical research Council (MRC) Centre for Global Infectious Disease Analysis,
12 Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College
13 London, London, UK

14

Corresponding author: joseph.biggs1@lshtm.ac.uk

15

Email contacts

16

- joseph.biggs1@lshtm.ac.uk

17

- j.challenger@imperial.ac.uk

18

- joel@ebi.ac.uk

19

- thomas.churcher@imperial.ac.uk

20

- jackie.cook@lshtm.ac.uk

21

22

23

24

25

26

27

28

29 **Abstract**

30

31 **Introduction**

32 Cluster randomised trials (CRTs) are the gold standard for measuring the community-wide impacts of
33 malaria control tools. CRTs rely on well-defined sample size estimations to detect statistically
34 significant effects of trialled interventions; however these are often predicted poorly by investigators.
35 Here, we review the accuracy of predicted parameters used in sample size calculations for malaria
36 CRTs with epidemiological outcomes.

37 **Methods**

38 We searched for published malaria CRTs using four online databases in March 2022. Eligible trials
39 included those with malaria-specific epidemiological outcomes which randomised at least six
40 geographical clusters to study arms. Predicted and observed sample size parameters were extracted
41 by reviewers for each trial. Pair-wise correlation coefficients were calculated to assess the accuracy
42 of predicted control-arm outcome estimates and desired relative reductions (effect sizes) between
43 arms compared to what was observed. Among trials which retrospectively calculated an estimate of
44 heterogeneity in cluster outcomes, we recalculated study power according to observed trial
45 estimates.

46 **Results**

47 Of the 1889 records identified and screened, 108 articles were eligible and comprised of 71 malaria
48 CRTs. Among trials that included sample size calculations (91.5%, 65/71), most estimated cluster
49 heterogeneity using the coefficient of variation (k) (80%, 52/65) which were often predicted without
50 using prior data (67.7%, 44/65). Predicted control-arm epidemiological outcomes correlated weakly
51 with those observed, with 61.2% (19/31) of prevalence estimates overestimated. Among the
52 minority of trials which retrospectively calculated cluster heterogeneity (20%, 13/65), empirical
53 values contrasted with those used in sample size estimations and often compromised study power.
54 Observed effect sizes were often smaller than had been predicted at the sample size stage (72.9%,
55 51/70) and were typically higher in the first, compared to the second, year of trials. Overall, effect
56 sizes achieved by malaria interventions tested in trials decreased between 1995 and 2021.

57 **Conclusions**

58 Study findings reveal sample size parameters in malaria CRTs were often inaccurate and resulted in
59 underpowered studies. Future trials must strive to obtain more representative epidemiological
60 sample size inputs to ensure interventions against malaria are adequately evaluated.

61 **Registration**

62 This review is registered with PROSPERO (CRD42022315741).

63

64 **Keywords**

65

66 Malaria; Cluster randomised trial; sample size; heterogeneity; coefficient of variation

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91 Introduction

92

93 Malaria is a parasitic disease that in 2022 was responsible for the deaths of 608,000 individuals
94 worldwide; most of whom were children in Sub-Saharan Africa [1]. There are numerous, effective
95 interventions that can be used to combat malaria transmission that are recommended by the World
96 Health Organisation (WHO). To generate evidence for the recommendation of these tools, cluster
97 randomised trials (CRTs) are conducted to demonstrate the community-wide impacts [2].
98 Historically, CRTs have been instrumental in demonstrating the mass effects of the first insecticide-
99 treated bed nets (ITNs) [3-5], mass chemoprevention strategies [4, 6], and most recently, dual-
100 action, long-lasting insecticide treated nets (LLINs) [7, 8]. Despite their necessity, CRTs are subject to
101 major constraints. Trialling interventions over large geographical areas is costly, logistically
102 challenging, and at the design stage, requires well-defined estimates of underlying transmission
103 patterns in the study setting [9]. Consequently, in recent years, some malaria CRTs have reported
104 being underpowered and have presented inconclusive findings [10-14].

105 Investigators power CRTs according to sample size estimations which account for cluster-level
106 randomisation, where groups of people, as opposed to individuals, are randomised to receive
107 interventions. This design can result in heterogeneity of outcomes between and within clusters
108 owing to groups of individuals, such as households, schools, and geographical areas, sharing similar
109 biological and socio-economic characteristics which introduces correlation in study outcomes [15,
110 16]. Consequently, cluster heterogeneity needs to be incorporated into sample size estimations,
111 along with expected control arm transmission and effect size estimates (relative percentage
112 reductions between arms), to compensate for the lower precision associated with this design. The
113 between- and within-cluster heterogeneity can be measured using the coefficient of variation (k) or
114 intra-cluster correlation coefficient (ICC), respectively, and heavily impacts trial size [15, 16]. Trialling
115 new interventions in areas with missing or inadequate data results in investigators having to rely on
116 judgement-based estimates for their sample size estimations which may be inaccurate.

117 Numerous reviews have evaluated sample size estimations in CRTs focused on cancer treatments
118 [17], school-based interventions [18], oral health [19], residential care [20] and CRTs in general [21].
119 These reviews highlighted that despite trials mostly including sample size estimations, not all
120 calculations accounted for cluster heterogeneity (73% [17], 78% [18], 71% [19] and 47% [20]). Two of
121 these reviews further explored whether trials also included empirical measures of cluster
122 heterogeneity and compared them to prior estimates [18, 21]. Both reviews highlighted trials rarely
123 provided retrospective estimates of cluster heterogeneity (<40%), and among trials that did, large

124 differences were identified between predicted and observed estimates. This suggests many trialists
125 misclassified the true degree of cluster heterogeneity at the design stage. Finally, one review
126 explored which trials stated their desired effect sizes and compared them to those observed [21].
127 They showed that 68% of desired effect sizes were overestimated. Interestingly, none of these
128 reviews compared the outcome measures predicted and observed in the control arms of their
129 included trials. This is crucial as misclassification of predicted effect size, cluster heterogeneity and
130 control-arm outcome measurements all impact study power [15, 16, 22].

131 Malaria transmission is driven by numerous environmental and socio-economic factors including
132 rainfall, temperature, vegetation cover, type of housing and provision of malaria interventions [23-
133 26]. Consequently, transmission is often spatially and temporally variable across various
134 geographical scales. This presents a challenge for malaria CRTs as heterogenous transmission in the
135 community may result in spatial/temporal variability in malaria-specific outcomes between
136 geographical clusters. Therefore, estimating the level of malaria transmission in the control arm and
137 the degree of cluster heterogeneity for malaria CRT sample size estimations is difficult in the
138 absence of baseline data.

139 In this review, our aim was to investigate the characteristics and quality of sample size estimations in
140 malaria CRTs that used geographical clusters. Specifically, we explored whether study investigators
141 accurately predicted sample size estimation parameters, including control-arm transmission, cluster
142 heterogeneity, and desired effect sizes, according to observed measurements during trials. It is
143 hoped results from this review will improve future study design and help ensure trialists are able to
144 accurately detect impacts of interventions that are vital in the fight against malaria.

145 **Methods**

146

147 **Search strategy and selection criteria**

148

149 We conducted a systematic review of published malaria CRTs with epidemiological outcomes. In
150 March 2022, we searched the database systems Pubmed, Web of Science, Embase and Cochrane
151 reviews using truncated versions of the terms 'malaria' and 'cluster randomised trial' for trials
152 published in English language. The bibliographies of identified reviews were additionally screened
153 according to title and abstract. Search results were imported into the reference manager Endnote
154 where digitally identified duplicates were removed. Manually identified duplicates were removed by
155 two reviewers (JB & JH). Pre-determined eligibility criteria were used to screen identified articles
156 based on title and abstract (JB & JH) while screening discordance was adjudicated by consensus (TC

157 & JC). Identified studies were eligible for inclusion if they met the following criteria: the study was a
158 CRT wherein at least six geographical clusters were randomised to intervention and control arms;
159 the study measured malaria-specific epidemiological outcomes. Such outcomes include malaria
160 prevalence or incidence according to microscopy, rapid diagnostic tests (RDTs), or molecular
161 methods. Trials that only measured anaemia and all-cause mortality were excluded as these
162 outcomes could be attributed to other conditions. Prior to study initiation, the review was registered
163 in PROSPERO on 9th March 2022 (CRD42022315741).

164 **Data extraction**

165

166 Two reviewers (JB & JH) independently extracted information from the final list of studies. Extraction
167 discrepancies were resolved by consensus with TC and JC. Data for sample size estimations and
168 empirical outcomes were extracted for all epidemiological outcomes measured at all trial timepoints.
169 For each trial, we extracted data on overall trial design, randomisation method and type of
170 intervention evaluated. For each sample size estimation, we extracted data on all assumptions
171 outlined as well as those used to estimate cluster heterogeneity. To compare sample size
172 assumptions to observed trial outcomes, where data was available, we extracted arm-aggregated
173 malaria prevalence (cases/survey population) and/or incidence data (cases/person-years) by trial
174 year and overall.

175

176 **Data analysis**

177

178 For each sample size estimation where observed prevalence/incidence data were available, we
179 calculated the relative reduction (effect size) between intervention arms for the duration of each
180 study and stratified by year. The effect size was calculated according to equations A and B where
181 subscript (1) and (2) represent the control and intervention arms, respectively, while (r) and (p)
182 correspond to malaria incidence per person year and malaria prevalence, respectively. In this
183 manner, the effect size represents the % relative reduction between the control and the intervention
184 arm.

$$185 \qquad \qquad \qquad \textit{Prevalence effect size} = 1 - p_2 / p_1 \qquad \qquad \qquad \textit{eq. A}$$

$$186 \qquad \qquad \qquad \textit{Incidence effect size} = 1 - r_2 / r_1 \qquad \qquad \qquad \textit{eq. B}$$

187 For sample size parameters including control-arm prevalence, incidence and effect size estimates,
188 pair-wise Pearson's correlation coefficients were used to calculate the strength of association

189 between those estimated and observed for trials (Rho , p -value <0.05). To determine whether
190 parameters were accurately estimated by investigators, we calculated the relative percentage
191 difference between estimates and observed outputs. Parameters were classified as
192 overestimated/underestimated if they exceeded a relative percentage difference of 10%. Regarding
193 cluster heterogeneity estimates (k/ICC) provided in trials, we first investigated whether estimates
194 based on pilot or baseline data differed to estimates based on no data. A pair-wise t-test was used to
195 determine whether the mean value difference equalled zero. Among trials that reported cluster
196 heterogeneity using observed trial data, we recalculated study power (%) according to the observed
197 k/ICC and year 1 control arm prevalence/incidence but using the desired effect size quoted in the
198 paper. The remaining sample size parameters used were identical to the original power calculations:
199 desired effect size (%), cluster size, cluster number and significance level (%). Study power for CRTs
200 was calculated according to methods described by Hayes and Moulton in [15]. All analyses were
201 conducted in STATA (v.17, Texas, USA).

202 **Results**

203

204 Our literature search yielded 1889 records from database searching and 145 records from the
205 bibliographies of Cochrane reviews (Figure 1). Following the removal of duplicates, a total of 1302
206 records were screened after which 991 were excluded as they were not concerned with malaria
207 CRTs. The remaining 311 records were assessed for eligibility resulting in 108 published articles being
208 included in this study, which comprised of 71 malaria CRTs (Additional file 1). The review PRISMA
209 2020 checklist is included in additional file 2.

210

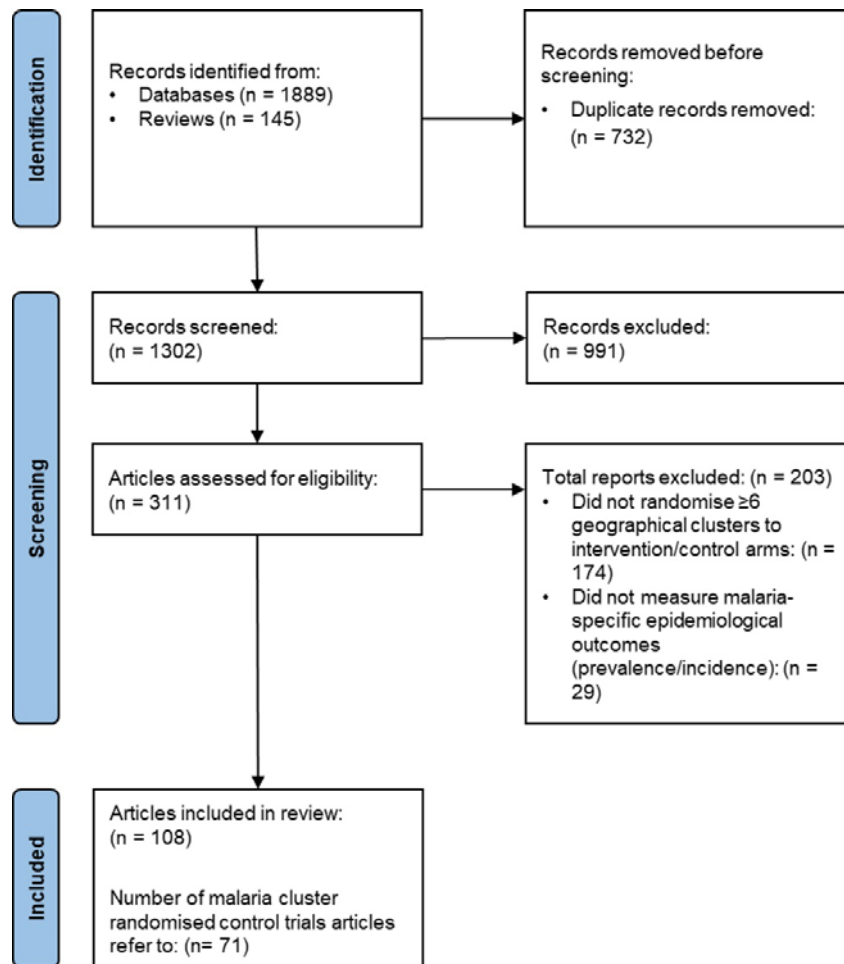


Figure 1: Study selection of included epidemiological malaria CRTs

211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226

The trial-level characteristics of the 71 malaria CRTs are shown in Table 1 and Figure 2A-G. Since 1995, the number of malaria-specific CRTs have increased in frequency and distribution across the world (Figure 2A&B). The 71 identified CRTs were conducted in a total of 78 countries, with study sites in Africa (N:53), Asia (N:21) and South America (N:4). 54.9% (39/71) of the trials evaluated vector control interventions, all measured *Plasmodium falciparum* outcomes while 26.8% (19/71) also measured *Plasmodium vivax* outcomes. Most trials adopted a parallel design (85.9%, 61/71) and consisted of two study arms (77.5%, 55/71). Concerning the cluster randomisation procedures, 39.4% (28/71) used simple randomisation to allocate clusters, 23.9% (17/71) implemented stratified randomisation, 22.5% (16/71) employed restricted randomisation, while 12.7% (9/71) randomised clusters within matched pairs. Among trials that randomised clusters through pair-matching or stratification, most restricted allocation based on a single criterion. For those that utilised restricted randomisation, most used between 3-4 restriction criteria (Figure 2F). The most common restriction

227 criteria used for randomisation included cluster transmission intensity (prevalence or incidence),
 228 cluster size, location, and historical intervention coverage (Figure 2G). Regarding cluster design,
 229 74.6% (53/71) adopted a basic cluster design, 14.1% (10/71) used a 'fried egg' design and 11.3%
 230 (8/71) reported ensuring a minimum buffer distance between clusters. Among trials that reported
 231 their minimum cluster buffer size, 72.7% (8/11) reported a minimum buffer size <2km while those
 232 who stated a minimum cluster separation reported a range between no separation and 3 km.

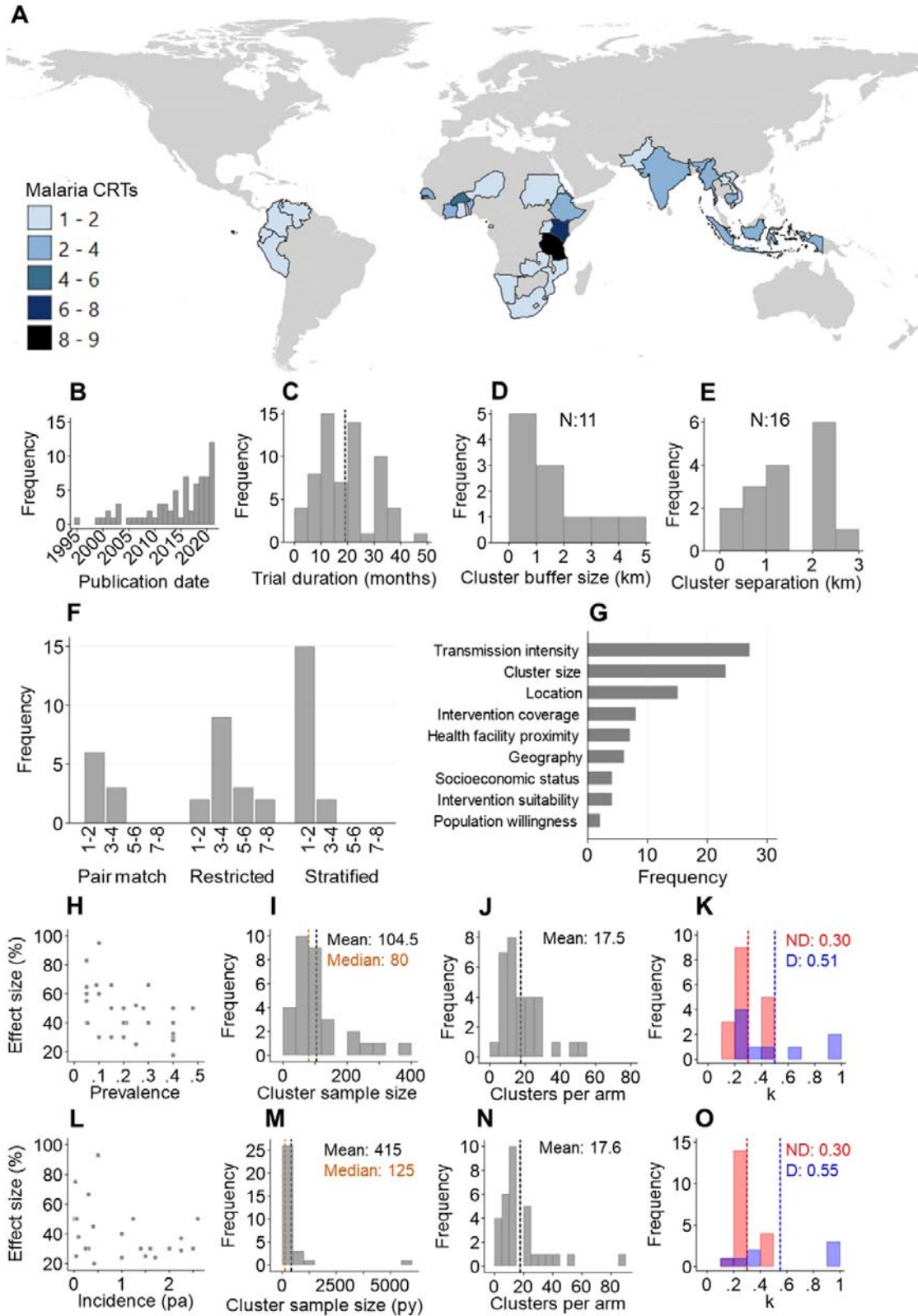
233 **Table 1: Overall characteristics of malaria CRTs identified in the systematic review (N:71)**

Trial characteristics	n	%
Intervention type		
Community engagement	1	1.4
Drug	26	36.6
Drug & Vector	5	7.0
Vector	39	54.9
Primary outcome		
Incidence	39	54.9
Prevalence	32	45.1
Malaria species		
<i>P. falciparum</i>	52	73.2
<i>P. falciparum</i> & <i>P. vivax</i>	19	26.8
Design		
Parallel	61	85.9
Factorial	4	5.6
Crossover	2	2.8
Stepped wedge	4	5.6
Randomisation method		
Block	1	1.4
Pair matched	9	12.7
Restricted	16	22.5
Simple	28	39.4
Stratified	17	23.9
Number of study arms		
2	55	77.5
3	9	12.7
4	7	9.9
Cluster type		
Basic	53	74.6
Buffered	8	11.3
Fried egg	10	14.1
Level of analysis		
Cluster	23	32.4
Individual	48	67.6
Formal sample size equation^a		
Yes	53	74.6

No	18	25.4
Reported being underpowered^b		
Yes	17	23.9
No	54	76.1

a: accounted for cluster heterogeneity

b: reported at the end of the trial



235

236 **Figure 2:** Characteristics of malaria CRTs identified in this review. **A:** Distribution of malaria CRTs. **B:**
 237 Annual frequency of malaria CRTs. **C:** Overall duration of malaria CRTs (dash line: mean). **D:** Size of
 238 buffers around study clusters. **E:** Minimum separation between study clusters. **F:** Number of

239 restriction criteria used according to the type of trial randomisation. **G:** The most utilised restriction
 240 criteria in malaria CRTs. Population willingness refers to population acceptance of trialled
 241 interventions. Sample size assumptions used in trials with prevalence as the outcome measure: **H:**
 242 Predicted control-arm prevalence compared to desired effect size. **I:** The desired total number of
 243 individuals surveyed per cluster. **J:** Required number of clusters per arm. **K:** the predicted coefficient of
 244 variation (*k*) for prevalence stratified by whether values were estimating using prior or baseline data
 245 (*D*) or assumed using no data (*ND*). Vertical dash: mean. Sample size assumptions used in trials with
 246 incidence as the outcome measure: **L:** Predicted control-arm incidence per person compared to
 247 desired effect size (*pa*: per annum). **M:** The desired person-years per cluster. **N:** Required number of
 248 clusters per arm. **O:** the predicted coefficient of variation (*k*) for incidence stratified by whether values
 249 were based on prior or baseline data (*D*) or assumed using no data (*ND*). Vertical dash: mean.

250

251 Among the included 71 CRTs, a total of 65 formal cluster sample size estimations were conducted
 252 that accounted for cluster heterogeneity by including a *k*, ICC or design effect component. Of these,
 253 34/65 were based on incidence while 31/65 were based on prevalence (Table 2 & Figure 2H-O). The
 254 remaining trials either did not account for cluster heterogeneity or lacked any sample size
 255 justification. Over 90% of all sample size estimations were calculated to achieve at 80% power
 256 (59/65) at the 5% significance level (60/65). Concerning the epidemiological outcome measures in
 257 the control arm, most investigators predicted incidence using prior data (70.6%, 24/34) while most
 258 investigators predicted prevalence without using prior data (54.8%, 17/31). Regarding sample size
 259 estimations based on prevalence (N: 31), investigators estimated a range of prevalences in the
 260 control arm (mean: 0.21, range: 0.05 – 0.48) and desired effect sizes (mean: 47.1% range: 17.5 - 95%)
 261 which tended to be higher in low prevalence settings (Figure 2H). The average required cluster
 262 sample size was 104.5 individuals (Median 80; Figure 2I) and average required number of clusters per
 263 arm equalled 17.5 (Figure 2J). For sample size estimations based on incidence (N: 34), a range of
 264 incidence estimates in the control arm were estimated (range: 0.002 - 2.6 cases per person per
 265 annum (*pa*)). Desired effect sizes (%) were similarly higher in lower incidence settings (mean: 41.1%,
 266 range 20 - 93%) (Figure 2L). The average cluster size for incidence was 415 (median: 125) person
 267 years (Figure 2M) while the mean number of required clusters per arm was 17.6 (Figure 2N).

268 **Table 2:** Characteristics of sample size estimations used in malaria CRTs stratified by outcome
 269 measure: prevalence or incidence. There were a total of 65 sample size estimations, 34 based on
 270 incidence outcomes while 31 based on prevalence outcomes.

Sample size characteristics	Overall N:65		Incidence N:34		Prevalence N:31	
	n	%	n	%	n	%
Desired power (%)						
80-90	59	90.8	29	85.3	30	96.8

90-100	6	9.2	5	14.7	1	3.2
Significance level (%)						
10	1	1.5	0	0.0	1	3.2
5	60	92.3	32	94.1	28	90.3
2.5	2	3.1	1	2.9	1	3.2
1.67	2	3.1	1	2.9	1	3.2
Cluster heterogeneity measure						
ICC	10	15.4	6	17.6	4	12.9
Design effect	3	4.6	2	5.9	1	3.2
k (CV)	52	80.0	26	76.5	26	83.9
Method used to estimate cluster heterogeneity						
With data ^a	21	32.3	9	26.5	12	38.7
Without data	44	67.7	25	73.5	19	61.3
Method used to estimate control-arm transmission						
With data ^a	38	58.5	24	70.6	14	45.2
Without data	27	41.5	10	29.4	17	54.8
Retrospectively estimated cluster heterogeneity						
Yes	13	20.0	5	14.7	8	25.8
No	52	80.0	29	85.3	23	74.2
Diagnostic used						
PCR	10	15.4	3	8.8	7	22.6
RDT	29	44.6	17	50.0	12	38.7
Microscopy	19	29.2	10	29.4	9	29.0
Mixed	7	10.8	4	11.8	3	9.7
Age range tested (years)						
<5	15	23.1	8	23.5	7	22.6
<10	14	21.5	9	26.5	8	25.8
<15	7	10.8	10	29.4	9	29.0
All ages	29	44.6	11	32.4	10	32.3

a: estimated using baseline or prior data

271

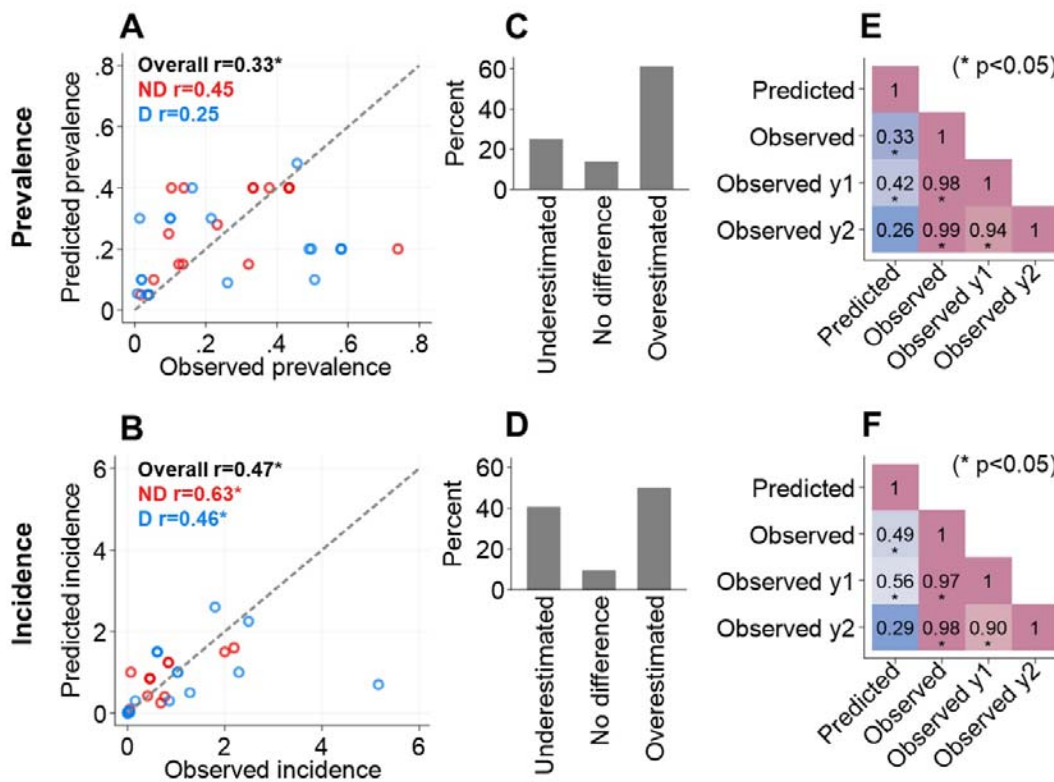
272 The most common cluster heterogeneity measure used in malaria CRT sample size calculations was
 273 the coefficient of variation (k) (80% 52/65). 67.7% (44/65) of estimated cluster heterogeneity
 274 measures were estimated with no prior data while only 32.3% (21/65) were estimated using baseline
 275 or pilot study data. Lastly, only a minority of investigators retrospectively calculated cluster
 276 heterogeneity using trial data (20%, 13/65) (Table 2).

277 **Control arm transmission intensity assumptions**

278

279 We explored how accurately epidemiological outcomes were predicted in the control arms of
 280 included trials (prevalence N: 31, incidence N: 34). Throughout each of the trials, predicted

281 prevalence and incidence only moderately positively correlated with corresponding observed values
 282 (incidence $r: 0.49$, $p\text{-value}<0.05$, prevalence $r: 0.39$, $p\text{-value}<0.05$) (Figure 3A&B). Moreover, most
 283 predicted prevalence and incidence estimates were overestimated by more than 10% according to
 284 observed estimates (prevalence overestimation: 61.1% (19/31); incidence overestimation: 50%
 285 (17/34)) (Figure 3C&D). We also assessed whether using prior data was better for predicting
 286 subsequent prevalence/incidence. Results revealed no different between using prior data or not
 287 (Figure 3A&B). Lastly, we found predicted prevalence/incidence correlated significantly with those
 288 observed in first year of each trial yet not the second year (Figure 3E&F). These results demonstrate
 289 investigators tended to poorly predict control-arm prevalence/incidence, particularly in the second
 290 year and that prevalence was often overestimated in malaria CRTs.
 291



292

293 **Figure 3:** Accuracy of predicted versus observed prevalence and incidence outcomes in malaria CRT
 294 control arms. **A & B:** Correlation between the predicted and overall observed prevalence/incidence
 295 stratified by method used to predict estimates: using data (D; blue) using no data (ND; red) and
 296 combined (black). r : correlation coefficient. **C & D:** The percentage of predicted prevalence/incidence
 297 estimates that were underestimated (relative percentage difference $<-10\%$), no difference (relative
 298 percentage difference -10% to 10%) or overestimated (relative percentage difference $>10\%$)
 299 according to overall observed estimates. **E & F:** Correlation matrix comparing the predicted
 300 prevalence/incidence with estimates observed throughout the trial (observed), in year 1 (Observed
 301 y1) and in year 2 (Observed y2). *: r $p\text{-value}<0.05$.

302 Cluster heterogeneity assumptions

303

304 Among trials that utilised the coefficient of variation (k) to account for cluster heterogeneity of
305 incidence/prevalence in their sample size estimations (Table 2), a range of values were used (mean:
306 0.37, range: 0.1 - 1.0) (Figure 2K&O). Values of k predicted using prior data were, on average,
307 statistically higher than those predicted with no prior data (no prior data mean k: 0.30; prior data
308 mean k: 0.52, t-test p-value<0.05). This suggests k was likely underestimated in many trials. A small
309 number of trials used the ICC to account for cluster heterogeneity, and they similarly had a large
310 range (mean: 0.12, range: 0.006 – 0.40) (Additional file 3).

311 Among the trials that additionally calculated k/ICC at the end of the study using empirical data (20%
312 13/65), we explored whether predicted cluster heterogeneity estimates were accurate and then used
313 the measured value for k/ICC to recalculate study power (Table 3). Empirical cluster heterogeneity
314 estimates often differed to those used in sample size estimations with the majority underestimating
315 k/ICC (61.5% 8/13). Recalculated power estimates, according to the stated desired effect, observed
316 control arm transmission intensity and k/ICC, were often lower than originally planned for. The
317 observed power for 7/11 trials was below 80%. For 4/11 trials, cluster heterogeneity was
318 overestimated which resulted in them remaining suitably powered to detect their desired effect
319 sizes. We were unable to replicate power calculations for two of the trials that provided retrospective
320 k/ICC estimates [27, 28]. It should be noted it was not always stated which timepoint/subset of trial
321 data was used to retrospectively calculate k/ICC.

322 **Table 3:** *The study power (%) to detect desired effect sizes according to predicted (left) and observed*
323 *(right) sample size parameters among trials the retrospectively calculated cluster heterogeneity. The*
324 *predicted sample size parameters include the predicted control-arm prevalence/incidence and the*
325 *k/ICC values stated in the article methods. The observed sample size parameters include the empirical*
326 *control-arm prevalence/incidence and k/ICC values in the first year of the trials. First year data was*
327 *utilised to estimate observed study power to account for temporal variations in transmission/cluster*
328 *heterogeneity. The remaining sample size parameters including clusters per arm, cluster size and*
329 *significance level were identical between the predicted and observed power calculations. Blue: study*
330 *power >80%. Red: Study power <80%.*

Trial	Desired effect size (%)	Predicted sample size parameters				Observed sample size parameters (year 1)				
		Control prevalence	Control incidence ^a	k	ICC	Power (%)	Control prevalence	Control incidence ^a	k ^b	ICC ^b
Samuels 2021	50	0.40	0.30		98	0.39	0.15			>99.9
Chaccour 2021	30		1.40	0.50	80		5.16	0.34		99
Staedke 2020	17.5	0.40	0.30		84	0.14	0.86			19
Bradley 2016	50	0.28	0.25		>99.9	0.23	0.59			45
Cisse 2016			0.30		n/a		0.66			n/a
Sluydts 2016	40	0.05	0.50		93	0.05	1.00			63
Bousema 2016	66.6	0.15	0.40		82	0.12	0.50			64
Pinder 2015	30		0.24	0.40	94		0.56	0.74		49
West 2014	33.3	0.09	0.25		84	0.26	0.26			98
Poespoprodjo 2021	66.6		0.30	0.05	89		0.86	0.07		>99.9
Foy 2019	28.8		2.25	0.02	87		2.49	0.059		61
Oldenburg 2018	30	0.10		0.075	n/a	0.51		0.1		n/a
von Seidlein 2019 (Vietnam) ^c	95	0.10		0.07	84	0.04		0.003		>99.9
von Seidlein 2019 (Cambodia) ^c	95	0.10		0.07	84	0.02		0.005		87
von Seidlein 2019 (Laos) ^c	95	0.10		0.07	84	0.11		0.035		99
von Seidlein 2019 (Myanmar) ^c	95	0.10		0.07	84	0.08		0.082		68

a: incidence per person per year

b: not always stated which subset of data used to calculate k/ICC

c: ICC retrospectively calculated at country level

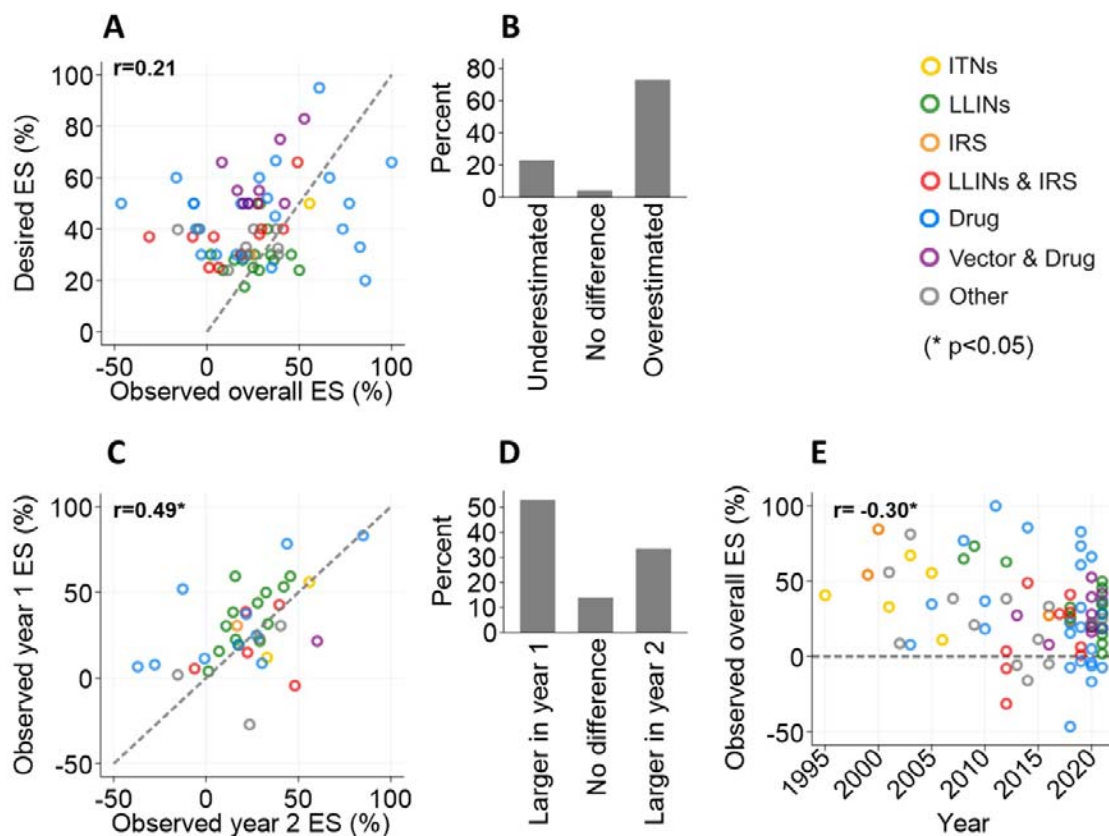
331

332 Effect size assumptions

333

334 Among the 71 included malaria CRTs, a total of 70 desired effect size estimates were accompanied
335 with empirical effect size estimates. We examined whether the desired effect sizes used in sample
336 size estimations corresponded to observed effect sizes in malaria CRTs. We identified no evidence of
337 a correlation between desired and observed effects sizes throughout the trial ($r=0.21$, $p: 0.09$) (Figure
338 4A). We also found that 72.7% (51/70) of desired effect sizes overestimated by more than a relative
339 10% difference (Figure 4B). We then explored factors that may have contributed to these findings.
340 Firstly, among trials that were conducted for at least 2 years (N:36), we found a strong positive
341 correlation between year 1 and year 2 observed effect sizes ($r=0.49$; $p:0.003$) (Figure 4C). Despite
342 this, effect sizes were more often higher in the first, compared to the second, year of each trial
343 (52.8% 19/36) (Figure 4C). Secondly, after comparing observed effect sizes to trial start date, a

344 negative correlation was seen ($r: -0.30, p < 0.05$), suggesting that effect sizes have been decreasing
 345 over time (Figure 4D).



346

347 **Figure 4:** Accuracy of desired versus observed effect size (ES) estimates in malaria CRTs. **A:** Correlation
 348 between the desired and overall observed effect size by type of intervention. Diagonal dash: line of
 349 equality. **B:** The percentage of desired effect size estimates that were underestimated (relative
 350 percentage difference $< 10\%$), no difference (relative percentage difference -10% to 10%) or
 351 overestimated (relative percentage difference $> 10\%$) according to overall observed effect size
 352 estimates. **C:** Correlation of observed effect size estimates by the 1st and 2nd year of the trial by type
 353 of intervention. Diagonal dash: line of equality **D:** The percentage of observed effect size estimates
 354 that were higher in the 1st or 2nd year of the trial (relative percentage difference $> 10\%$) or were no
 355 different (relative percentage difference $< 10\%$). **E:** **D:** Correlation between the overall observed effect
 356 size estimates versus the trial starting year by type of intervention. Horizontal dash: cut off line for
 357 positive effect size. Above this line epidemiological outcomes were lower in the intervention,
 358 compared to the control arms, of trials.

359 Discussion

360

361 Results from this review reveal malaria CRTs, measuring epidemiological outcomes, often rely on
 362 poorly defined sample size assumptions which results in compromised study power. Well powered
 363 trials need accurate information on predicted transmission intensity in the control arm, the

364 estimated heterogeneity of outcomes between or within clusters and desired effect size between
365 study arms. We found that transmission intensity and effect sizes were often over-estimated, with
366 measures of cluster heterogeneity commonly misclassified. To ensure future malaria CRTs are
367 adequately powered to detect the impacts of control interventions, efforts need to be made to
368 ensure sample size parameters are more reliably estimated at the trial design stage.

369 Our finding that most desired effect sizes in malaria CRTs were overestimated corresponds with
370 results from a separate review of 300 non-disease specific CRTs which found 68% of trials
371 experienced lower effect sizes than anticipated [21]. Authors speculated this over-estimation was
372 likely attributed to trials being powered to detect minimally important differences between study
373 arms and/or ineffective interventions being trialled. These are common challenges for malaria CRTs
374 too. A 30% effect size was previously documented as the threshold for an intervention to have public
375 health relevance and be cost-effective according to the WHO. These are likely highly ambitious
376 targets for certain interventions [2], particularly when being compared to already effective
377 interventions. Moreover, some trialists concluded their interventions were simply inadequate to curb
378 malaria transmission [12, 29, 30]. It should be noted that malaria CRTs have also suggested other
379 factors, unrelated to study power, that impeded their ability to demonstrate an impact including low
380 coverage/adherence [31, 32], inappropriate study settings [33, 34] and poor quality control [35].

381 In this review we further explored patterns in observed effect sizes among malaria CRTs and revealed
382 effect size estimates tended to be higher in the first compared to the second year of trials. This
383 implies the adherence and community-wide impact of certain trialled interventions wane over time.
384 For interventions such as bed nets, recent studies in Tanzania [36], Nigeria [37] and Nicaragua [38]
385 have demonstrated net coverage, usage, physical integrity and insecticidal activity all decreased
386 within a two-year period. Secondly, our results highlight observed effect sizes have, overall,
387 decreased since the 1990s. This is likely a consequence of trialled interventions being increasingly
388 layered over existing, widespread standard-of-care for malaria. Historically, control arms in malaria
389 CRTs consisted of either no or substandard interventions including untreated nets and placebo
390 treatments [3, 39, 40]. Recently however, control arms of trials typically include numerous, effective
391 malaria interventions [12, 41-43] and sometimes only differ from intervention arms with regards to
392 regimen [42, 44]. Together, these factors likely resulted in effect sizes being overstated.

393 Predicting malaria transmission intensity in the control arms of CRTs is challenging given the disease
394 is so spatially and temporally heterogeneous [24, 25]. Here, we found no evidence that estimating
395 control-arm transmission intensity using prior data provided a more accurate measures of prevalence
396 or incidence in control arms at the end of trials. Moreover, estimated transmission intensity

397 correlated with transmission more closely in the first, compared to the second, year of the trials. This
398 is likely the consequence of environmental, seasonal, socio-economic, and behavioural changes that
399 impact both human and mosquito populations [45, 46], and highlights the challenge in forecasting
400 short term malaria transmission patterns [47-51].

401 In this review only 20% of included malaria CRTs retrospectively calculated cluster heterogeneity
402 using trial data which resembles the previous review of 300 CRTs in general that found only 11%
403 provided empirical cluster heterogeneity estimates [21]. Moreover, the finding that the majority of
404 observed cluster heterogeneity measures differed to those inputted into sample size equations is
405 concerning as study power is so heavily impacted by between/within cluster correlation [15]. Future
406 malaria trials should adhere to CONSORT guidelines and provide empirical estimates of cluster
407 heterogeneity to both inform future trials and assist reviewers in determining whether trials are
408 adequately powered to detect their desired impact [21]. Moreover, given a recent secondary analysis
409 of a malaria CRT in Tanzania demonstrated temporal changes in within-cluster cluster heterogeneity
410 during the intervention period [52], providing empirical estimates of cluster heterogeneity at various
411 timepoints during trials may further help decipher whether trials were adequately powered
412 throughout the trial period [53]. As only a few malaria trials provided retrospective estimates of
413 k/ICC , we were unable to investigate whether basing estimates on prior data or not assists in
414 accurately characterising cluster heterogeneity. Moreover, it was not always clear which subset of
415 the data k/ICC measures referred to and investigators may have been prompted to present observed
416 cluster heterogeneity measures if the interventions failed to show impact. Consequently,
417 characterising the true degree of cluster heterogeneity among a representative sample of malaria
418 CRTs to inform future trials remains an imperative area of continued investigation.

419 **Conclusion**

420

421 Results from this review demonstrate the accuracy of epidemiological inputs in malaria CRT
422 sample/power size calculations require improvement. By simply reporting empirical cluster
423 heterogeneity measures alongside published results, in line with CONSORT guidelines, future trials
424 may be better informed to estimate suitable sample sizes. Determining trial transmission intensity
425 and heterogeneity in the control arm remains a larger challenge given the sporadic nature of malaria
426 transmission. Without more representative sample size parameters, future CRTs are at risk of being
427 underpowered to detect the impacts of vital, novel control tools against malaria.

428 **List of abbreviations**

429

430 CRT Cluster randomised trial
431 ES Effect size
432 ICC Intracluster correlation coefficient
433 IRS Indoor residual spraying
434 ITN Insecticide-treated bed nets
435 k coefficient of variation
436 LLIN Long-lasting insecticidal nets
437 pa Per annum
438 PCR Polymerase chain reaction
439 py Person-year
440 RDT Rapid diagnostic test
441 WHO World Health Organisation

442

443 **Declarations**

444

445 **Ethics approval and consent to participate**

446 Not applicable

447

448 **Consent for publication**

449 Not applicable

450

451 **Availability of data and materials**

452 The datasets generated and/or analysed during the current study are available from the
453 corresponding author on reasonable request.

454

455 **Competing interests**

456 The authors declare that they have no competing interests.

457 **Funding**

458 This research is supported by a grant to the London School of Hygiene and Tropical Medicine from
459 the Bill & Melinda Gates Foundation (INV-038132). JDC, JH, and TC also acknowledge funding from
460 the MRC Centre for Global Infectious Disease Analysis (reference MR/X020258/1), funded by the UK
461 Medical Research Council (MRC). This UK funded award is carried out in the frame of the Global
462 Health EDCTP3 Joint Undertaking. The funders had no role in study design, data collection and
463 analysis, decision to publish or preparation of the manuscript.

464

465 **Authors' contributions**

466 Study conception: JC and TC. Literature searching and extraction: JB and JH. Data analysis: JB and
467 JDC. Supervision: JC & TC. Manuscript preparation: JB. Manuscript editing and review: JB, JDC, TC and
468 JC. The authors read and approved the final manuscript.

469 **Acknowledgements**

470 Not applicable

471 **References**

472

- 473 1. WHO, *World malaria report 2023*. 2023, World Health Organisation: Geneva.
- 474 2. WHO, *How to design vector control efficacy trials: guidance on phase III vector control field*
475 *trial design*. 2017, World Health Organization.
- 476 3. Kroeger, A., et al., *Insecticide-impregnated bed nets for malaria control: varying experiences*
477 *from Ecuador, Colombia, and Peru concerning acceptability and effectiveness*. *Am J Trop Med*
478 *Hyg*, 1995. **53**(4): p. 313-23.
- 479 4. Nevill, C.G., et al., *Insecticide-treated bednets reduce mortality and severe morbidity from*
480 *malaria among children on the Kenyan coast*. *Tropical Medicine & International Health*,
481 2007. **1**(2): p. 139-146.
- 482 5. ter Kuile, F.O., et al., *Impact of permethrin-treated bed nets on malaria, anemia, and growth*
483 *in infants in an area of intense perennial malaria transmission in western Kenya*. *Am J Trop*
484 *Med Hyg*, 2003. **68**(4 Suppl): p. 68-77.
- 485 6. Tagbor, H., et al., *The clinical impact of combining intermittent preventive treatment with*
486 *home management of malaria in children aged below 5 years: cluster randomised trial*.
487 *Tropical Medicine & International Health*, 2011. **16**(3): p. 280-289.
- 488 7. Protopopoff, N., et al., *Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal*
489 *net and indoor residual spray interventions, separately and together, against malaria*
490 *transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two*
491 *factorial design trial*. *Lancet*, 2018. **391**(10130): p. 1577-1588.
- 492 8. Tiono, A.B., et al., *Efficacy of Olyset Duo, a bednet containing pyriproxyfen and permethrin,*
493 *versus a permethrin-only net against clinical malaria in an area with highly pyrethroid-*
494 *resistant vectors in rural Burkina Faso: a cluster-randomised controlled trial*. *Lancet*, 2018.
495 **392**(10147): p. 569-580.
- 496 9. Dron, L., et al., *The role and challenges of cluster randomised trials for global health*. *Lancet*
497 *Glob Health*, 2021. **9**(5): p. e701-e710.
- 498 10. Homan, T., et al., *The effect of mass mosquito trapping on malaria transmission and disease*
499 *burden (SolarMal): a stepped-wedge cluster-randomised trial*. *Lancet*, 2016. **388**(10050): p.
500 1193-201.

- 501 11. Samuels, A.M., et al., *Impact of Community-Based Mass Testing and Treatment on Malaria*
502 *Infection Prevalence in a High-Transmission Area of Western Kenya: A Cluster Randomized*
503 *Controlled Trial*. Clin Infect Dis, 2021. **72**(11): p. 1927-1935.
- 504 12. Sangoro, O., et al., *A cluster-randomized controlled trial to assess the effectiveness of using*
505 *15% DEET topical repellent with long-lasting insecticidal nets (LLINs) compared to a placebo*
506 *lotion on malaria transmission*. Malar J, 2014. **13**: p. 324.
- 507 13. Shekalaghe, S.A., et al., *A cluster-randomized trial of mass drug administration with a*
508 *gametocytocidal drug combination to interrupt malaria transmission in a low endemic area*
509 *in Tanzania*. Malar J, 2011. **10**: p. 247.
- 510 14. Sochantha, T., et al., *Insecticide-treated bednets for the prevention of Plasmodium falciparum*
511 *malaria in Cambodia: a cluster-randomized trial*. Trop Med Int Health, 2006. **11**(8): p. 1166-
512 77.
- 513 15. Hayes, R.J. and S. Bennett, *Simple sample size calculation for cluster-randomized trials*. Int J
514 Epidemiol, 1999. **28**(2): p. 319-26.
- 515 16. K Hemming, S.E., G Forbes, C Weijer, M Taljaard, *How to design efficient cluster randomised*
516 *trials*. Research methods and reporting, 2017. **358**.
- 517 17. Murray, D.M., et al., *Design and analysis of group-randomized trials in cancer: a review of*
518 *current practices*. J Natl Cancer Inst, 2008. **100**(7): p. 483-91.
- 519 18. Parker, K., et al., *Characteristics and practices of school-based cluster randomised controlled*
520 *trials for improving health outcomes in pupils in the UK: a systematic review protocol*. BMJ
521 Open, 2021. **11**(2): p. e044143.
- 522 19. Froud, R., et al., *Quality of cluster randomized controlled trials in oral health: a systematic*
523 *review of reports published between 2005 and 2009*. Community Dent Oral Epidemiol, 2012.
524 **40 Suppl 1**: p. 3-14.
- 525 20. Diaz-Ordaz, K., et al., *A systematic review of cluster randomised trials in residential facilities*
526 *for older people suggests how to improve quality*. BMC Med Res Methodol, 2013. **13**: p. 127.
- 527 21. Rutterford, C., et al., *Reporting and methodological quality of sample size calculations in*
528 *cluster randomized trials could be improved: a review*. J Clin Epidemiol, 2015. **68**(6): p. 716-
529 23.
- 530 22. Eldridge, S.M., D. Ashby, and S. Kerry, *Sample size for cluster randomized trials: effect of*
531 *coefficient of variation of cluster size and analysis method*. Int J Epidemiol, 2006. **35**(5): p.
532 1292-300.
- 533 23. Baidjoe, A.Y., et al., *Factors associated with high heterogeneity of malaria at fine spatial scale*
534 *in the Western Kenyan highlands*. Malar J, 2016. **15**: p. 307.
- 535 24. Gwitira, I., et al., *Spatial and spatio-temporal analysis of malaria cases in Zimbabwe*. Infect
536 Dis Poverty, 2020. **9**(1): p. 146.
- 537 25. Knudson, A., et al., *Spatio-temporal dynamics of Plasmodium falciparum transmission within*
538 *a spatial unit on the Colombian Pacific Coast*. Sci Rep, 2020. **10**(1): p. 3756.
- 539 26. Selvaraj, P., E.A. Wenger, and J. Gerardin, *Seasonality and heterogeneity of malaria*
540 *transmission determine success of interventions in high-endemic settings: a modeling study*.
541 BMC Infect Dis, 2018. **18**(1): p. 413.
- 542 27. Cisse, B., et al., *Effectiveness of Seasonal Malaria Chemoprevention in Children under Ten*
543 *Years of Age in Senegal: A Stepped-Wedge Cluster-Randomised Trial*. PLoS Med, 2016.
544 **13**(11): p. e1002175.
- 545 28. Oldenburg, C.E., et al., *Annual Versus Biannual Mass Azithromycin Distribution and Malaria*
546 *Parasitemia During the Peak Transmission Season Among Children in Niger*. Pediatr Infect Dis
547 J, 2018. **37**(6): p. 506-510.
- 548 29. Tiono, A.B., et al., *A controlled, parallel, cluster-randomized trial of community-wide*
549 *screening and treatment of asymptomatic carriers of Plasmodium falciparum in Burkina*
550 *Faso*. Malar J, 2013. **12**: p. 79.

- 551 30. Sutanto, I., et al., *Negligible Impact of Mass Screening and Treatment on Mesoendemic*
552 *Malaria Transmission at West Timor in Eastern Indonesia: A Cluster-Randomized Trial.* Clin
553 Infect Dis, 2018. **67**(9): p. 1364-1372.
- 554 31. Tagbor, H., et al., *The clinical impact of combining intermittent preventive treatment with*
555 *home management of malaria in children aged below 5 years: cluster randomised trial.* Trop
556 Med Int Health, 2011. **16**(3): p. 280-9.
- 557 32. Desai, M.R., et al., *Impact of Intermittent Mass Testing and Treatment on Incidence of*
558 *Malaria Infection in a High Transmission Area of Western Kenya.* Am J Trop Med Hyg, 2020.
559 **103**(1): p. 369-377.
- 560 33. Bousema, T., et al., *The Impact of Hotspot-Targeted Interventions on Malaria Transmission in*
561 *Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled*
562 *Trial.* PLoS Med, 2016. **13**(4): p. e1001993.
- 563 34. Magris, M., et al., *Community-randomized trial of lambda-cyhalothrin-treated hammock nets*
564 *for malaria control in Yanomami communities in the Amazon region of Venezuela.* Trop Med
565 Int Health, 2007. **12**(3): p. 392-403.
- 566 35. Keating, J., et al., *Evaluating indoor residual spray for reducing malaria infection prevalence*
567 *in Eritrea: results from a community randomized control trial.* Acta Trop, 2011. **119**(2-3): p.
568 107-13.
- 569 36. Lorenz, L.M., et al., *Comparative functional survival and equivalent annual cost of 3 long-*
570 *lasting insecticidal net (LLIN) products in Tanzania: A randomised trial with 3-year follow up.*
571 PLoS Med, 2020. **17**(9): p. e1003248.
- 572 37. Obi, E., et al., *Monitoring the physical and insecticidal durability of the long-lasting*
573 *insecticidal net DawaPlus((R)) 2.0 in three States in Nigeria.* Malar J, 2020. **19**(1): p. 124.
- 574 38. Villalta, E.L., et al., *Evaluation of the durability and use of long-lasting insecticidal nets in*
575 *Nicaragua.* Malar J, 2021. **20**(1): p. 106.
- 576 39. Misra, S.P., et al., *Malaria control: bednets or spraying? Spray versus treated nets using*
577 *deltamethrin--a community randomized trial in India.* Trans R Soc Trop Med Hyg, 1999. **93**(5):
578 p. 456-7.
- 579 40. von Seidlein, L., et al., *The effect of mass administration of sulfadoxine-pyrimethamine*
580 *combined with artesunate on malaria incidence: a double-blind, community-randomized,*
581 *placebo-controlled trial in The Gambia.* Trans R Soc Trop Med Hyg, 2003. **97**(2): p. 217-25.
- 582 41. Bradley, J., et al., *A cluster randomized trial comparing deltamethrin and bendiocarb as*
583 *insecticides for indoor residual spraying to control malaria on Bioko Island, Equatorial*
584 *Guinea.* Malar J, 2016. **15**(1): p. 378.
- 585 42. Poespoprodjo, J.R., et al., *Supervised versus unsupervised primaquine radical cure for the*
586 *treatment of falciparum and vivax malaria in Papua, Indonesia: a cluster-randomised,*
587 *controlled, open-label superiority trial.* Lancet Infect Dis, 2022. **22**(3): p. 367-376.
- 588 43. Chaccour, C., et al., *Incremental impact on malaria incidence following indoor residual*
589 *spraying in a highly endemic area with high standard ITN access in Mozambique: results from*
590 *a cluster-randomized study.* Malar J, 2021. **20**(1): p. 84.
- 591 44. Foy, B.D., et al., *Efficacy and risk of harms of repeat ivermectin mass drug administrations for*
592 *control of malaria (RIMDAMAL): a cluster-randomised trial.* Lancet, 2019. **393**(10180): p.
593 1517-1526.
- 594 45. Deutsch-Feldman, M., et al., *Spatial and epidemiological drivers of Plasmodium falciparum*
595 *malaria among adults in the Democratic Republic of the Congo.* BMJ Glob Health, 2020. **5**(6).
- 596 46. Yamba, E.I., et al., *Climate Drivers of Malaria Transmission Seasonality and Their Relative*
597 *Importance in Sub-Saharan Africa.* Geohealth, 2023. **7**(2): p. e2022GH000698.
- 598 47. Bath, D., et al., *Effectiveness and cost-effectiveness of reactive, targeted indoor residual*
599 *spraying for malaria control in low-transmission settings: a cluster-randomised, non-*
600 *inferiority trial in South Africa.* The Lancet, 2021. **397**(10276): p. 816-827.

- 601 48. Merkord, C.L., et al., *Integrating malaria surveillance with climate data for outbreak*
602 *detection and forecasting: the EPIDEMIA system*. Malar J, 2017. **16**(1): p. 89.
- 603 49. Pourtois, J.D., et al., *Climatic, land-use and socio-economic factors can predict malaria*
604 *dynamics at fine spatial scales relevant to local health actors: Evidence from rural*
605 *Madagascar*. PLOS Glob Public Health, 2023. **3**(2): p. e0001607.
- 606 50. Tompkins, A.M., et al., *Dynamical Malaria Forecasts Are Skillful at Regional and Local Scales*
607 *in Uganda up to 4 Months Ahead*. Geohealth, 2019. **3**(3): p. 58-66.
- 608 51. Zinszer, K., et al., *A scoping review of malaria forecasting: past work and future directions*.
609 *BMJ Open*, 2012. **2**(6).
- 610 52. Ouyang, Y., et al., *Accounting for complex intracluster correlations in longitudinal cluster*
611 *randomized trials: a case study in malaria vector control*. BMC Med Res Methodol, 2023.
612 **23**(1): p. 64.
- 613 53. Pinder, M., et al., *Efficacy of indoor residual spraying with dichlorodiphenyltrichloroethane*
614 *against malaria in Gambian communities with high usage of long-lasting insecticidal*
615 *mosquito nets: a cluster-randomised controlled trial*. Lancet, 2015. **385**(9976): p. 1436-46.
- 616