RESEARCH   **Open Access**

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

Joseph Biggs[1*], Joseph D. Challenger[2], Joel Hellewell[2], Thomas S. Churcher[2] and Jackie Cook[1]

Abstract

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

Methods We searched for published malaria CRTs using four online databases in March 2022. Eligible trials included those with malaria-specific epidemiological outcomes which randomised at least six geographical clusters to study arms. Predicted and observed sample size parameters were extracted by reviewers for each trial. Pair-wise Spearman's correlation coefficients (r_s) were calculated to assess the correlation between predicted and observed control-arm outcome measures and effect sizes (relative percentage reductions) between arms. Among trials which retrospectively calculated an estimate of heterogeneity in cluster outcomes, we recalculated study power according to observed trial estimates.

Results Of the 1889 records identified and screened, 108 articles were eligible and comprised of 71 malaria CRTs. Among 91.5% (65/71) of trials that included sample size calculations, most estimated cluster heterogeneity using the coefficient of variation (k) (80%, 52/65) which were often predicted without using prior data (67.7%, 44/65). Predicted control-arm prevalence moderately correlated with observed control-arm prevalence (r_s: 0.44, [95%CI: 0.12,0.68], p-value < 0.05], with 61.2% (19/31) of prevalence estimates overestimated. Among the minority of trials that retrospectively calculated cluster heterogeneity (20%, 13/65), empirical values contrasted with those used in sample size estimations and often compromised study power. Observed effect sizes were often smaller than had been predicted at the sample size stage (72.9%, 51/70) and were typically higher in the first, compared to the second, year of trials. Overall, effect sizes achieved by malaria interventions tested in trials decreased between 1995 and 2021.

*Correspondence:
Joseph Biggs
joseph.biggs1@lshtm.ac.uk
Full list of author information is available at the end of the article

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

Registration This review is registered with PROSPERO (CRD42022315741).

Keywords Malaria, Cluster randomised trial, Sample size, Heterogeneity, Coefficient of variation, Intracluster correlation coefficient

Introduction

Malaria is a parasitic disease that in 2022 was responsible for the deaths of 608,000 individuals worldwide, most of whom were children in Sub-Saharan Africa [1]. There are numerous, effective interventions that can be used to combat malaria transmission that are recommended by the World Health Organisation (WHO). To generate evidence for the recommendation of these tools, cluster randomised trials (CRTs) are conducted to demonstrate the community-wide effects [2]. Historically, CRTs have demonstrated the mass effects of insecticide-treated bed nets (ITNs) [3–5], mass chemoprevention strategies [4, 6], long-lasting insecticide treated nets (LLINs) [7, 8], and in the future, will be essential to evaluate the herd effect of novel malaria vaccines [1, 9, 10]. Despite their necessity, CRTs are subject to major constraints. Trialling interventions over large geographical areas is costly, logistically challenging, and at the design stage, requires well-defined estimates of underlying transmission patterns in the study setting [11]. Consequently, in recent decades, some malaria CRTs have reported being underpowered and have presented inconclusive findings [12–16].

Triallists determine the sample size of CRTs according to power calculations that consider cluster-randomisation, where groups of people, as opposed to individuals, are randomised to receive interventions. This design can result in heterogeneity of outcomes between and within clusters owing to groups of individuals, such as households, schools, and geographical areas, sharing similar biological and socio-economic characteristics which introduces correlation in study outcomes [17, 18]. Consequently, cluster heterogeneity needs to be incorporated into sample size estimations, along with expected control arm transmission and effect size estimates (relative percentage reductions between arms), to compensate for the lower precision associated with this design. The between- and within-cluster heterogeneity can be measured using the coefficient of variation (k) or intracluster correlation coefficient (ICC), respectively, and heavily impacts trial size [17, 18]. Trialling new interventions in areas with missing or inadequate data results in investigators having to rely on judgement-based estimates for their sample size estimations which may be inaccurate.

Numerous reviews have evaluated sample size estimations in CRTs focused on cancer treatments [19], school-based interventions [20], oral health [21], residential care [22] and CRTs in general [23]. These reviews highlighted that despite trials mostly including sample size estimations, not all calculations accounted for cluster heterogeneity (73% [19], 78% [20], 71% [21] and 47% [22]). Two of these reviews further explored whether trials included empirical measures of cluster heterogeneity and compared them to prior estimates [20, 23]. Reviews highlighted trials rarely provided retrospective estimates of cluster heterogeneity (<40%), and among trials that did, large differences were identified between predicted and observed estimates. This suggests many trialists misclassified the true degree of cluster heterogeneity at the design stage. Finally, one review explored which trials stated their desired effect sizes and compared them to those observed [23]. They showed that 68% of predicted effect sizes were overestimated which is a concerning finding given larger sample sizes are required to detect smaller effect sizes. Interestingly, none of these reviews compared the outcome measures predicted and observed in the control arms of their included trials. This is crucial as misclassification of predicted effect size, cluster heterogeneity and control-arm outcome measurements all impact study power [17, 18, 24].

Malaria transmission is driven by numerous environmental and socio-economic factors including rainfall, temperature, vegetation cover, type of housing and provision of malaria interventions [25–28]. Consequently, transmission is often spatially and temporally variable across various geographical scales. This presents a challenge for malaria CRTs as heterogenous transmission in the community may result in spatial/temporal variability in malaria-specific outcomes between geographical clusters. Therefore, estimating the level of malaria transmission in the control arm and the degree of cluster heterogeneity for malaria CRT sample size estimations is difficult in the absence of baseline data.

In this review, our aim was to investigate the characteristics and quality of sample size estimations in malaria CRTs that used geographical clusters. Specifically, we explored whether triallists accurately predicted sample size estimation parameters, including control-arm transmission, cluster heterogeneity, and predicted effect sizes,

according to observed measurements during trials. It is hoped results from this review will improve future study design and ensure trials are able to accurately detect statistically significant effects of interventions and guide evidence-based implementation.

Methods

Search strategy and selection criteria

We conducted a systematic review of published malaria CRTs with epidemiological outcomes. In March 2022, we searched the database systems Pubmed, Web of Science, Embase and Cochrane reviews using truncated versions of the terms 'malaria' and 'cluster randomised trial' for trials published in English language. The bibliographies of identified reviews were additionally screened according to title and abstract. Search results were imported into the reference manager Endnote where digitally identified duplicates were removed. Manually identified duplicates were removed by two reviewers (JB & JH). Pre-determined eligibility criteria were used to screen identified articles based on title and abstract (JB & JH) while screening discordance was adjudicated by consensus (TC & JC). Identified studies were eligible for inclusion if they met the following criteria: the study was a CRT wherein at least six geographical clusters were randomised to intervention/control arms and the trial measured malaria-specific epidemiological outcomes. A minimum of six clusters were chosen as this represents the approximate number that can be used to obtain a statistically significant result [17, 18]. Malaria-specific outcomes include malaria prevalence or incidence according to microscopy, rapid diagnostic tests (RDTs), or molecular methods. Trials that only measured anaemia and all-cause mortality were excluded as these outcomes could be attributed to other conditions. Prior to study initiation, the review was registered in PROSPERO on 9th March 2022 (CRD42022315741).

Data extraction

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

Data analysis

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

$$Prevalence\ effect\ size\ =\ 1 - p_2/p_1 \qquad (A)$$

$$Incidence\ effect\ size\ =\ 1 - r_2/r_1 \qquad (B)$$

To determine the accuracy of predicted sample size parameters in malaria CRTs, we estimated the strength of association and relative percentage difference between the predicted and observed control-arm prevalence/incidence and effect size estimates. To quantify the strength of association, we firstly used Shapiro-Wilk tests to investigate whether predicted and observed parameters followed a normal distribution using the 'swilk' command in STATA (v.18). Test p-values ($p < 0.05$) were used to reject the null hypothesis parameters were normally distributed. Pair-wise Spearman's rank correlation coefficients (r_s, p-values) were then estimated to quantify the strength of association between predicted and observed values using the 'spearman' command in STATA (v.18). Coefficient (r_s) 95% confidence intervals (95%CIs) referred to bias-corrected intervals derived from bootstraps with 2000 repetitions according to the STATA (v.13) 'bootstrap' command. To determine whether predicted sample size parameters were over or underestimated, we calculated the relative percentage difference between predicted and observed values. Relative percentage differences > 10% were considered truly different.

Regarding cluster heterogeneity estimates (k/ICC) provided in trials, we first investigated whether predicted estimates based on prior/baseline data differed to estimates predicted with no data. A pair-wise t-test was used to determine whether the mean value difference equalled zero ($p < 0.05$). Among trials that reported cluster heterogeneity using observed trial data, we recalculated study power (%) according to the observed k/ICC and year 1 control arm prevalence/incidence. The remaining sample size parameters used were identical to the original power calculations: predicted effect size (%), cluster size, cluster number and significance level (%). Study power for CRTs was calculated according to methods described by Hayes and Moulton in [17]. All analyses were conducted in STATA (v.18, Texas, USA).

Results

Our literature search yielded 1889 records from database searching and 145 records from the bibliographies of Cochrane reviews (Fig. 1). Following the removal of duplicates, a total of 1302 records were screened after which 991 were excluded as they were not concerned with malaria CRTs. The remaining 311 records were assessed for eligibility resulting in 108 published articles being included in this study. These articles included trial protocols (*n*=26), baseline results (*n*=3), main results of trials (*n*=71) and secondary results of trials (*n*=8). Together, included articles referred to 71 epidemiological malaria CRTs (Additional file 1). The review PRISMA 2020 checklist is included in additional file 2.

The trial-level characteristics of the 71 malaria CRTs are shown in Table 1; Fig. 2A-G. Since 1995, malaria-specific CRTs have increased in frequency and, overall, have been conducted in a total of 78 countries across Africa (n:53), Asia (n:21) and South America (n:4). 55% (39/71) of the trials evaluated vector control interventions, all measured *Plasmodium falciparum* outcomes while 27%

(19/71) also measured *Plasmodium vivax* outcomes. Most trials adopted a parallel design (86%, 61/71) and consisted of two study arms (76%, 55/71). Concerning the cluster randomisation procedures, 39% (28/71) used simple randomisation to allocate clusters, 24% (17/71) implemented stratified randomisation, 23% (16/71) employed restricted randomisation, while 13% (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 and 4 restriction criteria (Fig. 2F). The most common restriction criteria used for randomisation included cluster transmission intensity (prevalence or incidence), cluster size, location, and historical intervention coverage (Fig. 2G). Regarding cluster design, 75% (53/71) adopted a basic cluster design, 14% (10/71) used a 'fried egg' design and 11% (8/71) reported ensuring a minimum buffer distance between clusters. Among trials that reported their minimum cluster buffer size, 73% (8/11) reported a minimum buffer size<2 km while those
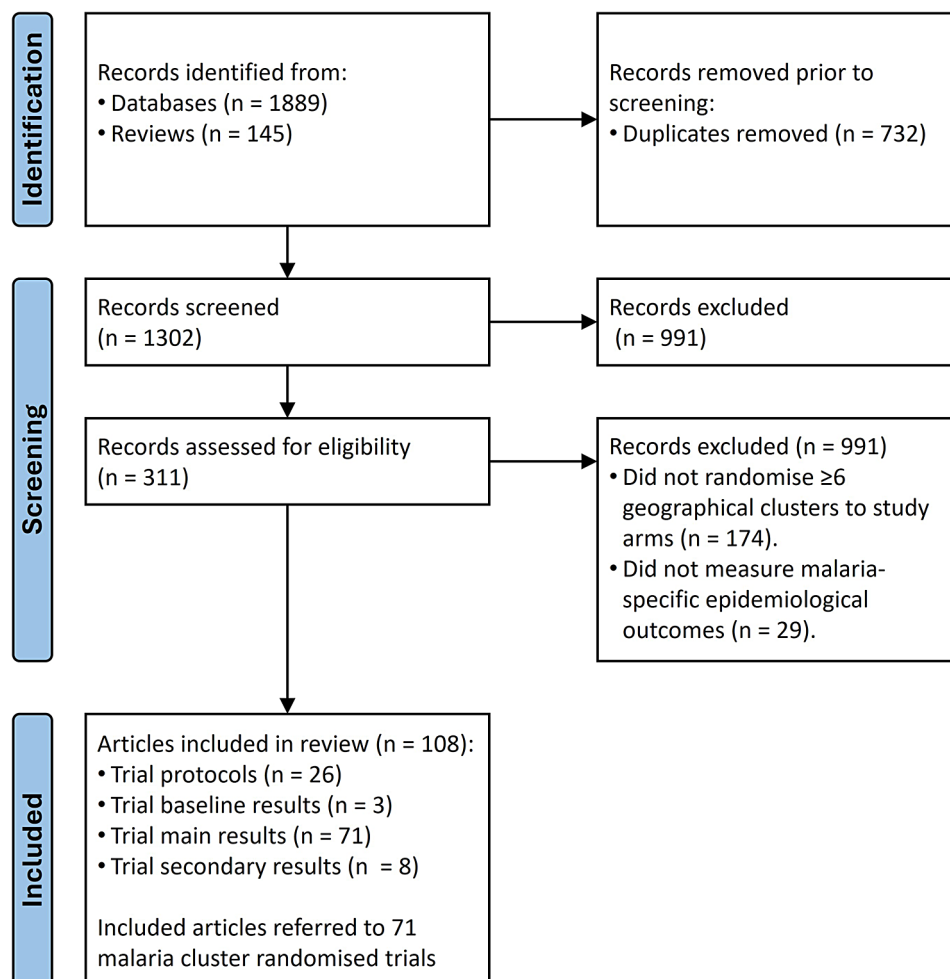

Fig. 1 Study selection of included epidemiological malaria CRTs

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		
P. falciparum	52	73.2
P. falciparum & P. vivax	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

who stated a minimum cluster separation reported a range between no separation and 3 km.

Among the included 71 CRTs, a total of 65 formal cluster sample size estimations were conducted that accounted for cluster heterogeneity by including a k, ICC or design effect component. Of these, 34/65 were based on incidence while 31/65 were based on prevalence (Table 2; Fig. 2H-O). The remaining trials either did not account for cluster heterogeneity or lacked any sample size justification. Over 90% of all sample size estimations were calculated to achieve power between 80 and 90% (59/65) at the 5% significance level (60/65). Concerning the epidemiological outcome measures in the control arm, most trials predicted incidence using prior data (71%, 24/34) or predicted prevalence without using prior data (55%, 17/31). Regarding sample size estimations based on prevalence (n: 31), investigators estimated a range of prevalences in the control arm (mean: 0.21, range: 0.05–0.48) and desired effect sizes (mean: 47.1% range: 17.5–95%) which tended to be higher in low prevalence settings (Fig. 2H). The average required cluster sample size was 104.5 individuals (Median 80; Fig. 2I) and average required number of clusters per arm equalled 17.5 (Fig. 2J). For sample size estimations based on incidence (n: 34), a range of incidence estimates in the control arm were estimated (range: 0.002–2.6 cases per person per annum). Desired effect sizes were similarly higher in lower incidence settings (mean: 41.1%, range 20–93%) (Fig. 2L). The average cluster size for incidence was 415 (median: 125) person years (Fig. 2M) while the mean number of required clusters per arm was 17.6 (Fig. 2N). Shapiro-Wilk tests provided strong evidence predicted prevalence, incidence and effect size distributions followed a non-normal distribution ($p < 0.05$) (Additional file 3).

The most common cluster heterogeneity measure used in malaria CRT sample size calculations was the coefficient of variation (k) (80% 52/65). 68% (44/65) of estimated cluster heterogeneity measures were estimated with no prior data while only 32% (21/65) were estimated using baseline or pilot study data. Lastly, only a minority of investigators retrospectively calculated cluster heterogeneity using trial data (20%, 13/65) (Table 2).

Control arm transmission intensity assumptions

We explored how accurately epidemiological outcomes were predicted in the control arms of included trials (prevalence n: 31, incidence n: 34). Overall, control-arm predicted prevalence was moderately positively correlated with observed prevalence (r_s: 0.44 [95%CI: 0.12–0.68], $p < 0.05$) (Fig. 3A) while predicted and empirical incidence was strongly positively correlated (r_s: 0.76 [95%CI: 0.49–0.90], $p < 0.05$) (Fig. 3B). Moreover, most predicted prevalence and incidence estimates were overestimated by more than 10% according to observed estimates (prevalence overestimation: 61% (19/31), incidence overestimation: 50% (17/34)) (Fig. 3C&D). We also assessed whether relying on prior data improved predictions of control-arm prevalence and incidence. Predicted control-arm incidence was slightly more associated with observed control-arm incidence when estimates were predicted with data, compared to, without data (Fig. 3B). However, this trend was unobserved for prevalence (Fig. 3A). Lastly, we found that predicted incidence
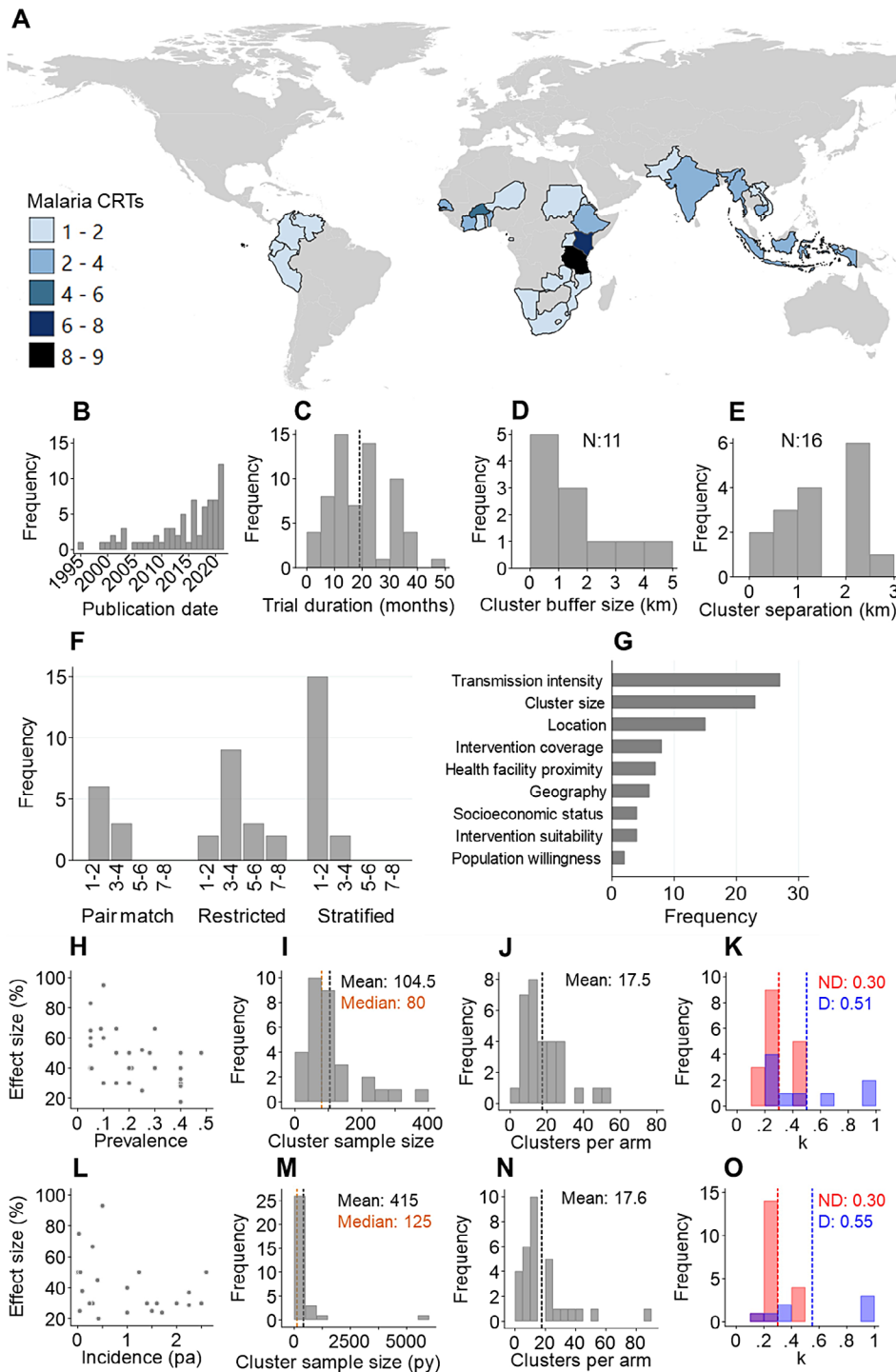
Fig. 2 **Characteristics of malaria CRTs identified in this review.A**: Distribution of malaria CRTs. **B**: Annual frequency of malaria CRTs. **C**: Overall duration of malaria CRTs (dash line: mean). **D**: Size of buffers around study clusters. **E**: Minimum separation between study clusters. **F**: Number of restriction criteria used according to the type of trial randomisation strategy. **G**: The most utilised restriction criteria in malaria CRTs. Population willingness refers to population acceptance of trialled interventions. **Sample size assumptions used in trials with prevalence as the outcome measure:H**: Predicted control-arm prevalence compared to predicted effect size. **I**: The desired total number of individuals surveyed per cluster. **J**: Required number of clusters per arm for prevalence outcomes. **K**: The predicted coefficient of variation (D) for prevalence sample size calculations among trials stratified by whether values were estimating using prior or baseline data (D) or assumed using no data (ND). Vertical dash: mean. **Sample size assumptions used in trials with incidence as the outcome measure:L**: Predicted control-arm incidence per person compared to predicted effect size (p.a.: per annum). **M**: The desired person-years per cluster. **N**: Required number of clusters per arm for incidence outcomes. **O**: The predicted coefficient of variation (K) for incidence sample size calculations among trials stratified by whether values were based on prior or baseline data (D) or assumed using no data (ND). Vertical dash: mean

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

Sample size	Overall		Incidence		Prevalence	
Characteristics	N:65		N:34		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

strongly correlated with empirical incidence in the first and second years of trials (Fig. 3F). In contrast, Predicted control-arm prevalence was only moderated correlated with observed prevalence in the first year of trials and weakly correlated with second year trial prevalence (Fig. 3E). These results demonstrate investigators tended to poorly predict and overestimate control-arm prevalence in malaria CRTs.

Cluster heterogeneity assumptions

Among trials that utilised the coefficient of variation (k) to account for cluster heterogeneity of incidence/prevalence in their sample size estimations (Table 2), a range of values were used (mean: 0.37, range: 0.1–1.0)

(Fig. 2K&O). Values of k predicted using prior data were, on average, statistically higher than those predicted with no prior data (no prior data mean k: 0.30; prior data mean k: 0.52, t-test p-value<0.05). This suggests k was likely underestimated in many trials. A small number of trials used the ICC to account for cluster heterogeneity, which similarly had a large range (mean: 0.12, range: 0.006–0.40) (Additional file 4).

Among the trials that additionally calculated k/ICC at the end of the study using empirical data (20% 13/65), we explored whether predicted cluster heterogeneity estimates were accurate and used empirical values to recalculate study power (Table 3). Empirical cluster heterogeneity estimates often differed to those used in
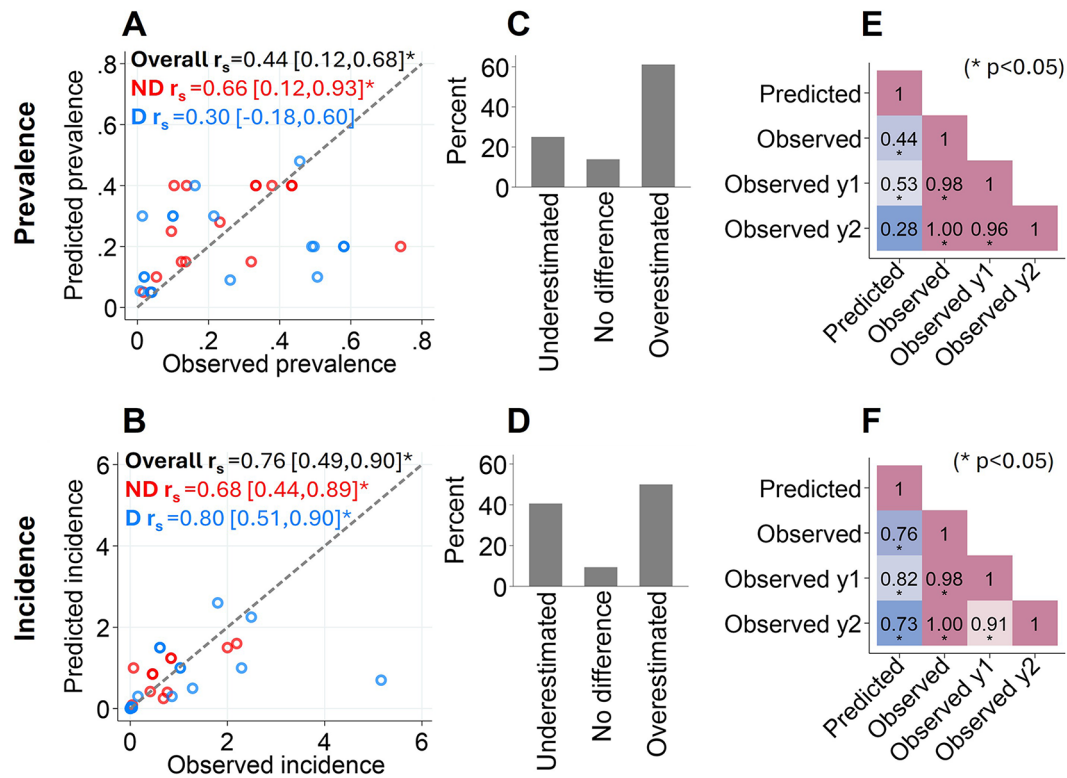
Fig. 3 Accuracy of predicted versus observed prevalence and incidence outcomes in malaria CRT control arms. **A** & **B**: Correlation between the predicted and overall observed prevalence/incidence stratified by method used to predict estimates: using data (D; blue) using no data (ND; red) and overall (black). **C** & **D**: The percentage of predicted prevalence/incidence estimates that were underestimated (relative percentage difference <-10%), no difference (relative percentage difference − 10–10%) or overestimated (relative percentage difference > 10%) according to overall observed estimates. **E** & **F**: Correlation matrix comparing the predicted prevalence/incidence with estimates observed throughout the trial (observed), in year 1 (Observed y1) and in year 2 (Observed y2). r_s: Spearman's rank correlation coefficient. Brackets: r_s 95%CIs. *: r_s p-value < 0.05

sample size estimations with the majority underestimating k/ICC (62% 8/13). Among 11/13 of these trials we were able to replicate original power calculations, we additionally recalculated study power according to empirical k/ICC values and control-arm prevalence/incidence. Recalculated power for 7/11 trials was below 80%. For 4/11 trials, cluster heterogeneity was overestimated which resulted in them remaining suitably powered to detect their desired effect sizes. It should be noted it was not always stated which timepoint/subset of trial data was used to retrospectively calculate k/ICC.

Effect size assumptions

Among the 71 included malaria CRTs, a total of 70 desired sample size effect size estimates were accompanied with empirical effect size estimates. We examined how accurately trials predicted these measures according to corresponded to observed effect sizes. Overall, we identified a weak, statistical insignificant, positive correlation between predicted and observed effect size estimates (r_s: 0.19, [95%CI: -0.04-0.42], p-value:0.11) (Fig. 4A). Furthermore, 73% (51/70) of desired effect sizes were overestimated by > 10% (Fig. 4B). We explored

factors that may have contributed to this overestimation. Firstly, among trials that were conducted for at least 2 years (N:36), we found a moderate positive correlation between year 1 and 2 observed effect sizes (r_s:0.44, [95%CI:0.06–0.74], p-value:0.007) (Fig. 4C) which were typically larger in year 1 (52.8% 19/36) (Fig. 4C). Secondly, we identified a weak, yet statistically significant, negative correlation between overall observed effect sizes by trial start date (r_s: -0.22, [95%CI:-0.42,-0.01], p-value:0.045) revealing effect sizes have decreased over previous decades (Fig. 4D).

Discussion

Results from this review reveal malaria CRTs, measuring epidemiological outcomes, often rely on poorly defined sample size assumptions which results in compromised study power. Well powered trials need accurate information on predicted transmission intensity in the control arm, the estimated heterogeneity of outcomes between or within clusters and desired effect size between study arms. We found that transmission intensity and effect sizes were often over-estimated, with measures of cluster heterogeneity commonly misclassified. To ensure future

Table 3 The study power (%) to detect desired effect sizes according to predicted (left) and observed (right) sample size parameters among trials that retrospectively calculated cluster heterogeneity. The predicted sample size parameters include the predicted control-arm prevalence/incidence and the k/ICC values stated in the article methods. The observed sample size parameters include the empirical control-arm prevalence/incidence and k/ICC values in the first year of the trials. First year data was utilised to estimate observed study power to account for temporal variations in transmission/cluster heterogeneity. The remaining sample size parameters including clusters per arm, cluster size and significance level were identical between the predicted and observed power calculations. Blue: study 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]	Power (%)
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

malaria CRTs are adequately powered to detect the impacts of control interventions, efforts need to be made to ensure sample size parameters are more reliably estimated at the trial design stage.

Our finding that most desired effect sizes in malaria CRTs were overestimated corresponds with results from a separate review of 300 non-disease specific CRTs which found 68% of trials measured lower effect sizes than anticipated [23]. Authors speculated this over-estimation was likely attributed to trials being powered to detect minimally important differences between study arms and/or ineffective interventions being trialled. These are common challenges for malaria CRTs too. A 30% effect size was previously documented as the threshold for an intervention to have public health relevance and be cost-effective according to the WHO. These are likely highly ambitious targets for certain interventions [2], particularly when being compared to already effective interventions. We speculate ambitious desired effect sizes estimates in some cases may have been necessary to acquire funding, which in turn, resulted in null results that we argue should not be interpreted as failure to demonstrate effect, but a consequence of unrealistic expectation. However, it should be noted some trials did conclude their interventions were simply inadequate to curb malaria transmission [14, 29, 30]. While other trials suggested null results were a consequence of low coverage/adherence [31, 32], inappropriate study settings [33, 34] and poor quality control [35].

In this review we further explored patterns in observed effect sizes among malaria CRTs and revealed effect size estimates tended to be higher in the first compared to the second year of trials. This implies the adherence and community-wide impact of certain trialled interventions wane over time. For interventions such as bed nets, recent studies in Tanzania [36], Nigeria [37] and Nicaragua [38] have demonstrated net coverage, usage, physical integrity and insecticidal activity all decreased within a two-year period. Secondly, our results highlight observed effect sizes have, overall, decreased since the 1990s. This is likely a consequence of trialled interventions being increasingly layered over existing, widespread standard-of-care for malaria. Historically, control arms in malaria CRTs consisted of either no or substandard interventions including untreated nets and placebo treatments [3, 39, 40]. Recently however, control arms of trials typically include numerous, effective malaria interventions [14, 41–43] and sometimes only differ from intervention arms with regards to regimen [42, 44]. Together, these factors likely resulted in effect sizes being overstated and put into question the suitability of superiority trials for evaluating some malaria interventions. For interventions that differ slightly from existing practise, non-inferiority trials may be worth consideration. Although non-inferiority
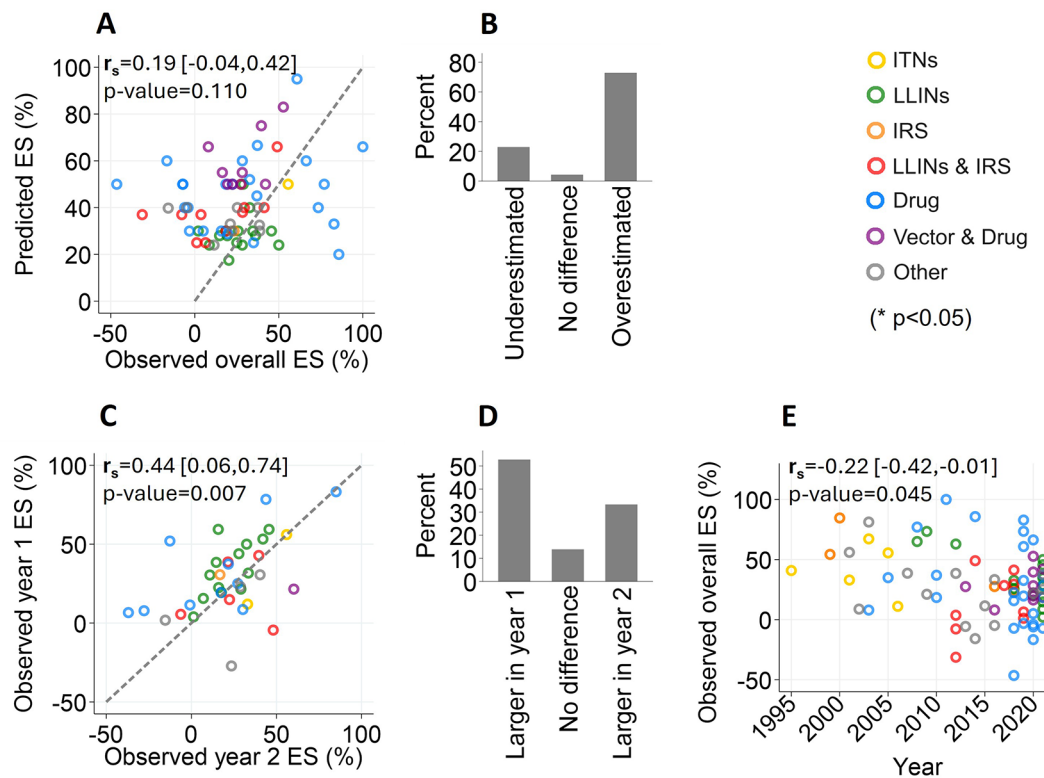
Fig. 4 **Accuracy of predicted versus observed effect size (ES) estimates in malaria CRTs**. **A**: Correlation between the predicted and overall observed effect size by type of intervention. Diagonal dash: line of equality. **B**: The percentage of predicted effect size estimates that were underestimated (relative percentage difference <-10%), no difference (relative percentage difference −10–10%) or overestimated (relative percentage difference >10%) according to overall observed effect size estimates. **C**: Correlation of observed effect size estimates by the 1st and 2nd year of the trial by type of intervention. Diagonal dash: line of equality **D**: The percentage of observed effect size estimates that were higher in the 1st or 2nd year of the trial (relative percentage difference > 10%) or were no different (relative percentage difference < 10%). **E**: **D**: Correlation between the overall observed effect size estimates versus the trial starting year by type of intervention. r_s: Spearman's rank correlation coefficient. Brackets: r_s 95%CIs

margins should be carefully informed by clinically and economically relevant guidelines in study settings [45].

Predicting malaria transmission intensity in the control arms of CRTs is challenging given the disease is so spatially and temporally heterogeneous [26, 27]. Here, we found weak evidence that estimating control-arm transmission intensity using prior data improved predictions of prevalence or incidence. Moreover, estimated transmission intensity correlated with transmission more closely in the first, compared to the second, year of the trials. This is likely the consequence of environmental, seasonal, socio-economic, and behavioural changes that impact both human and mosquito populations [46, 47], and highlights the challenge in forecasting short term malaria transmission patterns [48–52]. Moreover, contamination of interventions between arms [53], increased trial participation, higher intervention uptake and improved availability of existing control tools over time may have also influenced control-arm outcome measurements. These trends similarly could explain why effect sizes typically decreased between the first and second years of malaria CRTs.

In this review only 20% of included malaria CRTs retrospectively calculated cluster heterogeneity using trial data which resembles the previous review of 300 CRTs in general that found only 11% provided empirical cluster heterogeneity estimates [23]. Moreover, the finding that the majority of observed cluster heterogeneity measures differed to those inputted into sample size equations is concerning as this resulted in trials being either over or underpowered [17]. Both scenarios are problematic as overpowered trials are unnecessarily large and thus waste a proportion of their resources and needlessly expose a larger proportion of the community to the intervention. Underpowered studies however have a low chance of identifying statistically significant effects, which again represents a waste of resources but also is associated with serious ethical implications as communities are enrolled into a trial that will likely produce null results [11, 54]. Future malaria trials should therefore adhere to CON-SORT guidelines and provide empirical estimates of cluster heterogeneity to both inform future trials and assist reviewers in determining whether trials are adequately powered to detect their desired impact [23]. Moreover,

given a recent secondary analysis of a malaria CRT in Tanzania demonstrated temporal changes in within-cluster cluster heterogeneity during the intervention period [55], providing empirical estimates of cluster heterogeneity at various timepoints during trials may further help decipher whether trials were adequately powered throughout [56]. As only a few malaria trials provided retrospective estimates of k/ICC, we were unable to investigate whether basing estimates on prior data assists in accurately characterising cluster heterogeneity. Although, we did reveal assumed ICC/K estimates were lower, on average, than calculated ICC/k values in sample size calculations inferring trialists underestimated values. However, this finding could be biased by trialists being more prompted to empirically calculate cluster heterogeneity if they anticipated elevated values. Consequently, characterising the true degree of cluster heterogeneity among a representative sample of malaria CRTs to inform future trials remains an imperative area of continued investigation.

Conclusion

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

Abbreviations
CRT   Cluster randomised trial
ES   Effect size
ICC   Intracluster correlation coefficient
IRS   Indoor residual spraying
ITN   Insecticide-treated bed nets
k   Coefficient of variation
LLIN   Long-lasting insecticidal nets
pa   Per annum
PCR   Polymerase chain reaction
py   Person-year
RDT   Rapid diagnostic test
WHO   World Health Organisation

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12874-024-02361-9.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Medical Research Council (MRC) International Statistics and Epidemiology Group, Department of Infectious Disease Epidemiology and International Health, London School of Hygiene and Tropical Medicine, London, UK
[2]Medical Research Council (MRC) Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, London, UK

References
1. WHO. World malaria report 2023. World Health Organisation: Geneva; 2023.
2. WHO. How to design vector control efficacy trials: guidance on phase III vector control field trial design. World Health Organization; 2017.
3. Kroeger A, et al. Insecticide-impregnated bed nets for malaria control: varying experiences from Ecuador, Colombia, and Peru concerning acceptability and effectiveness. Am J Trop Med Hyg. 1995;53(4):313–23.
4. Nevill CG, et al. Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast. Tropical Med Int Health. 2007;1(2):139–46.
5. ter Kuile FO, et al. Impact of permethrin-treated bed nets on malaria, anemia, and growth in infants in an area of intense perennial malaria transmission in western Kenya. Am J Trop Med Hyg. 2003;68(4 Suppl):68–77.
6. Tagbor H, et al. The clinical impact of combining intermittent preventive treatment with home management of malaria in children aged below 5 years: cluster randomised trial. Tropical Med Int Health. 2011;16(3):280–9.
7. Protopopoff N, et al. Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two factorial design trial. Lancet. 2018;391(10130):1577–88.
8. Tiono AB, et al. Efficacy of Olyset Duo, a bednet containing pyriproxyfen and permethrin, versus a permethrin-only net against clinical malaria in an

area with highly pyrethroid-resistant vectors in rural Burkina Faso: a cluster-randomised controlled trial. Lancet. 2018;392(10147):569–80.

9. Asante KP, et al. Feasibility, safety, and impact of the RTS,S/AS01(E) malaria vaccine when implemented through national immunisation programmes: evaluation of cluster-randomised introduction of the vaccine in Ghana, Kenya, and Malawi. Lancet. 2024;403(10437):1660–70.

10. Delrieu I, et al. Design of a phase III cluster randomized trial to assess the efficacy and safety of a malaria transmission blocking vaccine. Vaccine. 2015;33(13):1518–26.

11. Dron L, et al. The role and challenges of cluster randomised trials for global health. Lancet Glob Health. 2021;9(5):e701–10.

12. Homan T, et al. The effect of mass mosquito trapping on malaria transmission and disease burden (SolarMal): a stepped-wedge cluster-randomised trial. Lancet. 2016;388(10050):1193–201.

13. Samuels AM, et al. Impact of community-based Mass Testing and Treatment on Malaria infection prevalence in a high-transmission area of Western Kenya: a Cluster Randomized Controlled Trial. Clin Infect Dis. 2021;72(11):1927–35.

14. Sangoro O, et al. A cluster-randomized controlled trial to assess the effectiveness of using 15% DEET topical repellent with long-lasting insecticidal nets (LLINs) compared to a placebo lotion on malaria transmission. Malar J. 2014;13:324.

15. Shekalaghe SA, et al. A cluster-randomized trial of mass drug administration with a gametocytocidal drug combination to interrupt malaria transmission in a low endemic area in Tanzania. Malar J. 2011;10:247.

16. Sochantha T, et al. Insecticide-treated bednets for the prevention of Plasmodium Falciparum malaria in Cambodia: a cluster-randomized trial. Trop Med Int Health. 2006;11(8):1166–77.

17. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28(2):319–26.

18. Hemming K, Forbes SEG, Weijer C, Taljaard M. How to design efficient cluster randomised trials. Res Methods Report, 2017. 358.

19. Murray DM, et al. Design and analysis of group-randomized trials in cancer: a review of current practices. J Natl Cancer Inst. 2008;100(7):483–91.

20. Parker K, et al. Characteristics and practices of school-based cluster randomised controlled trials for improving health outcomes in pupils in the UK: a systematic review protocol. BMJ Open. 2021;11(2):e044143.

21. Froud R, et al. Quality of cluster randomized controlled trials in oral health: a systematic review of reports published between 2005 and 2009. Community Dent Oral Epidemiol. 2012;40(Suppl 1):3–14.

22. Diaz-Ordaz K, et al. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. BMC Med Res Methodol. 2013;13:127.

23. Rutterford C, et al. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. J Clin Epidemiol. 2015;68(6):716–23.

24. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol. 2006;35(5):1292–300.

25. Baidjoe AY, et al. Factors associated with high heterogeneity of malaria at fine spatial scale in the western Kenyan highlands. Malar J. 2016;15:307.

26. Gwitira I, et al. Spatial and spatio-temporal analysis of malaria cases in Zimbabwe. Infect Dis Poverty. 2020;9(1):146.

27. Knudson A, et al. Spatio-temporal dynamics of Plasmodium Falciparum transmission within a spatial unit on the Colombian Pacific Coast. Sci Rep. 2020;10(1):3756.

28. Selvaraj P, Wenger EA, Gerardin J. Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: a modeling study. BMC Infect Dis. 2018;18(1):413.

29. Tiono AB, et al. A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of Plasmodium Falciparum in Burkina Faso. Malar J. 2013;12:79.

30. Sutanto I, et al. Negligible impact of Mass Screening and Treatment on Mesoendemic Malaria Transmission at West Timor in Eastern Indonesia: a cluster-randomized trial. Clin Infect Dis. 2018;67(9):1364–72.

31. Tagbor H, et al. The clinical impact of combining intermittent preventive treatment with home management of malaria in children aged below 5 years: cluster randomised trial. Trop Med Int Health. 2011;16(3):280–9.

32. Desai MR, et al. Impact of intermittent Mass Testing and Treatment on incidence of Malaria infection in a high transmission area of Western Kenya. Am J Trop Med Hyg. 2020;103(1):369–77.

33. Bousema T, et al. The impact of Hotspot-targeted interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: a cluster-randomized controlled trial. PLoS Med. 2016;13(4):e1001993.

34. Magris M, et al. Community-randomized trial of lambdacyhalothrin-treated hammock nets for malaria control in Yanomami communities in the Amazon region of Venezuela. Trop Med Int Health. 2007;12(3):392–403.

35. Keating J, et al. Evaluating indoor residual spray for reducing malaria infection prevalence in Eritrea: results from a community randomized control trial. Acta Trop. 2011;119(2–3):107–13.

36. Lorenz LM, et al. Comparative functional survival and equivalent annual cost of 3 long-lasting insecticidal net (LLIN) products in Tanzania: a randomised trial with 3-year follow up. PLoS Med. 2020;17(9):e1003248.

37. Obi E, et al. Monitoring the physical and insecticidal durability of the long-lasting insecticidal net DawaPlus((R)) 2.0 in three States in Nigeria. Malar J. 2020;19(1):124.

38. Villalta EL, et al. Evaluation of the durability and use of long-lasting insecticidal nets in Nicaragua. Malar J. 2021;20(1):106.

39. Misra SP, et al. Malaria control: bednets or spraying? Spray versus treated nets using deltamethrin–a community randomized trial in India. Trans R Soc Trop Med Hyg. 1999;93(5):456–7.

40. von Seidlein L, et al. The effect of mass administration of sulfadoxine-pyrimethamine combined with artesunate on malaria incidence: a double-blind, community-randomized, placebo-controlled trial in the Gambia. Trans R Soc Trop Med Hyg. 2003;97(2):217–25.

41. Bradley J, et al. A cluster randomized trial comparing deltamethrin and bendiocarb as insecticides for indoor residual spraying to control malaria on Bioko Island, Equatorial Guinea. Malar J. 2016;15(1):378.

42. Poespoprodjo JR, et al. Supervised versus unsupervised primaquine radical cure for the treatment of falciparum and vivax malaria in Papua, Indonesia: a cluster-randomised, controlled, open-label superiority trial. Lancet Infect Dis. 2022;22(3):367–76.

43. Chaccour C, et al. Incremental impact on malaria incidence following indoor residual spraying in a highly endemic area with high standard ITN access in Mozambique: results from a cluster-randomized study. Malar J. 2021;20(1):84.

44. Foy BD, et al. Efficacy and risk of harms of repeat ivermectin mass drug administrations for control of malaria (RIMDAMAL): a cluster-randomised trial. Lancet. 2019;393(10180):1517–26.

45. WHO. *Data requirements and protocol for determining non-inferiority of insecticide-treated net and indoor residual spraying products within an established WHO intervention class.* 2019.

46. Deutsch-Feldman M et al. Spatial and epidemiological drivers of Plasmodium Falciparum malaria among adults in the Democratic Republic of the Congo. BMJ Glob Health, 2020. 5(6).

47. Yamba EI, et al. Climate drivers of Malaria Transmission Seasonality and their relative importance in Sub-saharan Africa. Geohealth. 2023;7(2):e2022GH000698.

48. Bath D, et al. Effectiveness and cost-effectiveness of reactive, targeted indoor residual spraying for malaria control in low-transmission settings: a cluster-randomised, non-inferiority trial in South Africa. Lancet. 2021;397(10276):816–27.

49. Merkord CL, et al. Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. Malar J. 2017;16(1):89.

50. Pourtois JD, et al. Climatic, land-use and socio-economic factors can predict malaria dynamics at fine spatial scales relevant to local health actors: evidence from rural Madagascar. PLOS Glob Public Health. 2023;3(2):e0001607.

51. Tompkins AM, et al. Dynamical Malaria forecasts are Skillful at Regional and local scales in Uganda up to 4 months ahead. Geohealth. 2019;3(3):58–66.

52. Zinszer K et al. A scoping review of malaria forecasting: past work and future directions. BMJ Open, 2012. 2(6).

53. Multerer L, et al. Analysis of contamination in cluster randomized trials of malaria interventions. Trials. 2021;22(1):613.

54. Hemming K, et al. Ethical implications of excessive cluster sizes in cluster randomised trials. BMJ Qual Saf. 2018;27(8):664–70.

55. Ouyang Y, et al. Accounting for complex intracluster correlations in longitudinal cluster randomized trials: a case study in malaria vector control. BMC Med Res Methodol. 2023;23(1):64.

56. Pinder M, et al. Efficacy of indoor residual spraying with dichlorodiphenyltrichloroethane against malaria in Gambian communities with high usage of long-lasting insecticidal mosquito nets: a cluster-randomised controlled trial. Lancet. 2015;385(9976):1436–46.

Publisher's note