HOW TO DESIGN VECTOR CONTROL EFFICACY TRIALS

Guidance on phase III vector control field trial design
provided by the Vector Control Advisory Group

World Health Organization

Acknowledgements

The main text was prepared by Anne Wilson and Steve Lindsay (Durham University, England), with major contributions from Immo Kleinschmidt (London School of Hygiene and Tropical Medicine, England), Tom Smith (Swiss Tropical and Public Health Institute, Switzerland) and Ryan Wiegand (Centers for Disease Control and Prevention, United States of America). The document is based on a peer-reviewed manuscript (*14*). The authors would like to thank the following experts, including those who attended the Expert Advisory Group on Design of Epidemiological Trials for Vector Control Products, held 24–25 April 2017, and the Vector Control Advisory Group for their contributions to the guidance document: Salim Abdulla, Till Baernighausen, Christian Bottomley, John Bradley, Vincent Corbel, Heather Ferguson, Paul Garner, Azra Ghani, David Jeffries, Sarah Moore, Robert Reiner, Molly Robertson, Mark Rowland, Tom Scott, Joao Bosco Siqueira Jr., and Hassan Vatandoost. We would also like to acknowledge the following individuals for significant contributions to reviewing and editing this work: Anna Drexler, Tessa Knox, Jan Kolaczinski, Emmanuel Temu and Raman Velayudhan.

WHO/HTM/NTD/VEM/2017.03

Design and layout: Patrick Tissot, WHO Neglected Tropical Diseases.

Please consult the WHO Neglected Tropical Diseases website for the most up-to-date version of all documents (www.who.int/neglected_diseases/en)

Printed in France.

Contents

Preface v
Glossary vi
Abbreviations xv
Executive summary xvi

1. Why is the design of vector control studies so important? 1
2. Framework for designing and conducting an efficacy study 4
   - 2.1 Step 1: define the research question 4
     - 2.1.1 Superiority and non-inferiority studies 6
     - 2.1.2 Choosing outcome measures 7
     - 2.1.3 Choosing a comparator 8
   - 2.2 Step 2: decide on the experimental unit 9
     - 2.2.1 Replication 11
   - 2.3 Step 3: decide on the study's design 12
     - 2.3.1 Bias and confounding 12
     - 2.3.2 Types of study designs 14
     - 2.3.3 Randomized controlled trials 15
     - 2.3.4 Controlled before-and-after studies, time series and interrupted time series studies 17
     - 2.3.5 Crossover studies 18
     - 2.3.6 Observational studies 19
     - 2.3.7 Hierarchy of study designs for estimating intervention efficacy 22
   - 2.4 Step 4: calculate the sample size 25
   - 2.5 Step 5: evaluate contamination effects 28
   - 2.6 Step 6: blinding 31
   - 2.7 Step 7: implement the intervention 32
   - 2.8 Step 8: determine how to measure the outcome 33
     - 2.8.1 Choosing sources of epidemiological data 33
     - 2.8.2 Choosing epidemiological outcomes 34
     - 2.8.3 Choosing entomological outcomes 34
     - 2.8.4 Selecting sites for entomological monitoring 34
     - 2.8.5 Duration of follow-up 35

2.8.6 Loss to follow-up 36
2.9 Step 9: write the protocol and register the study 36
2.10 Step 10: ethical review and study oversight 38
2.10.1 Ethical review 38
2.10.2 Study oversight 38
2.11 Step 11: randomization and study conduct 38
2.11.1 The two-step randomization process 39
2.11.2 Randomization for cluster-randomized trials 39
2.11.3 Data capture 40
2.12 Step 12: data management and statistical analysis 40
2.12.1 Analysis plan 41
2.12.2 Data management 42
2.12.3 Statistical analysis: outcome measures and their relation to techniques for statistical analysis 42
2.13 Step 13: reporting and disseminating the study 43
2.13.1 Study reporting 43
2.13.2 Research dissemination 43
3. Conclusions 45

References 46
Annexes 54
Annex 1. Suggested further reading 54
Annex 2. Summary of the GRADE approach to rating the quality of evidence 56
Annex 3. Items to include in a sample size calculation 57
Annex 4. Items that should be included in a clinical trial protocol and related documents 59

Preface

Vector-borne diseases (VBDs), including malaria, dengue and leishmaniasis, cause considerable morbidity and mortality in tropical regions of the world. Vector control methods, such as long-lasting insecticidal nets (LLINs), indoor residual spraying (IRS) and larval source management, are typically the most effective interventions against VBDs. For some diseases, such as dengue, chikungunya and Zika virus disease, vector control is the only method available to protect populations.

Historically, vector control has been important to control and eliminate many VBDs, including malaria, onchocerciasis, leishmaniasis, Chagas disease and Aedes-borne diseases (*1–9*). Widespread deployment of LLINs and, to a lesser extent, IRS have also contributed to an unprecedented decline in malaria infection of more than 50% in sub-Saharan Africa (*10*). Although these programmes have been successful at reducing morbidity and mortality from VBDs, vector control programmes face a number of challenges, including financial and human resource constraints, widespread insecticide resistance, the spread of pathogens and vectors, and uncontrolled urbanization.

Maintaining the gains achieved by using vector control measures urgently requires new tools, which must be shown to be efficacious before they can be broadly used for public health. For many vector control interventions against malaria or neglected tropical diseases (known as NTDs), such as dengue, leishmanaisis and Chagas disease, there is a dearth of evidence to support their use in public health approaches to vector control because few epidemiological field trials have been conducted, and those that have been were often poorly designed and implemented. To develop new vector control tools, epidemiological (phase III) efficacy trials are critically important to measure the efficacy of such interventions against epidemiological end points. These trials provide the main body of evidence reviewed by policymakers and help drive public health decision-making because an intervention that does not reduce infection, morbidity or mortality will not be recommended.

The World Health Organization (WHO) leads policy-making on new vector control tools. For more than 20 years, WHO's evidence-based policy-making has leveraged evidence from field efficacy studies to make recommendations about vector control that are adopted at the country level. For example, WHO's review of evidence from large-scale epidemiological field trials on LLINs and IRS led to policy recommendations and in-country use (*11–13*) that have contributed to significant declines in malaria (*10*). For new vector control tools, technologies and approaches, WHO's Vector Control Advisory Group (VCAG) plays a key part in policy development. Initiated in 2012, this advisory body assesses the potential public health efficacy of new products and assists WHO in formulating new policy recommendations for vector control. To facilitate evidence-based policy-making, VCAG also advises product developers about how to best generate efficacy data for VCAG review.

The purpose of this document is to provide guidance about designing and conducting phase III epidemiological field trials of new vector control interventions. In this document, phase III field trials refers to epidemiological trials used to measure the efficacy of vector control interventions against epidemiological

end points, and they are distinct from WHO Pesticide Evaluation Scheme (known as WHOPES) phase III studies, which are large-scale field studies that measure the longevity of insecticidal activity and, in the case of LLINs, the physical durability and acceptance of the insecticidal nets. Many of the principles that apply to designing phase III studies also apply to phase IV studies, which measure the effectiveness of vector control interventions under routine conditions.

This guidance does not replace training, consultation with statisticians or epidemiologists, or the extensive range of books covering this subject (many of which are suggested as further reading, see **Annex 1**). Instead, it is intended to give indications from which investigators can delve further into the literature or consult with experts, as appropriate. Epidemiological field trials are significant and complex undertakings and, therefore, it is essential that entomologists and epidemiologists are involved in planning and conducting these studies.

This manual sets out a framework of steps to take and concerns to be considered when designing and conducting an epidemiological trial to assess the public health value of a new vector control intervention; these steps include defining the research question, choosing the study's design, randomizing the interventions, and calculating the sample size. The guidance has been formulated to aid innovators and researchers, both from academic institutions and country programmes. It is hoped that this guidance will lead to more carefully considered and rigorously designed vector control studies, so that the results from these studies can be used to recommend new interventions for vector control deployed at the country level.

VCAG is available to provide technical guidance to product developers and manufacturers about how to design and conduct epidemiological trials for new vector control products that are in keeping with the guidance provided in this manual.

Glossary

(adapted from references *15–20*)

Active case detection	This refers to cases detected by study staff who visit participants in their homes and screen them for disease.
Allocation concealment	This refers to ensuring that the investigator is unaware of which group (that is, treatment or control) an individual or cluster will be assigned to on enrolment. Selection bias can be introduced if the investigator or participant knows to which group a participant (or cluster) is assigned. For example, this may occur if alternation is used to allocate participants to treatment or control groups, or if assignment envelopes are not sealed and can be tampered with. Allocation concealment is more crucial in studies using sequential randomization rather than those in which randomization occurs at a single time point.
Attrition bias	This refers to systematic differences occurring between those individuals or communities that drop out of a study versus those that continue.
Bias	Bias is an overestimate or underestimate of the effect of an intervention.
Blinding	Blinding is used in trials in which participants or investigators or outcome assessors, or a combination of these, do not know to which group the individual or cluster has been assigned. Single-blind refers to either the participant or investigator (or outcome assessor) being blinded to the assignment, and double-blind refers to both the participant and the investigator (or outcome assessor) being blinded to the assignment.
Case–control study	This type of study compares the prevalence of an exposure (for example, the use of a protective intervention) between a group of people with the disease of interest (cases) and a group of people without the disease (controls). In a study of this type, the controls should be selected so that they represent the population from which the cases originated.
Cluster-randomized controlled trial (CRT)	This is a randomized controlled trial in which groups of individuals, for example, a household, village or larger geographical area are Randomly allocated to receive either intervention or control.
Cohort study (observational)	This is a type of observational study in which groups of disease-free individuals are identified: those in the exposed group (using the protective intervention) and those who are unexposed (not using the protective intervention). The groups are then followed over a period of time to evaluate the outcome of interest (usually disease or infection). In this study type, individuals are not allocated to the intervention of interest by the investigators.

Cohort study (randomized)

A randomized controlled trial in which a cohort of individuals is randomized to receive either the intervention or control intervention, and the cohorts are followed up for the outcome of interest.

Confounding bias

This type of bias arises whenever a cause is common to both the outcome and the exposure (*20*). For example, in an observational study of the association between using house screening and the incidence of malaria, the relationship is likely to be confounded by socioeconomic status because people in houses that use screening are likely to have higher socioeconomic status, and they may have greater access to other protective measures against malaria, such as long-lasting insecticidal nets. This is a common type of bias in observational studies and non-randomized trials, but it can also occur in poorly randomized studies. A variable that is on the causal pathway between the exposure and the outcome is not a confounder. For example, if indoor residual spraying reduces mosquito density, which results in lower malaria incidence, then mosquito density does not confound the relationship between exposure and outcome, even though it is associated with both.

Control group

This is the group of participants that receives no intervention, a placebo or the current standard of care (depending on the study design), and this group thereby serves as a comparison group when the intervention results are evaluated.

Courtesy bias

This refers to the tendency for participants to give favourable answers out of courtesy to the investigator, for example, by incorrectly reporting high compliance with an intervention.

Cross-sectional study

In an analytical cross-sectional study, information is collected at one point in time on the prevalence of the outcome of interest (for example, a disease or infection) and exposure (for example, the use of a protective intervention).

Cluster randomization

In this type of study, clusters are randomly assigned to either control or intervention groups. Clusters can be geographical areas (for example, sectors of a large city), communities (for example, villages), administrative units (for example, districts or regions), institutions (for example, schools), health facilities or households.

Confounding variables

The variables that are associated with both the exposure and the outcome but that do not lie on the causal pathway between exposure and outcome.

Consistency

Consistency refers to the level of heterogeneity in a study's results that remains after exploration of a priori hypotheses that might explain heterogeneity. The GRADE system (see below) suggests rating downgrading the quality of the evidence if a high level of inconsistency (heterogeneity) is present.

Controlled before-and-after study	A study in which observations are made before and after the implementation of an intervention in both the intervention group and a control group that does not receive the intervention; this is also known as a pre–post study.
Crossover study	Study in which individuals or clusters are allocated to the intervention or control group for a period of time before switching (or crossing over) to the other group. There is usually a washout period before the switch is made to avoid carry-over effects from the intervention.
Detection bias	This type of bias refers to systematic differences between the intervention and comparator groups in how a study's outcomes are determined. It can be reduced by blinding the clinician (or outcome assessor) to the group to which each participant is assigned.
Direct effect	In a cluster-randomized controlled trial, the direct effect of the intervention can be estimated within the intervention clusters by comparing the incidence of the outcome in those receiving the intervention with that in those not receiving it.
Directness	This refers to the generalizability of the population, intervention, comparator and outcomes from each study to the population of interest. For example, evidence from a study conducted in soldiers may be downgraded when applying this evidence to the general population. The GRADE approach suggests downgrading the quality of evidence if the study population, intervention, comparator or outcomes differ from those in the population of interest.
Dose response	This term refers to the relationship between the exposure and the outcome. There is a dose response when there is a decreasing or increasing trend in the outcome in relation to increasing levels of exposure.
Double blind	In this type of blinding, neither the investigator nor the participant knows to which treatment group a participant has been randomly assigned.
Effectiveness study	These studies estimate the effect of an intervention under pragmatic (or real-life) conditions (for example, interventions delivered under routine conditions) so that the relevance of the findings for policy and practice is maximized.
Effect size	This refers to the magnitude of difference between the treatment and control groups, expressed, for example, as the risk, rate ratio or percentage reduction in prevalence.
Efficacy trial	These studies estimate the effect of an intervention under the ideal conditions that can usually be achieved only in a trial, for example, by ensuring maximal coverage of the target population and adherence to the intervention.
Eligibility criteria	These are requirements that must be fulfilled for individuals to be able to participate in a study; for example, this could include restricting participation to people of a specific age.

Exclusion criteria	These are characteristics that disqualify an individual from participating in a study; for example, people who have participated in another clinical trial may be excluded, or those who anticipate travelling out of the study area during the study period.
Experimental study	In this type of study design, the interventions are allocated to study participants and the outcomes are observed.
Experimental unit	This refers to a participant or group of participants exposed to a particular treatment.
Good Clinical Practice	Good Clinical Practice (GCP) is a set of internationally recognized ethical and scientific quality requirements developed by the International Council for Harmonisation (known as the ICH) of Technical Requirements for Pharmaceuticals for Human Use that must be followed when designing, conducting, recording and reporting clinical trials that involve people.
GRADE	The Grading of Recommendations Assessment, Development and Evaluation method is a systematic and explicit approach to making judgements about the quality of a body of evidence and the strength of recommendations made from that evidence.
Indirect effect	In a cluster-randomized controlled trial, the indirect effect of the intervention can be estimated by comparing the incidence of the outcome among those in the intervention cluster that did not receive the intervention with that in the control clusters.
Information bias	Information bias occurs when inaccurate information is collected about the exposure (intervention) or the disease, or both, which can lead to a biased estimate of the association between the exposure (or intervention) and the outcome. Important types of information bias are recall bias and observer bias.
Intention-to-treat analysis	In this type of evaluation, data from participants are analysed according to the randomization scheme, and any changes that occurred following randomization – such as non-compliance, protocol deviations and withdrawal – are ignored.
Interrupted time series	This is a type of study in which the outcome (for example, disease incidence) is measured on a number of occasions, both before and following the introduction of an intervention. This allows an investigator to determine whether an intervention has had an impact greater than any underlying trend in the data. This design may include a parallel control group.
Loss to follow-up	This refers to participants who are not able to be followed up during the study, generally because they are cannot be reached. High levels of loss to follow-up can introduce bias if the loss differs between study arms.
Margin of non-inferiority	This is pertinent to non-inferiority trials: the margin of non-inferiority (delta; δ) is an arbitrarily defined margin within which the difference between the efficacy of the tested product and the comparator must lie.

Mass effect	These are the additional effects of an intervention that are achieved when a substantial proportion of the population receives the intervention. For example, in some circumstances, mass killing of mosquitoes coming into contact with long-lasting insecticidal nets can reduce the transmission of a disease so that indirect protection is provided to those individuals not using the nets.
Matching	In this technique, clusters are formed into groups such that only one cluster in each group is assigned to each study arm. Matching is typically done using the baseline values of the end point of interest or a surrogate variable that is expected to be correlated with the end point.
Non-inferiority study	A non-inferiority trial aims to demonstrate that the tested product is not worse than the comparator by more than a small, pre-specified amount, which is known as the non-inferiority margin (delta; δ). The difference between the effect of the test product (T) and the effect of the comparator (C) must be less than δ – that is, the upper bound of the 95% confidence Interval of C – T must be less than δ. The choice of δ is a clinical (or entomological) judgement, not a statistical one. The smaller the δ, the less T is inferior to C, but the larger the required sample size.
Observational study	A type of study in which the effect of the exposure on the participants is observed, but the investigator has no role in assigning participants to the exposure.
Observational unit	The unit in which a particular outcome is measured, for example, dengue infection may be measured in an individual, or mosquito density may be measured in a community.
Observer bias	Observer bias occurs when knowledge of exposure status influences the classification of disease status or vice versa. For example, clinicians assessing patients may be more or less likely to diagnose a particular disease if they know that a person received a protective intervention in the study. Observer bias can be reduced by ensuring that investigators and outcome assessors are not aware of to which group (intervention or control) participants have been assigned, or by having outcome measures that do not rely on subjective judgements, or both.
Odds ratio	This is the ratio of the odds of a disease occurring in the exposed (intervention) group compared with the unexposed (control) group.
Outcome	This refers to a parameter that the study sets out to measure; it should be defined in advance of the study being conducted and reflect the question the study sets out to answer. The primary outcome is the outcome of greatest importance, and a secondary outcome is typically an additional effect of an intervention that is of lesser importance or is less optimal for assessing the question asked by the study. There can be multiple primary and secondary outcomes. Primary outcomes can be epidemiological or both epidemiological and entomological.
Overall effect	The overall effect of the intervention is obtained by comparing the overall incidence in the intervention and control arms.

Passive case detection	This is when a case is detected because a participant attends a health provider.
Performance bias	This refers to "systematic differences in the care provided to members of the different study groups other than the intervention under investigation" (*18*). For example, if participants know they are in the control group of a trial of insect repellent, they may be more likely to use other forms of vector control, such as protective clothing. Alternatively, healthcare providers may care for patients differently if they are aware of the study group to which the participant is assigned. Performance bias can be reduced by using blinding to ensure that participants, healthcare providers and researchers are not aware of which intervention participants have received, although this is not always possible.
Per-protocol	This term refers to the study population who completed the treatment as originally allocated. For example, in a per-protocol analysis of a study of long-lasting insecticidal nets, individuals who travelled out of the area during the study would be excluded.
Placebo	This is an intervention given to participants in the control group that appears similar to the real intervention but, in fact, has no therapeutic benefit. Placebo interventions can be used to blind participants and reduce performance bias.
Precision	This refers to the level of confidence in an effect estimate. The confidence intervals around the effect estimate indicate the precision of the estimate. Random error can lead to large confidence intervals if the sample size is small or the number of events (or cases) is small. GRADE recommends downgrading for imprecision if (i) a recommendation would differ if the upper or lower boundary of a 95% confidence interval represented the truth or (ii) the effects are large and both the sample size and number of events are modest even if the 95% confidence intervals appear satisfactorily narrow.
Prevalence ratio	This is the prevalence of disease (or infection) among those who received the intervention compared with the prevalence of disease (or infection) among those who did not receive the intervention.
Protective efficacy (PE)	This is the percentage reduction in morbidity or mortality in those receiving the vector control intervention compared with that in the control group. It can be calculated as (i) 1 – the rate ratio, (ii) 1 – the risk ratio or (iii) 1 – the odds ratio.
Pseudoreplication	This refers to using inferential statistics to test for treatment effects with data from experiments or observational studies in which either the treatments are not replicated (although samples may be) or the replicates are not statistically independent.
Publication bias	This refers to bias introduced into a systematic review because positive study results are more likely to be published than negative ones.
Public health value	An intervention has public health value if it has proven protective efficacy to reduce or prevent infection or disease, or both, in humans.

Randomization	Individuals or clusters are allocated to intervention and control groups at random. Randomization consists of two interrelated steps: sequence generation and allocation concealment (not to be confused with blinding).
Randomized controlled trial (RCT)	In this study design, individuals are randomly allocated to either the intervention or control group. The intervention and control groups are then followed up for the outcome of interest.
Rate ratio	This is the ratio of the rate of disease (or infection) among those who received the intervention compared with the rate of disease (or infection) among those who did not receive the intervention
Recall bias	This refers to systematic differences between groups in recalling information about exposures. It is a particular problem in case–control studies in which surveys are used to gather information about past exposures.
Risk difference	This is the risk of disease (or infection) in the intervention group minus the risk of disease (or infection) in the control group.
Risk ratio	This is the ratio of the risk of disease (or infection) among those who received the intervention compared with the risk of disease (or infection) among those who did not receive the intervention
Sequence generation	This refers to a method of generating an allocation sequence. The method can be non-random (for example, whether a participant has an odd or even date of birth, or the investigator's preference) or random (for example, using a random number generator, drawing lots, tossing a coin).
Selection bias	This refers to "bias in the estimated association or effect of an exposure on an outcome that arises from the procedures used to select individuals into the study or the analysis" (*18*). Often the term is used to refer to systematic differences among the characteristics of the study population and those of other populations (that is, highlighting a lack of generalizability). In randomized controlled trials and cohort studies, selection bias can occur when there is loss to follow-up. In case–control studies, selection bias is introduced if cases are selected who are not representative of all cases within the population, or if controls are selected who are not representative of the population that produced the cases.
Single blind	This refers to a participant being unaware of which treatment group they have been randomized to, although the investigator is aware.
Step-wedge design	This is a type of study in which the intervention is rolled out to different clusters in a staged fashion. At the end of the study, all clusters will have received the intervention. The order in which clusters receive the intervention is usually determined at random.
Stratification or stratified randomization	This technique is used to ensure that study arms are balanced with regard to a characteristic thought to affect response to the vector control intervention (for example, baseline incidence). Individuals or clusters are grouped to form strata based on a characteristic (for example, clusters with a low incidence versus high incidence of the disease) and are randomly allocated to the intervention or control group within the stratum such that equal numbers are assigned to each group in each stratum.

Study sponsor	Some regulatory bodies or funders require a study's sponsor to be identified. The sponsor is an individual, company, institution or organization that takes ultimate responsibility for the initiation of, management (or arranging the initiation and management) or financing (or arranging the financing), or a combination of these, for the research. The sponsor takes primary responsibility for ensuring that the study's design, conduct and reporting meet appropriate standards.
Superiority study	This type of study aims to show that one vector control intervention is more efficacious than another. This requires a one-sided test of statistical significance.
Systematic review	A systematic review uses rigorous methods to identify studies and synthesize the results of these studies to answer a specific question. The Cochrane Collaboration produces gold standard systematic reviews that are conducted in a highly rigorous fashion.
Target product profile	This is a strategic document that guides the process of developing a new vector control product by outlining the features of and performance targets for the intended product.
Test-negative case–control study	This is a type of case–control design whereby the use of an intervention is compared between cases who test positive and those who test negative (controls) who present to a health facility. The advantage of this design is that cases and controls are recruited in a single step and there is no need to spend time testing individuals to identify controls from the community.
Time series	In this type of study, the outcome (for example, the incidence of disease) is measured on a number of occasions following the introduction of an intervention. Typically, measurements are made at equally spaced time points, for example, monthly or yearly. In some cases, there may also be a control time series of people who have not received the intervention, in which the same measurements are made, although some time series studies do not have a control group.
Total effect	In a cluster-randomized controlled trial, the total effect of the intervention is calculated by comparing the incidence among the in the intervention clusters with the incidence in the control clusters.
Triple blind	This refers to the study participants, investigators, laboratory staff and those analysing data being blinded to which treatment group participants have been randomized.
Trial steering committee	Group of individuals responsible for providing overall supervision of the trial including ensuring that the trial is conducted according to Good Clinical Practice.

Abbreviations

CONSORT	Consolidated Standards of Reporting Trials
CRT	cluster-randomized controlled trial
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IRS	indoor residual spraying
ITT	intention to treat
LLIN	long-lasting insecticidal net
NTD	neglected tropical disease
PE	protective efficacy
PICO	Population, Intervention, Comparator, Outcome
RCT	randomized controlled trial
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
SWCRT	step-wedge cluster-randomized trial
TPP	target product profile
VBD	vector-borne disease
VCAG	Vector Control Advisory Group at WHO
WHO	World Health Organization
WHOPES	WHO Pesticide Evaluation Scheme

Executive summary

Vector control is the main method by which vector-borne diseases (VBDs) can be controlled and eliminated. Member States rely on the World Health Organization (WHO) for guidance about new vector control tools, technologies and approaches, including which new products are safe, effective and of high enough quality to meet the needs of disease control programmes. Two parallel assessment pathways provide guidance for Member States: the prequalification pathway, for products already covered by a policy recommendation, and the new intervention pathway, which supports the development of new policies for products not in an already established class. Through the new intervention pathway, WHO's Vector Control Advisory Group (VCAG) evaluates evidence and validates the public health value of new vector control tools to support WHO in formulating policies. However, WHO's policy-making for new tools relies on well-designed and well-conducted entomological and epidemiological trials. To ensure that studies deliver robust and policy-relevant evidence, this manual provides detailed guidance about how to design phase III efficacy trials for vector control. The recommendations are intended to help develop the best evidence base and facilitate the most expeditious assessment of the public health value of new vector control tools. In addition to using this document, it is recommended that innovators partner with experienced epidemiologists and statisticians when conducting trials and draw on the VCAG for guidance about trial design.

The first step in designing a trial is to define a clear and valid research question that does not duplicate existing efforts. The study population, intervention, comparator and outcomes should be clearly defined. Depending on the question, studies may be designed to show comparative effectiveness, to demonstrate the superiority of the new intervention or to show non-inferiority to an existing intervention. Phase III studies should be designed around epidemiological end points to demonstrate the public health value of the intervention. Entomological outcomes cannot be used on their own for this purpose, although they can be combined with epidemiological outcomes to evaluate a claimed entomological effect.

Typically, researchers will allocate vector control interventions to groups of individuals (termed clusters), which can be households, villages or larger geographical areas, although some study designs will use individual-level allocation. Study designs can be observational, in which the investigator's role is restricted to observing the relationship between the intervention and outcome, or experimental, in which investigators assign treatments to experimental units and observe the outcome.

Trials should be designed to minimize the risk of bias. Randomized controlled trials provide the best means of minimizing bias. If the number of experimental units is sufficiently large, randomization will ensure that there are no systematic differences between study arms caused by possible confounding variables. Studies without a contemporaneous control group are not recommended because there may be secular trends or changes in the study outcome that are not due to the intervention. Sample size should be determined by a power calculation for the primary study outcomes to ensure that the study can answer the study question. The movement of vectors or humans (for example, in dengue studies with day-biting vectors), or changes to the intervention (for example, sharing topical repellent with relatives) make vector control trials prone to

contamination effects and make findings difficult to interpret. Careful study design and careful analysis of the data are needed to counter these issues. When possible, study participants and outcome assessors should be blinded to help reduce performance and detection biases.

For phase III trials, interventions should be optimally implemented, with high coverage of the study population and high adherence by study participants, as appropriate. Without optimal implementation, researchers cannot know if a lack of effect is due to the poor efficacy of the intervention or simply due to poor implementation.

Epidemiological outcomes should include parasitological confirmation when relevant, such as for suspected cases of malaria, and conform to agreed definitions. Common data collection approaches include passive and active case detection, and cross-sectional surveys. Entomological sampling methods should be selected to minimize bias; for example, automated sampling tools may be less prone to bias than specimen collection by fieldworkers. Vector populations can be highly variable over time and space, therefore entomological sampling sites should be selected randomly, and repeat measures should be taken to capture transmission over time. The intensities of VBDs are typically seasonal and often highly variable among years. For this reason, for any new vector control tools in new product classes, WHO requires evidence from at least two well-conducted, randomized controlled trials that use epidemiological outcomes and follow-up over at least two transmission seasons.

Once the overall design of the study has been finalized, the protocol should be written following established WHO guidelines, and a brief analytical plan should be developed. The protocol should be publicly available, and studies on humans should be included in a trials registry. Studies involving humans or that raise ethical issues need to be reviewed by an ethics committee, and written informed consent must be sought from participants prior to enrolment. Strong oversight for the study should be provided through a trial steering committee, a board to monitor data collection and analysis, and, for studies involving humans, a safety-monitoring board.

For several types of studies, the first step is to randomize individuals or groups of individuals to treatment and control arms. Epidemiological data are likely to include both responses to questionnaires and the collection and testing of clinical samples. A plan for statistical analysis is required before the end of field data collection. The measure of the effect of the intervention is typically expressed as a ratio (for example, the incidence rate ratio or risk ratio) comparing the incidence or prevalence of the outcome in the intervention and control groups. Cluster-randomized trials and repeated measures require sophisticated analytical methods due to the correlation of outcomes within clusters or observational units.

Regular updates on the study's progress should be made to VCAG, and final data shared to enable VCAG to assess a product's public health value. Established guidelines, such as the CONSORT (Consolidated Standards of Reporting Trials) Statement, may be followed for reporting. Although such trials are primarily designed to aid the policy-making process, ultimately researchers should disseminate the results to appropriate national and international audiences, for example, through community meetings at the study site, peer-reviewed publications and conferences.

New vector control tools are urgently needed to reduce the burden of VBDs and mitigate threats to existing methods, such as resistance to insecticides. Improving the quality of future vector control trials by adopting these guidelines will save valuable resources and expedite the global policy-making process, making new and effective interventions accessible for broad, programmatic public health use.

1. Why is the design of vector control studies so important?

Vector-borne diseases (VBDs), such as malaria, dengue and leishmaniasis, are responsible for considerable morbidity and mortality, and fall disproportionately on the poorest communities in low- and middle-income countries (*21–24*). The main method for controlling and eliminating VBDs is through vector control (*3,8, 25–28*). For example, the widespread deployment of long-lasting insecticidal nets (LLINs) and, to a lesser extent, indoor residual spraying (IRS) have been instrumental in the declines in malaria observed during the past 15 years (*10*).

Although vector control tools have a major role in controlling and eliminating VBDs, in many cases the evidence for assessing the efficacy of vector control interventions is limited or non-existent, particularly for VBDs other than malaria, some of which are neglected tropical diseases (NTDs). This is because there have been few field trials, and those that have been performed were often poorly designed and conducted. Poorly conducted studies are unacceptable for several reasons. First, they are difficult to interpret and may give misleading results. Robust designs are needed to provide strong evidence of causality. Evidence-based policy-making has been adopted by the World Health Organization (WHO), and researchers are working with WHO to help establish appropriate ways of summarizing and assessing the evidence about vector control measures (*29,30*). Most guideline review panels, including those at WHO, assess the certainty of the body of evidence derived from systematic reviews using the GRADE approach (Grading of Recommendations Assessment, Development and Evaluation) (**Box 1** and **Annex 2**). These assessments are used by the panels to make policy recommendations. Poorly designed and conducted studies reduce the certainty of the evidence, and the guideline panel takes this into account.

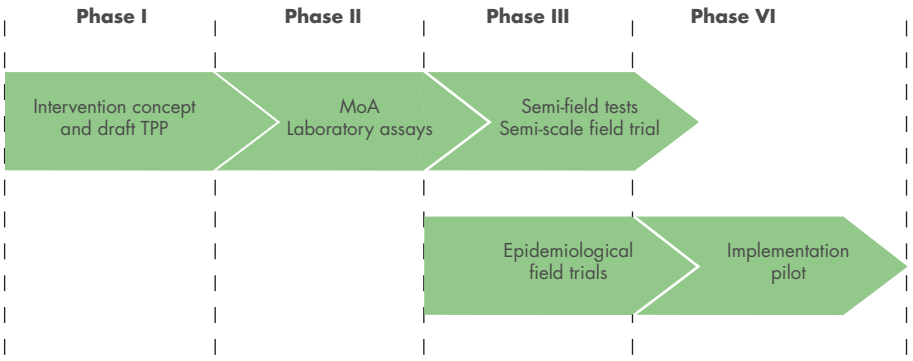
Box 1. GRADE methodology for evaluating evidence when making recommendations for policies and guidelines

Since 2008, WHO has adopted the Grading of Recommendations Assessment, Development and Evaluation system (GRADE) methodology when evaluating evidence and making recommendations for policies and guidelines (*31–33*). The GRADE system gives an initial rating of the certainty of the evidence based on the risk of bias in the study's design (Annex 2), with randomized controlled trials (RCTs) rated at the start as having high certainty, and observational studies starting as lower certainty. Studies are then upgraded or downgraded based on several factors. RCTs can be downgraded depending on the risk of bias, inconsistency, indirectness, imprecision or publication bias. Studies other than RCTs can be upgraded if there are clear, large, obvious effects, a dose response relationship, or limited residual confounding. The final score can range from high (a lot of confidence in the estimate of the effect that is unlikely to change with further research) to very low (that is, there is uncertainty about the estimate of the effect).

Second, poorly conducted studies limit the return on investment of research funding, as highlighted in a Lancet series that called for the better design, conduct, analysis and reporting of studies (*34, 35*). Wasting valuable time testing products while using inadequate study designs means that there are delays in getting potentially life-saving products to market. Last, it is unethical to conduct studies on humans if these studies are unlikely to yield reliable evidence due to poor design and execution.

The development of vector control interventions starts with the development of a target product profile (TPP) and ends with the deployment of an effective tool for public health benefit (Fig. 1). The TPP is a strategic document that guides the development process by outlining the features of and performance targets for the intended vector control tool. After the TPP is developed, phase I studies are conducted to determine the

Fig. 1. Stages in developing a new vector control product (adapted from reference *36*).



TPP, target product profile; MoA, mode of action

mode of action of the product; these are followed by phase II semi-field and small-scale field studies, such as experimental hut studies, which generally have entomological end points. Following phase II studies, phase III trials are conducted using epidemiological outcomes to measure the efficacy of the vector control tool when implemented under optimal conditions.[1] Phase III epidemiological field trials provide the main evidence that is reviewed by policymakers, and these trials are critical in driving public health policy-making because interventions without public health value will not be recommended. In this report, the public health value of a product is defined as a product having proven protective efficacy (PE) to prevent infection or disease, or both, in humans. Interventions that have no benefit can cause serious harm.

The last stage of developing vector control interventions is the phase IV studies that measure the effectiveness of the vector control tool when it is delivered and used operationally (that is, under real-world conditions), as well as collecting information on the feasibility, distribution mechanisms, acceptability, economics (e.g. cost effectiveness analysis) and safety of the tool.

The principles of good study design and conduct outlined in this document can be applied to both phase III and phase IV trials. This document outlines some general concepts relevant to designing studies for phase III epidemiological field trials that aim to measure the efficacy of vector control interventions, as well as discussing bias and the different options for study designs. It then sets out a framework of steps to take and concerns to be considered when designing and conducting a study; these steps include defining the research question, randomly assigning the experimental conditions, and calculating the sample size. A glossary outlines some key definitions, and **Annex 1** provides sources for further reading. It is hoped that this guidance document will lead to more carefully considered and rigorously designed vector control trials. Innovators are strongly advised to work closely with statisticians and epidemiologists to conduct phase III trials and engage with WHO's Vector Control Advisory Group (VCAG) early on during protocol development to ensure that trial data meet WHO's needs for determining public health value.

1. The phase III epidemiological field trials described in this document are distinct from the WHO Pesticide Evaluation Scheme (known as WHOPES) phase III trials that form part of the public health testing and evaluation programme for pesticides.

2. Framework for designing and conducting an efficacy study

Table 1 shows the steps required to design and conduct an efficacy study for vector control interventions.

2.1 Step 1: define the research question

Prior to embarking on a study, researchers should review the published literature on the subject, including systematic reviews, and review trial registries (for example, at https://www.clinicaltrials.gov or http://www.isrctn.com/) to ensure that the study question is valid and does not duplicate existing efforts. Researchers and product developers are encouraged to submit a concept note to VCAG before writing a full study proposal. The research question should be clearly defined. The Population, Intervention, Comparator and Outcome (PICO) process can be used to frame and formulate a research question (Table 2). The question should state clearly the population of interest, the intervention, the comparator and the outcome of interest. The research question should ideally grow out of issues raised either in the literature or in practice, and it should be manageable within the given time frame and resources.

When conducting a study, there are two hypotheses that need to be considered: the null hypothesis (there is no difference between the two interventions) and the alternative hypothesis (there is a difference between the two interventions or, more commonly for superiority trials, that the novel intervention is more protective than standard practice). When testing a hypothesis, there are two types of error possible. A type I error (the probability of which is termed alpha, α) is one in which the null hypothesis is incorrectly rejected – that is, there is no effect but it is reported that there is (a false positive). A type II error (the probability of which is termed beta, β) is one in which the null hypothesis is incorrectly not rejected – that is, there is an effect but it is not detected (a false negative).

Table 1. Process of carrying out an efficacy study for vector control interventions

Step	Process	
Step 1	Develop the research question	• Define your research question • What are the study Population, Intervention, Comparator and Outcome (PICO) ?
Step 2	Decide on the experimental unit	• Decide which experimental unit is appropriate for the intervention and study question
Step 3	Design the study	• Decide on a suitable study design
Step 4	Determine the sample size	• Calculate the sample size required
Step 5	Evaluate contamination effects	• Could contamination or spillover effects due to movement of the vector or study participants, or changes to the intervention (such as the sharing of insect repellent) be reduced through the design of the study or data analysis?
Step 6	Assess blinding	• Is it possible to use blinding in the study? • Who will be blinded (participants, clinicians, investigators, outcome assessors)?
Step 7	Determine how to implement the intervention	• Put measures in place to ensure high-quality implementation and adherence to the intervention
Step 8	Determine how to measure the outcome(s)	• How will outcomes be measured? • How often or over what period will outcomes be measured? • What measures will be put in place to reduce loss to follow-up? • Establish a trial steering committee and data and safety monitoring board (DSMB)
Step 9	Write the protocol and register the study	• Write the protocol; include a description of the sample size calculation and a brief statistical analysis plan, the standard operating procedures, and information sheets and informed consent forms for participants • Register the study and publish or make the protocol publicly available before recruitment is completed
Step 10	Obtain ethical review and oversight of the study	• Obtain relevant ethical clearances for the study and import permit for test products • Establish a study steering committee and a board to monitor data safety
Step 11	Randomize the experimental units and conduct the study	• Randomize the experimental units • Conduct the study and generate data • Ensure high-quality data recording
Step 12	Ensure good data management and perform statistical analyses	• Write the plan for statistical analysis • Clean and finalize the data; have an independent statistician un-blind the data if appropriate • Perform the statistical analysis
Step 13	Report the study's findings	• Write up and disseminate the study's findings

Table 2. Components of good research questions based on defining the Population, Intervention, Comparator and Outcome (PICO)

Parameter	Detail	Examples	
		References *37, 38*	Reference *39*
P – Population	What study population are you interested in?	In a rural population living in the Upper River Region of The Gambia…	…do long-lasting insecticidal nets and routine control measures (case detection and treatment, and indoor residual spraying)…
I – Intervention	Which intervention are you interested in?	…do long-lasting insecticidal nets and indoor residual spraying ….	…do long-lasting insecticidal nets and routine control measures (case detection and treatment, and indoor residual spraying)…
C – Comparator	Which intervention or strategy do you want to compare?	…compared with using long-lasting insecticidal treated nets alone…	… compared with using routine control measures alone…
O – Outcome	What is the effect of the intervention on public health outcomes?	… provide additional protection against clinical malaria?	.. reduce the incidence of visceral leishmaniasis?

2.2.1 Superiority and non-inferiority studies

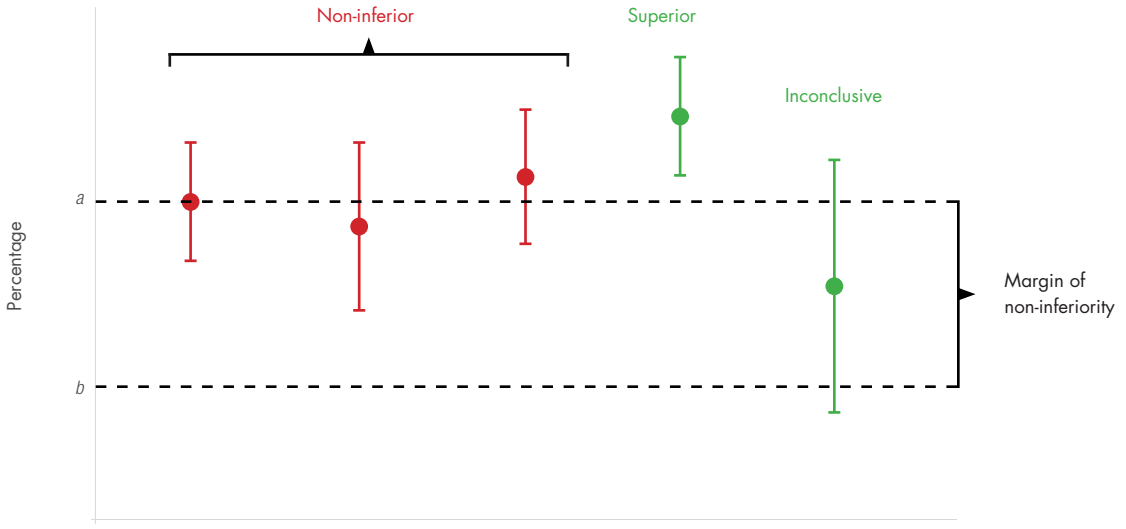
Studies can set out to determine whether the new vector control intervention is superior to the control intervention in a superiority study or non-inferior to the control (standard best practice) in a non-inferiority study (*40*). A non-inferiority study can be used to determine whether a new product is as effective as an existing product at a lower cost or with fewer unwanted side effects. Fig. 2 illustrates how these types of study differ.

The non-inferiority margin is the largest difference between the control and the intervention by which the intervention would be considered to be non-inferior by expert opinion; it is context dependent and not a statistical consideration. It must be specified before the study is conducted. The non-inferiority margin influences the study's sample size: the smaller this margin, the larger the required sample size. In practice, a 10% difference is often used as a margin (M), although this is arbitrary.

The null hypothesis in standard superiority trials states that there is no difference between the interventions. In contrast, non-inferiority trials have as their null hypothesis that the experimental treatment is inferior to the standard treatment by at least the margin of non-inferiority. The treatment (T) is superior to the control or comparator (C) if the lower bound of the confidence interval of the difference between the treatment and control is greater than zero. The treatment is non-inferior to the control if the confidence interval of the difference between the treatment and control does not encompass the lower bound of the non-inferiority margin.

If a superiority study is conducted badly (for example, the intervention is poor; there are high contamination effects in cluster-randomized controlled trials [CRTs], there is high loss to follow-up, or poor quality data) it makes it more difficult to show superiority – that is, a type II error is more likely. However, if a non-inferiority study is conducted poorly, it is easier to show non-inferiority, making the rejection of the null hypothesis more likely – that is, increasing the chances of a type I error. Similarly, if both interventions are ineffective, the study will demonstrate non-inferiority. Therefore, it is imperative that non-inferiority studies are shown to be well conducted.

Fig. 2. Schematic illustrating superiority and non-inferiority studies. Treatment is non-inferior to control if the lower bound of the confidence interval of PE falls within the margin of non-inferiority, and it is superior if the lower bound is above the margin of non-inferiority. The study is inconclusive if the upper and lower bounds of the confidence intervals are outside the margin of non-inferiority. On the y axis, a% and b% represent, respectively, the upper and lower bounds of the margin of non-inferiority.



2.1.2 Choosing outcome measures

Studies generally have primary outcomes and secondary outcomes (also known as primary and secondary end points). The primary outcomes should be clear, well-defined and assessable measures of effect, and usually one or two primary outcomes are chosen. There can be any number of secondary measures, although they should all be relevant to the declared aims of the study. Secondary outcomes are effects of secondary importance.

Given the limited funds available for disease control, stakeholders, including most importantly programme managers, need to implement interventions that demonstrate public health value. Epidemiological outcomes are necessary to demonstrate that a vector control intervention is efficacious in preventing infection or disease, or both, in humans. Therefore, it is essential that there is always an epidemiological primary outcome for phase III studies. Entomological outcomes can be used in combination with an epidemiological outcome to evaluate the claimed entomological effect. A phase III study should also collect data about the physical and chemical durability of the product, as this remains a requirement under WHO's efficacy testing guidelines for large-scale field trials (WHOPES phase III).

Phase III studies need to demonstrate public health value in order for an intervention to be recommended by WHO. Therefore, it is imperative that studies have epidemiological outcomes.

The best epidemiological measure is generally the incidence of clinical disease, disease-specific mortality or the prevalence of infection because these are relevant to public health. For some diseases, such as dengue, validated correlates, such as seroconversion measured by sequential blood samples (known as the seroincidence), can be good proxies for past disease or infection (*41,42*). Studies should use WHO's recommended case definitions and parasitological diagnosis or serological or molecular verification (*43–47*) to allow data to be compared among sites and studies.

Entomological outcomes are not always good predictors of epidemiological outcomes. For example, although topical repellents are effective in preventing mosquito bites at the individual level, community trials have generally not shown their efficacy in preventing malaria (*48*). When possible, it is preferable to use entomological outcomes that relate to disease transmission, such as vector densities, vector biting rate or the entomological inoculation rate. Larval surveys are not recommended for evaluating the effect of vector control on Aedes-borne diseases. These traditional indicators of immature *Aedes* abundance (for example, house index = percentage of houses with larvae or pupae, or both) give poor indications of adult populations because the majority of larvae and pupal stages do not develop into adult mosquitoes (*49,50*). Better measures to use include pupal demographic surveys or measurement of the density of adult vectors, although these are more labour intensive than larval surveys (*50–53*).

2.1.2.1 Problems of multiplicity

Many trials test multiple hypotheses, such as by comparing a new treatment with several other treatments or comparing a new treatment with a control intervention repeatedly over time (common in time series studies). This can be problematic because testing multiple hypotheses can inflate the chance of making a type I error (in which the null hypothesis is incorrectly rejected – that is, there is no effect but an effect is reported). The degree of multiplicity can be reduced by prioritizing the questions the study sets out to address rather than by trying to test everything. If multiplicity persists, then statistical techniques (for example, the Bonferroni correction, false discovery rate procedures) can be used to adjust the results.

2.1.3 Choosing a comparator

Studies should always have a control group from which data are collected contemporaneously with data from the intervention group. Studies without a control group or those using a non-contemporaneous control group provide lower-quality evidence and are not recommended (see **section 2.3.7** and **Table 3** for more details). This is because longitudinal changes, such as rainfall, or changes in diagnostic practices may have an impact on entomological and epidemiological outcomes and can exaggerate or mask the effect of an intervention.

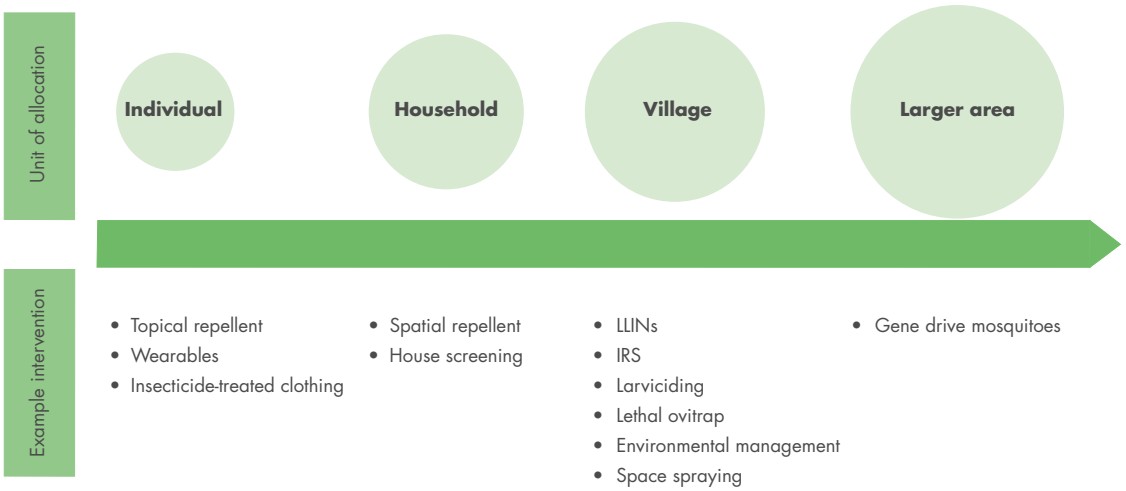
For ethical reasons the control group must receive care that reflects the standard best practice interventions (that is, routine vector control measures, such as the mass deployment of LLINs). Asking whether an intervention plus standard practice is better than standard practice alone is a good operational question from a policy perspective because a new intervention should not be recommended if it does not provide protection that is better than standard best practice. For blinding purposes, a placebo, such as an untreated bed net or insect repellent without an active ingredient, is sometimes given in addition to standard best practice (see, for example, references *54–56*).

2.2 Step 2: decide on the experimental unit

A key concept in study design is the choice of experimental unit – that is, the participant or group of participants exposed to a particular treatment. Vector control interventions may be utilized by individuals or groups of individuals, depending on the intervention in question (Fig. 3). For example, topical insect repellents or insecticide-treated clothing may be allocated at the individual level, spatial repellents and house screening at the household level, LLINs or IRS at the village level, and environmental management or the release of genetically modified mosquitoes over larger geographical areas. In fact, the allocation of interventions to groups of individuals known as clusters is common in vector control studies. There are several reasons for this (*15*). First, many vector control tools are, by their nature, applied to groups of people or communities. For example, spatial repellent may be allocated to a household, or an environmental sanitation intervention against dengue may be allocated at the community level. Second, cluster allocation can help reduce the contamination between study arms that might occur if individuals within the same community receive different interventions (for example, insect repellent might be shared with family members within the same household or village); additionally, most vector control interventions have an individual-level as well as a community-level effect. Last, cluster allocation allows assessment of the indirect effects of an intervention (see Section 2.3.3, Box 2).

In studies of individuals, the observational and experimental unit are typically the same. However, for CRTs, the observational unit and experimental unit may be the same (for example, in a household study of LLINs, mosquito density in houses may also be measured) or different (for example, in a household or village study an epidemiological outcome in an individual or individuals from within the household or village may be measured).

Fig. 3. Schematic illustrating how levels of allocation differ depending on the type of vector control intervention being studied[a]



a LLINs, long-lasting insecticidal nets; IRS, indoor residual spraying

2.2.1 Replication

Replication is when multiple experimental units receive the same particular intervention, for example, when multiple households receive LLINs. Whenever an intervention is implemented, there is likely to be natural variability in the response. Thus, by having replicates one can estimate the variability within a treatment and increase the precision of the estimate.

Replicates are not the same as repeated measurements of the same item (which are known as repeats). Repeats are measurements taken during the same experimental run or consecutive runs, but replicate measurements are taken during identical but different experimental runs.

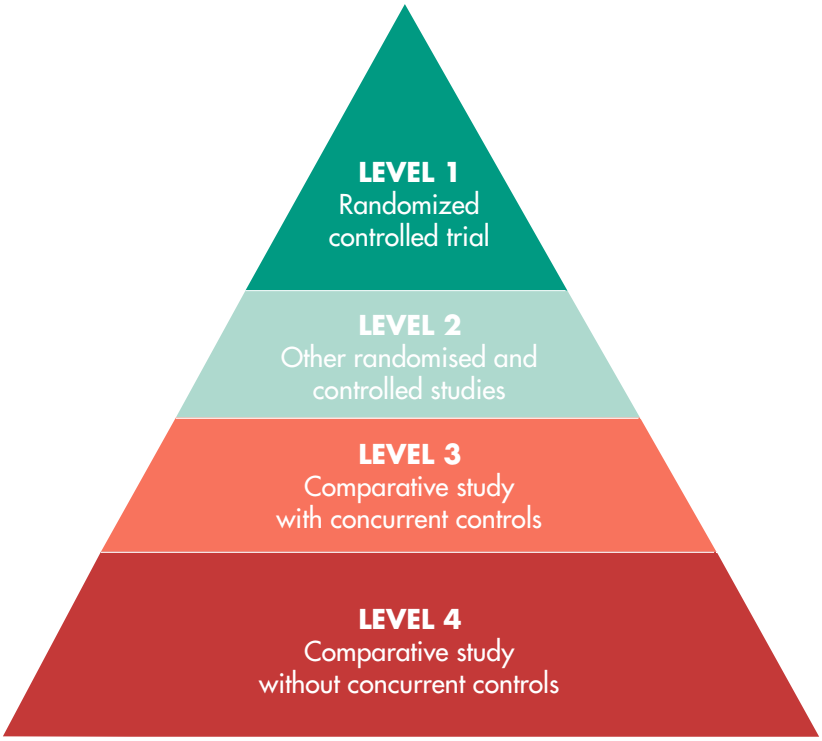
Repeat measurements taken from the same observational units are likely to be highly correlated, such as when using a light trap in the same house for repeated collection. As a result, repeated measures of the same observational unit give information about variability within that experimental unit rather than about variability in the effect of the treatment. Pseudoreplication should be avoided when possible. This refers to using inferential statistics to test for treatment effects with data from experiments or observational studies in which either the treatments are not replicated (although the samples may be) or the replicates are not statistically independent. Unfortunately, pseudoreplication is a common problem with vector control trials, many of which use two villages (*57,58*) or two areas (*59,60*), with one in which the intervention was introduced and the other acting as a control. This is a poor design because using only two clusters means that inferential statistics cannot separate variability caused by treatment from variability in the experimental units (*61,62*). Pseudoreplication can also occur when, for example, households are used as experimental units but then the outcome is measured for all individuals within the household.

It is important to have sufficient replicates (that is, multiple experimental units) that receive the intervention to increase the precision of the estimate of the effect size. This is not the same as measuring the outcome multiple times within the same experimental unit.

2.3 Step 3: decide on the study's design

An overview of the hierarchy of study designs is presented in Fig. 4.

Fig . 4. Hierarchy of study designs reflecting the strength of evidence when assessing the efficacy of vector control interventions (adapted from references *63, 64*)



Study designs recommended by WHO to assess the public health value of a new intervention or product	
Level 1. **Randomized controlled trial**: individual or cluster randomized	Recommended
Level 2. **Randomized controlled trial**: step-wedge, cross-over, factorial design	Recommended
Level 3. **Non-randomized trial with control**: before-and-after studies, cohort study, case–control study, cross-sectional study, time-series or interrupted time-series	Recommended on a case-by-case basis
Level 4. **Trials without a control or using a historical control group**: such as time series or interrupted time series without control group	Not recommended

2.3.1 Bias and confounding

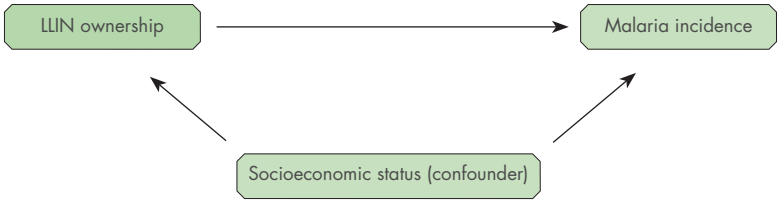
Bias or systematic error is the risk that a study will overestimate or underestimate the effect of an intervention, and this risk does not depend on the sample size. The findings of a study with a high risk of bias are likely to be invalid. However, different aspects of a study's design, which are discussed in more detail in this section, can reduce the risk of introducing bias.

There are many types of bias; the most common and important types are selection bias, information bias and confounding bias. Selection bias occurs when those included in a study are not representative of all those eligible for inclusion. This can lead to an incorrect estimate of the association between the exposure (or intervention) and the outcome.

Information bias occurs when inaccurate information is collected about either the exposure (intervention) or the disease, or both, and this can lead to a biased estimate of the association between the exposure (intervention) and the outcome. Important subtypes of information bias are recall bias and observer bias, both of which are more common in particular study designs.

Confounding is a type of bias that arises whenever there is a common association between the outcome and the exposure. This is common in observational studies and non-randomized trials. For example, in an observational study of the association between LLIN use and malaria incidence, the relationship is likely to be confounded by socioeconomic status because those of higher socioeconomic status may be more likely to own an LLIN and may have a reduced risk of malaria independent of LLIN ownership, due to living in better housing or having access to diagnosis and treatment (Fig. 5). A variable that is on the causal pathway between the exposure and the outcome is not a confounder. For example, if IRS reduces mosquito density and this results in a lower incidence of malaria, then mosquito density is not a confounder despite being associated with both the exposure and the outcome. One way of dealing with confounding is to restrict comparisons to individuals who have the same or similar values for a confounding variable (that is, individuals in the same stratum). This is known as a stratified analysis. For example, the association between LLINs and malaria incidence could be assessed separately for males and females or according to socioeconomic status. Alternatively, more sophisticated methods are available that allow adjustments to be made for confounding variables without the need for stratification. Treating a variable on the causal pathway as a confounder can introduce bias. Both positive confounding (when the observed association is biased away from the null) and negative confounding (when the observed association is biased towards the null) can occur.

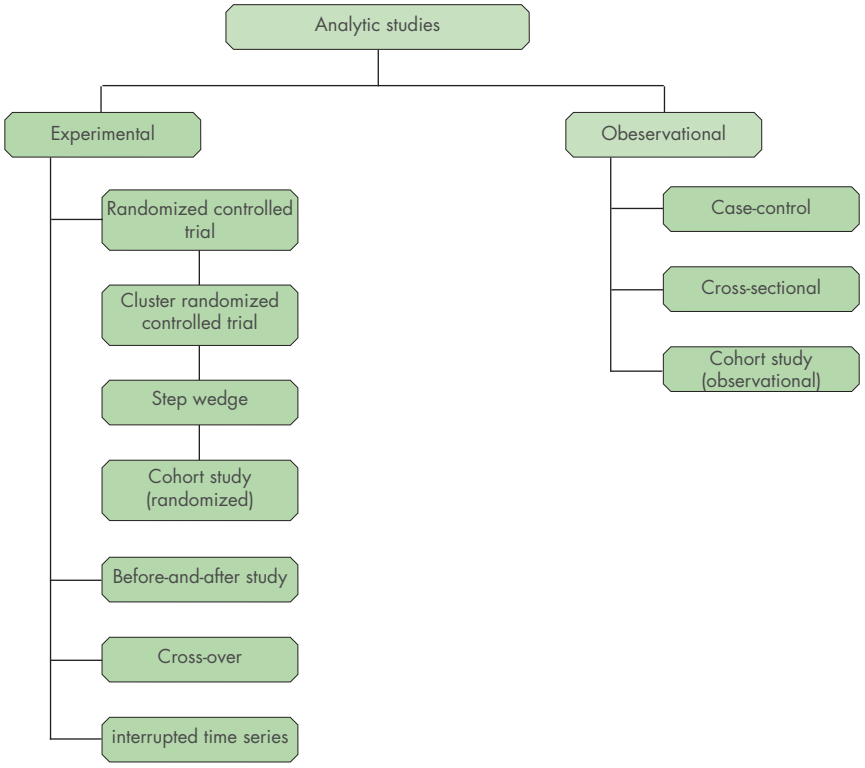
Fig. 5. Example of confounding bias when investigating the association between ownership of a long-lasting insecticidal net (LLIN) and malaria. In an observational study, factors of socioeconomic status may confound the observed relationship between LLIN use and malaria incidence. Persons of higher socioeconomic status may be more likely to own an LLIN, and independent of LLIN ownership, better housing and access to diagnosis and treatment will reduce risks of malaria.



2.3.2 Types of study designs

Analytical studies, in marked contrast to descriptive studies, try to quantify the relationship between two factors – that is, the effect of an intervention (or exposure) on an outcome (Fig. 6). Analytical studies can be further separated into experimental studies or observational studies. In experimental studies, investigators assign treatments to experimental units and observe the outcome. Experimental studies are carefully controlled so that they provide generalizable, statistically rigorous answers to a specific question. In observational studies, the investigator measures the association between the intervention and the outcome, but does not determine the treatment that each participant receives. Although observational studies are often more practical to carry out and less resource intensive, they provide only correlations, not generalizable conclusions.

Fig. 6 . Analytical study designs for vector control interventions



2.3.3 Randomized controlled trials

Randomized controlled trials (RCTs) provide the least biased, most robust estimate of an intervention's efficacy, and they are generally considered the gold standard in study design for evaluating the efficacy of a protective intervention. If the number of experimental units is sufficiently large, randomization (the allocation to treatment and control groups in a random fashion) is expected to provide no systematic differences between groups that could be caused by confounding variables (*15*). Due to the law of large numbers, randomization with a sufficiently large number of experimental units is expected to achieve balance between the treatment and control groups.

Even if randomization is used, it is good practice to check that the baseline characteristics of the groups are similar (*65*). If there are differences in the characteristics of the intervention and control communities, potential biases can be reduced by adjusting for these differences using stratification or regression models that allow adjustment for potential confounding variables (see, for example, reference *66*). However, there is no guarantee that this will fully control for confounders that may be unknown or unmeasured.

2.3.3.1 Cluster-randomized controlled trials

Many vector control interventions are implemented at a cluster level, meaning at the household, village- or district level. Cluster-randomized controlled trials (CRTs) of vector control interventions – trials in which the intervention or control is randomly allocated to clusters – are therefore relatively common. There are many examples of CRTs, including a study measuring the efficacy of LLINs against mortality in children in western Kenya (*67*). CRTs are also able to capture the indirect effects of an intervention, such as those that occur through reductions in transmission or mass killing of the vector population (**Box 2**).

Box 2. Direct, indirect, total and overall effects

Conventional individually randomized trials are designed to capture the direct effect of an intervention on the individuals who receive it. Cluster-randomized controlled trials (CRTs) also allow measurement of the indirect effect of an intervention – that is, the benefit that individuals may obtain due to other members of the community receiving the intervention. Indirect effects occur in two ways. First, vector control interventions can reduce the pool of infected individuals in the community and, therefore, reduce the transmission of infections to other people in the community. Second, vector control interventions can result in mass effects, which are additional effects achieved when a substantial proportion of the population receives the intervention. For example, the mass killing of mosquitoes that come into contact with LLINs can reduce the survival of the vector population, leading to indirect protection and reducing transmission to individuals not using the LLINs (*68*).

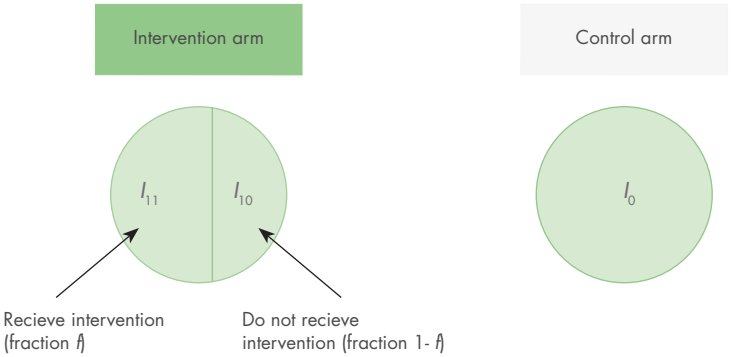
When considering designing a CRT to measure the direct and indirect effects of an intervention, it is useful to distinguish between the different effects that it is possible to measure using this design (**Fig. 7**) (*15*). In intervention clusters, the intervention is assumed to be delivered to a randomly sampled proportion of the population (fraction f), while the remaining individuals do not receive the intervention (fraction 1 – f). An example of this would be a CRT of improved housing in which improved housing is randomly allocated to only a proportion of houses in the intervention arm (*69*).

.../...

The direct effect of the intervention (the protective efficacy, or PE_{direct}) can be estimated by comparing the incidence of the outcome in those receiving the intervention and those not receiving it within the intervention clusters, denoted I_{11} and I_{10}, respectively:

$$PE_{direct} = 1 - \frac{I_{11}}{I_{10}}$$

Fig . 7. Diagram showing incidence rates in intervention and control arms used to determine measures of direct, indirect, total and overall effects (after reference 15)



The indirect effect of the intervention can be estimated by comparing the incidence of the outcome among those in the intervention cluster who did not receive the intervention with the outcome of those in the control clusters, denoted I_{10} and I_0, respectively:

$$PE_{indirect} = 1 - \frac{I_{10}}{I_0}$$

The indirect effect can result from the effects of the intervention on infectiousness or, alternatively, from mass effects.

The total effect of the intervention is calculated by comparing the incidence among those receiving the intervention in the intervention clusters (I_{11}) with the incidence in the control clusters (I_0):

$$PE_{total} = 1 - \frac{I_{11}}{I_0}$$

Finally, the overall effect of the intervention is obtained by comparing the overall incidence in the intervention and control arms:

$$PE_{overall} = 1 - \frac{I_1}{I_0} = 1 - \frac{f \chi I_{11} + (1-f) \chi I_{10}}{I_0}$$

2.3.3.2 Step-wedge cluster-randomized trials

In a step-wedge CRT (SWCRT), the intervention is rolled out to clusters in a staged fashion and the order in which clusters receive the intervention is determined by randomization (**Fig. 8**) (*70–72*). The design includes an initial period during which no clusters are exposed to the intervention. Each time point at which a new cluster receives the intervention is called a step, and all clusters will have received the intervention after the final step. Before and after each step, the dependent variable of interest is measured in all clusters. There should be multiple clusters in each step so that a cluster does not become confounded by the step. The number of clusters, the intervals between steps and the total duration of the trial need to be specified in the study protocol prior to the start of the study

An SWCRT may be used when logistical, practical or financial constraints make the staged roll out of an intervention desirable or when an intervention is being rolled out by a programme and researchers want to use the roll out to collect data . SWCRTs are also able to test whether local elimination of transmission is possible because the intervention expands to complete coverage (*73*). Examples of an SWCRT design include a trial that used odour-baited traps for mosquitoes (*74*) and a trial comparing whether combination LLINs (impregnated with pyriproxyfen and pyrethroids) provide additional protection compared with standard pyrethroid-impregnated LLINs (*75*).

Because clusters act as their own controls, SWCRTs are less sensitive to between-cluster variation and, thus, in some cases they may require a smaller sample size than parallel designs (*76*). However, although they are useful in some circumstances, step-wedge studies may be more difficult to interpret than CRTs and, therefore, their use should be carefully considered. For example, SWCRTs are subject to bias if there are temporal trends in a disease, as one of the assumptions in an SWCRT is that the risk of disease does not change over time. This can be a problem even if randomization ensures a

Fig. 8. Designing step-wedge cluster-randomized studies for vector control interventions. These studies start with a period during which no clusters are exposed to the intervention. After this initial period, the intervention is rolled out to the clusters in a staged fashion, so that by the end of the study all clusters have received the intervention. The order in which clusters receive the intervention is determined by randomization.

E3						
E2						
E1						
D3						
D2						
D1						
C3						
C2						
C1						
B3						
B2						
B1						
A3						
A2						
A1						
	0	1	2	3	4	5

Time period (step)

balance between arms at baseline. Therefore, this design should be used cautiously if the incidence of a VBD is highly variable (*77*). SWCRTs can also be prone to imbalance between clusters, although constrained randomization could be used to minimize this. SWCRTs are less adaptable, for example, if a follow-up period needs to be extended because the observed VBD incidence is lower than expected as all clusters will have already crossed over to the intervention by the end of the study (*77*).

Step-wedge studies are a type of cluster-randomized trial that has utility in some specific circumstances. However, they should generally be performed only if a standard cluster-randomized trial cannot be and only if there is already good evidence that the intervention is effective and should be rolled out to the entire population.

2.3.3.3 Randomized cohort design

RCTs can include a cohort design if the intervention and control cohorts are established in each study arm (see, for example, references (*37,38,69*). As with an observational cohort study, the outcome of interest is measured in the study cohorts allocated to the control and intervention groups, and the outcomes are compared to estimate the efficacy or effectiveness of the intervention.

2.3.3.4 Factorial design

In a factorial trial, two (or more) interventions can be compared simultaneously. For example, if one wishes to study both the use of larvicide and the early diagnosis and treatment of malaria, instead of conducting two separate trials one could conduct a study with four groups consisting of a control group; a group receiving early diagnosis and treatment; a group benefiting from larviciding; and a group that receives early diagnosis and treatment plus benefits from larviciding (Fig. 9) (see, for example, reference *78*). Clusters or individuals should be randomized to an intervention or control group. As well as measuring the separate effects of each intervention, factorial designs also permit an assessment of potential interactions between treatments. The factorial design estimates the effect of each intervention on its own as well as the effect of the two interventions combined.

Fig. 9. Factorial design trial comparing four different combinations of interventions to prevent malaria

		Disease management intervention	
		No early detection and treatment	Early detection and treatment
Vector control intervention	No larviciding	Arm 1	Arm 2
	larviciding	Arm 3	Arm 4

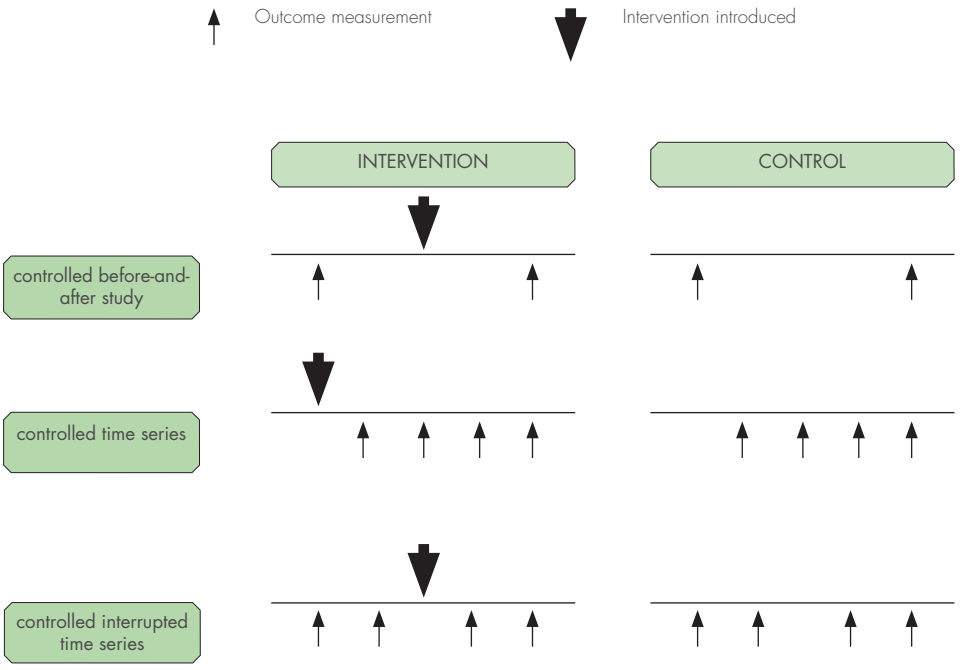
2.3.4 Controlled before-and-after studies, time series and interrupted time series studies

In controlled before-and-after studies, the outcome is measured on one occasion prior to and then again after the introduction of the intervention, and it is measured at the same time points in the control arm (Fig. 10). Controlled before-and-after studies are typically non-randomized.

In time series studies, data on outcome measures are collected at several time points once the intervention has been implemented in the intervention group. Interrupted time series studies involve collecting data on outcome measures at several time points in the intervention group before and after the intervention is implemented. In time series and interrupted time series studies there may also be a control group in which the same measurements are made at the same time points, although quite often there is no control in a time series study. Even if there is a control group, there are unlikely to be sufficient experimental units for randomization to be effective. Time series and interrupted time series designs are useful for evaluating the efficacy of some vector control tools, such as when control efforts need to be implemented over wide areas, as in

studies of human African trypanosomiasis in which vectors are highly mobile (see, for example, reference *79*), or following population-wide (for example, national) introduction of a control tool. They can be used prospectively (when the intervention is allocated by the investigator, such as in an experimental design) and also retrospectively to evaluate the efficacy or effectiveness of an intervention (when the intervention is not allocated by the investigator, such as an observational design).

Fig. 10. Schematic of controlled before-and-after, controlled time series and controlled interrupted time series studies for vector control interventions (adapted from reference *80*)

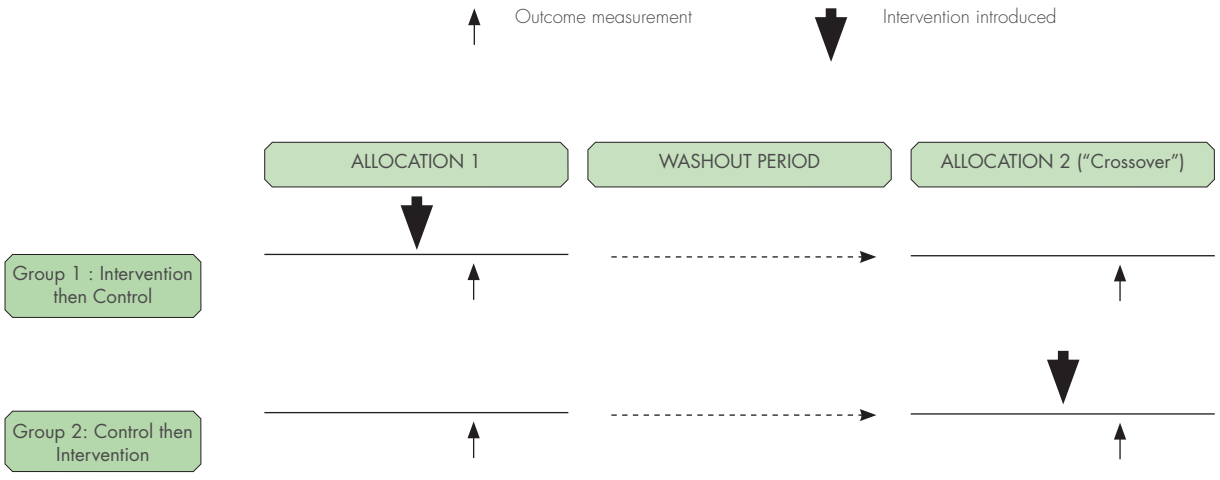


2.3.5 Crossover studies

Crossover studies can be used to assess interventions whose effect disappears when the intervention is removed. In this type of study, the experimental units are randomly allocated to the order in which they receive the intervention or control (for example, the units are allocated to control first then intervention, or to intervention first then control) (Fig. 11). The outcome measures are assessed and then, following a suitable washout period (during which indices have returned to baseline), the groups are switched (that is, they crossover) to the remaining group – either the intervention or control – and the outcome measures are assessed again. The main advantage of crossover studies is that each unit acts as its own control, thus reducing variation in outcome and allowing for much smaller sample sizes. Despite being an efficient

design, crossover studies are uncommon in vector control. This is because many vector control interventions have long-lasting effects, although crossover studies have been used to evaluate short-acting interventions, such as some larvicides (*81*), fly control (*82*), house improvements in a phase II study (*83*) and cattle sponging with insecticide (*84*).

Fig. 11. Schematic design of crossover studies for vector control interventions



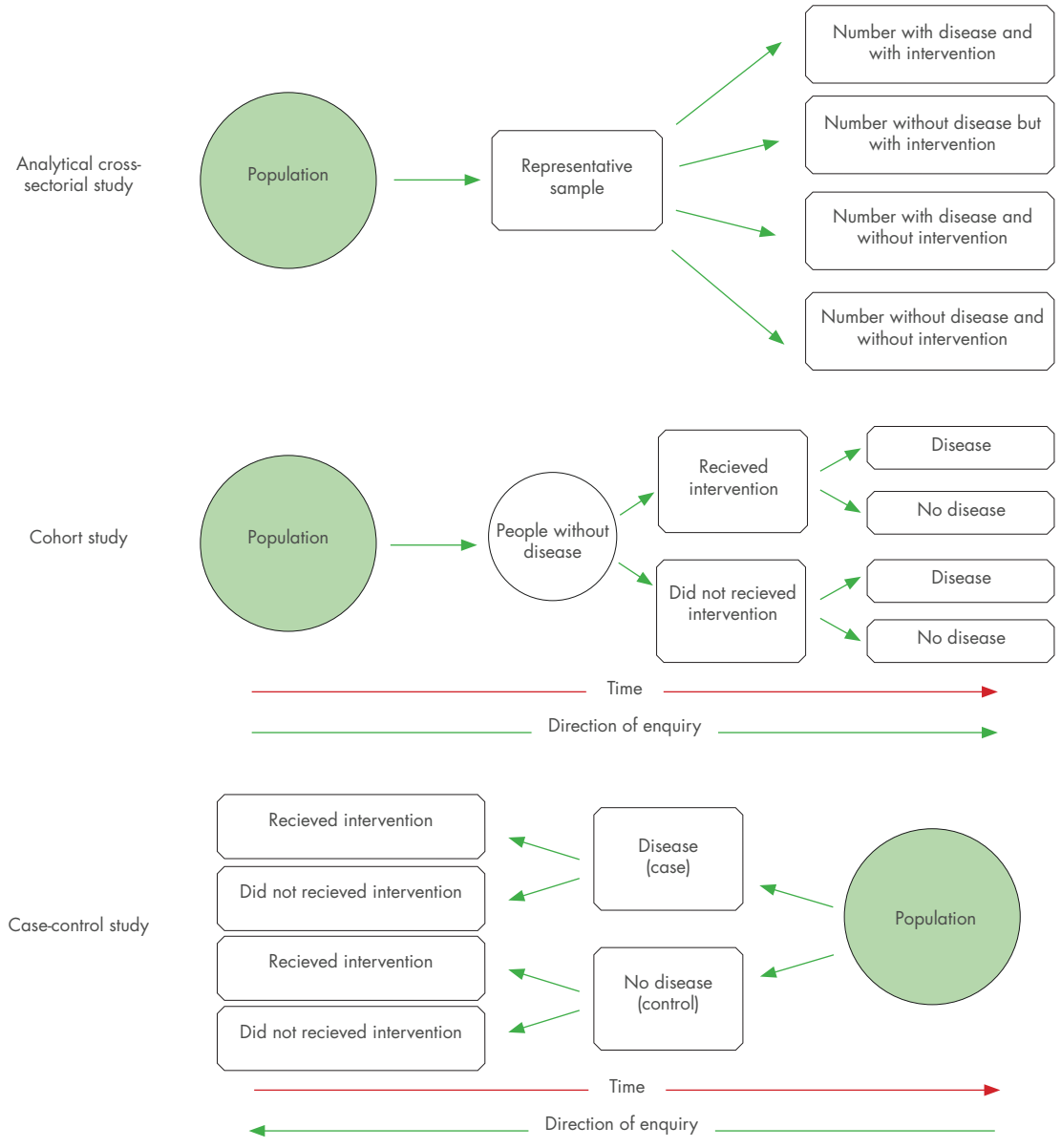
2.3.6 Observational studies

Observational studies, such as cross-sectional, cohort or case–control studies (Fig. 12), are sometimes used to generate evidence of the efficacy of vector control interventions. However, these designs should not be used when an experimental design is possible. Observational studies provide weaker evidence than experimental (that is, randomized) designs because they can be subject to bias, in particular selection bias, recall bias, observer bias and confounding.

In an analytical cross-sectional study, information is collected at one time point about the prevalence of the outcome of interest (for example, a disease or infection) and exposure (for example, the use of a protective intervention). Cross-sectional studies can be prone to recall bias because participants' recall of using an intervention may vary depending on whether they have the infection or disease of interest or not (*85*). For example, people who have had malaria (cases) might be more likely to recall using malaria control interventions than people who have not had malaria (controls) (*86*). The underreporting of exposure in one of the groups can lead to bias towards (or away from) the null. If the recall of using a protective intervention is equally poor in both controls and cases, then this biases effect estimates towards the null (*85*).

In a cohort study, two groups of disease-free people are identified: exposed (typically using a protective intervention of interest) and unexposed (not using a protective intervention of interest). The groups are then followed during a period of time to assess the outcome of interest (usually a disease or infection). In this study type, participants are not allocated to the intervention of interest. Cohort studies follow individuals for

Fig. 12 . Schematic design of observational studies for evaluating vector control interventions



an extended period during which they may die, migrate or refuse to continue to participate in the study. In addition, losses to follow-up may be related to the exposure, the outcome, or both. For example, individuals who develop the disease or infection may be less likely to continue to participate. The degree to which losses to follow-up are correlated with exposure and outcome can lead to attrition bias in the measures of the effects of exposure and outcome.

In a case–control study, a group of people with the disease of interest (cases) and a group of people without the disease (controls) are identified. Exposure is then retrospectively assessed, often through self-reporting. The prevalence of the exposure of interest (that is, the use of a protective intervention) is compared between the two groups. A study by Mathanga and colleagues in Malawi used a case–control design to measure the effectiveness of LLINs in a setting of pyrethroid resistance (*87*). Selecting controls can be problematic in a case–control study because they should represent the population from which the cases originated. Cases can be identified through either active or passive methods, although engaging in passive case detection at facilities is generally less arduous and costly than implementing active case detection in the community. If facility-based case ascertainment is used, controls can be selected from individuals attending the same facility but with illnesses other than the VBD of interest or from the community, which is generally the preferred approach, although this depends on having a good understanding of care-seeking behaviours in the community. It is possible to match cases and controls for confounding factors such as age or sex, however, matching in this way does not eliminate confounding completely. Once matched on a variable, it is no longer possible to evaluate that variable as a risk factor for disease or exposure, and a matched-pair analysis must be performed, thus reducing the overall statistical power. Case–control studies typically assess risk at the individual level. Therefore, case–control studies may incompletely measure the effect of interventions that may have a community-level effect (for example, the use of LLINs in circumstances that result in the mass killing of mosquitoes).

A specific type of case–control study is the test-negative case–control design, originally used for evaluating vaccine effectiveness against viral pathogens (*88*), but currently being used to evaluate the efficacy of *Wolbachia* against arboviruses in a trial in Yogyakarta, Indonesia. In this type of study, the use of an intervention is compared between cases who test positive and controls who test negative and present to a health facility with an arboviral-like illness. A major advantage of this design is that by recruiting controls from those who present with arboviral-like illness but test negative for arboviruses, both money and time are saved by not having to screen large cohorts of individuals for arboviruses. This design differs from a standard case–control study in that the marginal ratio of cases to non-cases (or controls) is not specified or even known during enrolment, which occurs prior to testing.

The aim of a case–control study is to select controls who are representative of the population that produced the cases. Controls are used to provide an estimate of the exposure rate in the population. Therefore, case–control studies can be prone to selection bias when those individuals selected as controls are unrepresentative of the population that produced the cases. Case–control studies can also be prone to recall bias if the recall of using an intervention differs between cases and controls.

Observational studies do not provide as strong evidence about the efficacy of an intervention as randomized controlled trials because they are more prone to bias. Observational studies are not recommended if it possible to randomize the assignment of the intervention.

2.3.7 Hierarchy of study designs for estimating intervention efficacy

In summary, the methodological quality of study designs varies, so some are better able than others to answer the question, "Does the intervention work?" or "Does this intervention work better than that intervention?" (*89*). In other words, study designs vary in the degree to which they allow observed effects to be attributed to the intervention with confidence. **Table 3** summarizes the pros and cons of the different types of study designs and their utility. In **Fig. 4** we provide a hierarchy of study designs that can be used to evaluate the efficacy of vector control interventions, ranking types of studies as level 1, 2, 3 or 4 according to their methodological quality. This figure is based on previous work by the Oxford Centre for Evidence Based Medicine's levels of evidence from 2011 and guidance from the National Health and Medical Research Council of Australia (*63, 64*).

RCTs and CRTs are the gold standards of study designs for evaluating the efficacy of vector control interventions (level 1). Randomization reduces the risk of confounding by ensuring that the background characteristics of the control and intervention groups – such as disease prevalence, demographic variables and socioeconomic status – are similar to one another. Other types of studies that are both randomized and controlled, such as SWRCTs and crossover designs, are ranked as level 2. Level 3 studies are comparative studies with concurrent controls that are non-randomized, for example, before-and-after, cohort, case–control, and cross-sectional studies; and time series or interrupted time series studies. These studies have a higher risk of bias. Studies without a contemporaneous control group are not recommended (ranked as level 4). In these studies, it is difficult to attribute observed changes to the intervention because there may be secular trends or changes in disease incidence that are not due to the intervention.

Table 3. Pros and cons of different study designs for evaluating the efficacy of vector control interventions

Category of study design	Study design	Pro	Con	Main use of study design
Randomized controlled trial	Individual randomized controlled trial	Randomization ensures low risk of selection bias if the number of randomization units is sufficiently large; Blinding is more likely	Can be prone to observer bias or performance bias if study is unblinded or becomes unblinded Can be expensive	Rigorous evaluation of an intervention allocated at the individual level for which no community-level protection is expected
	Cluster-randomized controlled trial	Randomization ensures low risk of selection bias if the number of randomization units is sufficiently large; Blinding is more likely	Can be prone to observer bias or performance bias if study is unblinded or becomes unblinded; Potential for spillover effects; Can be expensive	Rigorous evaluation of an intervention allocated at cluster level
Other randomized and controlled studies	Step-wedge	Useful when logistical, practical or financial constraints make the staged roll out of the intervention desirable; May require a smaller sample size compared with standard parallel designs	Can be expensive; Can be biased if there are temporal trends in the disease	Evaluation of logistically or practically complicated interventions that cannot be rolled out everywhere simultaneously
	Crossover	An efficient design because participants or clusters serve as their own controls, thus reducing variation and sample size	Potential for carry-over effects if the washout period between the reallocation to the intervention and the control groups is too short	Can be used when the intervention has a short-lasting effect
Comparative studies with concurrent controls	Before-and-after	Relatively cheap, simple and convenient	Typically non-randomized and so weaker than RCTs in establishing causality; Difficult to identify a suitable control group with comparable disease situation at baseline; Not able to control for secular trends in the data due to two periods of measurements (pre- and post-); Usually involve a few clusters (a CRT with a sample size calculation would be better);	Suitable for natural experiment–type studies and when interventions need to be implemented over large areas so that cluster randomization is not feasible
	Time series	Use of a control group and multiple time points allows for the evaluation of secular trends	Typically non-randomized and so weaker than RCTs in establishing causality; No pre-intervention data, so unable to control for baseline differences; Need sufficient data points to do time series analysis (typically at least three)	

Category of study design	Study design	Pro	Con	Main use of study design
Comparative studies with concurrent controls	Interrupted time series		Typically non-randomized and so weaker than RCTs in establishing causality; Need sufficient data points to do time series analysis (typically at least three before and after)	
	Case–control study	Good for evaluating rare outcomes or those with a long lag time between exposure and disease;	Prone to selection bias due to difficulty in selecting controls who are representative of the population from which cases arise;	Can be used for phase IV studies of interventions when it is ethically unacceptable to use unexposed individuals for the control group
		Quick and relatively cheap	Potential for recall bias in recollection of intervention use; High risk of confounding; Can give only the relative risk of the outcome	
	Cohort study	Good for evaluating rare exposures;	Loss to follow-up and attrition bias are potential problems;	Good for studying the effects of risk factors
		Can establish timing and directionality of events (unlike case–control studies)	High risk of confounding because exposure may be linked to a hidden confounder; For rare diseases or those that take time to manifest, the follow-up period may be lengthy; Can be expensive	
	Cross-sectional study	Cheap and simple	At most, establishes association, but does not establish causality; Potential for recall bias in recollection of intervention use; High risk of confounding	Good for studying the effects of risk factors

CRT, cluster-randomized controlled trial; RCT, randomized controlled trial.

2.4 Step 4: calculate the sample size

Sample size calculations are carried out prior to conducting a study to determine the number of participants that need to be recruited to answer the study's question, assuming a given power and level of statistical significance (*90–92*). The sample size determines the size of the standard error of the effect estimate. Small sample sizes lead to wide confidence intervals around the estimated effect and, hence, poor precision. Sample size calculations are required to quantify the probability of a trial detecting an intervention effect of a minimum size, given the number of experimental units and the variability of the effect among them.

Vector control studies can have primary outcomes that are both clinical and entomological, and it is important to ensure that a sample size calculation is performed for all primary outcome measures, regardless of whether they are epidemiological or entomological. Unfortunately, it is common to find vector control studies without a sample size calculation.

All study protocols should include a justification of the chosen sample size.

Sample size calculations should be performed for both epidemiological and entomological primary outcomes.

For individually randomized trials, a number of parameters are required to calculate sample sizes. First, the prevalence or incidence of the outcome (or other outcome measure, for example, the hazard rate) in the control group is needed. This quantity can be obtained from previous studies conducted in the area or from baseline data. Second, the smallest expected effect size of the intervention should be defined that is relevant from a public health or clinical perspective. Generally, an intervention would be recommended for operational deployment if it resulted in at least a 30% reduction in epidemiological outcomes, but this cut-off obviously depends on the intervention and the setting in which it would be used. For example, a study assessing the effect of house screening against exposure to malaria vectors aimed to detect a reduction in house entry by malaria mosquitoes of at least 50% (*93*). Third, the variance around the effect size must be calculated: for binomial and for count data, the variance is implicit from the expected prevalence or incidence in the control group; for continuous data, it needs to be specified independently. Fourth, the significance level (or P value) that represents the probability of a type I error must be selected. Generally P = 0.05 is used, which means that there is a 5% probability of a type I error (that is, of falsely concluding that an intervention effect exists). Last, the power of the study or the probability of finding an intervention effect that actually exists must be calculated. The sample size needs to be large enough to ensure that the probability of a type II error is reasonably small, at most 20% (which is 80% power), but preferably 10% (which is 90% power).

Many vector control trials use a cluster-randomized design. Because outcomes measured in individuals or sampling sites within the same cluster are likely to be more similar than they are between clusters, the sample size calculation needs to account for the additional variation in the outcome between clusters. The level of between-cluster variation in the outcome is measured by the coefficient of variation (k), which is defined as the ratio of between-cluster standard deviation to the mean. If k = 0.25, the between-cluster variation in outcomes varies from about half the mean to about 1.5 times the mean; if k =0.5, the variation is from about zero to twice the mean. A large k value implies substantial between-cluster variation in the outcome, which makes it harder to show an intervention effect unless the sample size is increased (**Box 3**). If it is not possible to estimate k, then k = 0.5 is often assumed as a conservative value. When publishing CRTs it is important to include the k value in the manuscript so that other researchers can use it. Ten clusters per arm are recommended as an absolute minimum (*94*), and it is generally better for CRTs to have a larger number of

small clusters than fewer large clusters, with the proviso that the clusters should not be so small that spillover effects become too large (*15*). Hooper and colleagues describe sample size calculations for longitudinal studies with a step-wedge design (*95*). For entomological outcomes, the same principle applies: it is better to sample from many different clusters (see, for example, reference *38*) than conduct intensive collections in one or two clusters (see, for example, reference *96*). **Box 4** provides an example of how two typical village studies without sufficient replication and without a sample size calculation can be improved.

Box 3. Simulated sample size calculations for a cluster-randomized controlled trial looking at outcomes in dengue infection (92).

The table illustrates how changing the parameter values for a sample size calculation affects the cluster sample size and overall sample size, assuming a fixed power of 80% and two-sided significance level of 5% (P = 0.05).

	Protective efficacy (%)	Disease in control arm (%)	Disease in intervention arm (%)	Coefficient of variation (k)	No. people/ cluster	No. clusters/ arm	No. people/ arm	Comment
Parameter values for a sample size calculation	50	10	5	0.5	100	12	1 200	Exemplar
	50	10	5	**0.25**	100	5	500	**Reducing k to 0.25** (indicating there is less between-cluster variation in the outcome) means that fewer clusters per arm are needed with the same sample size per cluster.
	50	10	5	0.5	**200**	12	2 400	**Increasing the cluster size to 200**: Many investigators believe that the bigger the cluster size the better. However, in this example the number of people per cluster has been doubled but the power remains at 80%, illustrating that there is no benefit in increasing the sample size per cluster.
	33	10	7	0.5	100	36	3 600	**Reducing the desired protective efficacy**: In this example, the effect size (or protective efficacy) sought has been reduced from 50% to 33%. A larger number of clusters (and a larger overall sample size) are required to detect a smaller difference.
	50	**5**	**2.5**	0.5	100	13	1 300	**Assuming a rarer outcome (% disease)**: In this example, the outcome (infection) is assumed to be rarer in the control arm. However, the number of clusters needed is similar to that in the exemplar (13 versus 12), and the overall sample size has not increased appreciably (1 200 versus 1 300).

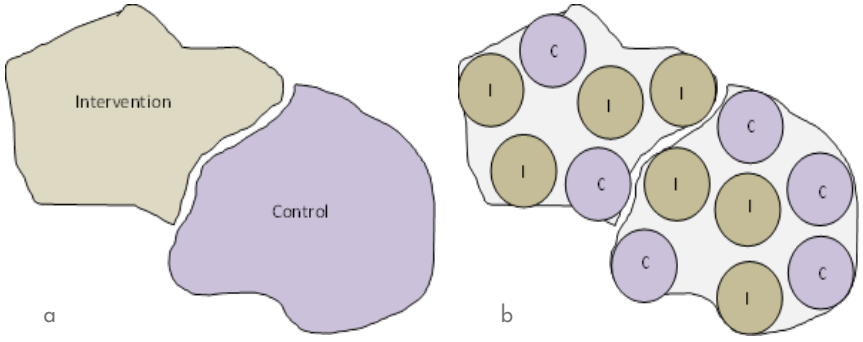
Box 4. How to improve the typical design of two-village pilot studies

Studies comparing two sites (one acting as the control and one receiving the intervention) should not be performed (Fig. 13a). This is a poor design because the intervention effect is completely confounded by the study site. These studies are generally non-randomized, and clearly no sample size calculation has been performed.

These types of studies could be improved through three simple steps.

1. Divide the study area into **smaller clusters**, although not too small.
2. Perform a **sample size calculation** to determine the number of clusters and the population per cluster to measure the outcome with sufficient power (> 80%): at least six per intervention and six per control is recommended.
3. **Randomize these clusters** to either the intervention or control group. (Fig. 13b)

Fig. 13. (a) Typical two-village comparison found in many published studies, and (b) an improved design for a study in same area



In terms of statistical power, it is better for cluster-randomized trials to have a larger number of small clusters than fewer large clusters.

Sample size calculations are relatively simple when the data are analysed with statistical methods such as a comparison of means, using t-tests or an analysis of variance (known as ANOVA), or proportions, using $\chi 2$ tests. However, vector control studies are rarely analysed with such simple methods. Therefore, these standard sample size calculation methods are inadequate when confronted with complex designs, such as CRTs and data from vector control studies, such as count data and data with multiple sources of random variation (for example, within and between study sites). It is strongly recommended that an experienced statistician is consulted to assist with sample size calculations, particularly for CRTs. At a minimum, sample size calculations in protocols should report the information contained in **Annex 3**.

2.5 Step 5: evaluate contamination effects

Contamination (or spillover effects) between different study arms caused by the movement of vectors (*97, 98*) or humans or sharing of the interventions (such as the sharing of insect repellent) between clusters can make interpreting study findings difficult. Spillover that has a conservative effect (that is, it biases results towards the null) can occur through one of two routes. First, community-level effects of the intervention can reduce the transmission intensity in neighbouring control clusters, as occurred in a study using insecticide-treated covers for water jars and window curtains against dengue in Mexico and the Bolivarian Republic of Venezuela (*99*). Second, the movement of people between intervention and control clusters can also dilute the intervention effect because a person's risk of infection is proportional to the amount of time they spend in a treatment area versus out of a treatment area. If the protective effect of an intervention or the sample size of the study is sufficiently large, a positive result can still be demonstrated in a superiority trial, albeit with a reduced intervention effect. However, a negative finding of no difference in such a trial is harder to interpret, and a critical question arises: is the lack of effect due to spillover or due to the absence of efficacy of the new intervention?

A more serious problem arises if the spillover effect is anti-conservative because it exaggerates the difference in outcomes between the intervention and control arms. For example, topical repellents, spatial repellents or structural changes made to houses that have no killing effect on mosquitoes may divert vectors to participants in the control arm of the study who are not using the intervention, putting them at a higher risk of infection than they would have been otherwise (*100–102*).

Hayes and Moulton (*15*) outline a number of methods for reducing contamination, including ensuring that clusters are well separated by using a buffer zone so there is no common boundary between intervention and control clusters, as was done in a larval source management study conducted in the United Republic of Tanzania (*103*), or by using a "fried egg design", in which the intervention and control are administered throughout the cluster, but the outcome is measured only in the central portion of the cluster (*104*). When designing these types of studies it is important to estimate how far the vector is likely to fly in seeking a blood meal or a breeding site, although in practice this is difficult to determine with precision. Records of georeference information for cases that make up the outcome measure can be used to determine whether there were edge effects due to contamination. This technique has been used to estimate the size of area-wide effects in studies of using LLINs for malaria control (*105*). In addition, the travel histories of participants should be collected so that they can be excluded from the per-protocol study analysis (see **section 2.12.1.1**) if they have travelled for significant periods of time and, therefore, spent a relatively brief time being exposed to the intervention (see, for example, reference *38*).

One advantage of an SWCRT design is that any spillover effect is less important because, theoretically, as the intervention clusters coalesce and grow in size, the participants furthest from the edge will be better protected than those on the edge. This analysis requires a statistical model that considers both the position of the participant and the timing of the intervention. Rather than treating contamination effects as a nuisance, it is possible to build an evaluation of their effect into a study's design. This has been tried previously (see, for example, reference *73*), and it is advantageous in that it allows evaluation of the effect of different levels of coverage on outcomes, as well as the dispersal of the vector.

Tackling the problem of human movement in dengue studies is more difficult because the major vectors bite during the day when people are engaged in their daily activities. Some strategies for avoiding this problem would be to use larger cluster areas or to monitor epidemiological outcomes in a sentinel cohort that is less mobile (for example, young children) (*106,107*). Still, even a paediatric cohort study would require a sample size of several thousand to detect a moderate intervention effect (*77*), given that the dengue seroprevalence rate is generally 5–10% in endemic countries (*108*). There are some useful references on this subject (*77,107, 109*), and potential study designs are summarized in **Box 5**. These designs should also be transferrable to other Aedes-borne diseases, including yellow fever, chikungunya and Zika virus disease.

The unintended consequences of using topical repellents or house screening can be minimized by randomizing only a relatively low proportion of individuals or households in a village to receive the intervention (*55,93,*

110, 111). Studies of interventions that can be allocated at the individual level (such as, topical repellents) are probably better allocated at a cluster level (for example, households or, better still, communities) to avoid the intervention being shared among participants in different study arms. Contamination can also be a problem in crossover trials if the washout period is insufficient. Although crossover trials may be suitable when the washout period is short (for example, for a larvicide with a short half-life) (*81*), they should be used with caution in cases in which interventions persist (for example, DDT [dichlorodiphenyltrichloroethane] or habitat manipulation).

Box 5. Studies designed to measure the epidemiological impact of dengue
(adapted from references *77, 107, 109*)

Dengue trials need to take into account several ecological complexities of dengue transmission, which makes designing these trials challenging.

1. Multiple vector species are able to transmit dengue. Although *Aedes aegypti* is the primary urban vector, secondary vectors include *Ae. albopictus* and *Ae. polynesiensis*, which have markedly different ecologies.
2. Dengue transmission is local, focal and heterogeneous. Transmission foci vary in space and time. Foci are connected at short distances by a combination of human and mosquito movement patterns (mosquitoes seldom fly more than 100 m) and at longer distances by human movement alone.
3. *Ae. aegypti* is a day-biting mosquito, so people are at risk of infection at home and when they go about their daily activities.

Trial sites should be chosen where (i) *Ae. aegypti* is the only vector species to simplify the interpretation of the results, and (ii) there are historically high levels of human infection. A cluster-randomized controlled trial can be used with a few adaptations to take into account the specific nature of dengue. Both epidemiological and entomological end points should be measured. Measuring the effect of the intervention on entomological parameters is necessary to confirm that the intervention affected the mosquito population as expected and to assess the relative transmission risks inherent in each cluster.

Baseline virology and vector sampling should be carried out 1 year before the intervention is implemented, and the intervention should last for at least two transmission seasons to ensure there is a sufficient number of cases if there is interannual variation in transmission and mosquito population densities.

The epidemiological effect should be measured using multiple complementary approaches. For example.

- Longitudinal cohort

  To minimize the movement between the treatment and control arms, a paediatric cohort should be used because the participants enter the study either immunologically naive or with monotypic antibody response.

- Surveillance for fever

  Conduct routine surveillance for fever, consisting of one to three visits per household each week, of people living near cohort participants to allow longitudinal comparisons of people with documented dengue illness.

- Geographical clusters

  Conduct geographical cluster studies that screen people living within a designated radius (approximately 100 m) of a person with a laboratory-diagnosed dengue virus infection (the index case) to measure variation in fine-scale spatial patterns of dengue virus transmission.

.../...

Fig. 14. Geographical cluster methods for measuring the epidemiological impact of dengue. The central coloured circle represents the home of an individual with confirmed dengue (red, area with Wolbachia-infected vectors; green, area with uninfected vectors). People living within a 100 m radius (black dots) are screened for concomitant or secondary dengue virus infections (crosses denote homes of additional individuals infected with dengue) (adapted from reference *107*).

The number of participants needed for the study depends on background transmission rates, local herd immunity, movement between clusters, the anticipated effect size of the intervention, between-cluster variation and the logistics capacity. Typically, the sample size will be in the range of 2 000 to 3 000 participants, with at least 5 times that number being under surveillance for fever.

If coverage is not absolute, the individuals or homes within treatment areas that do not accept the intervention can be documented and still provide data on infection status; this can be accounted for in the analysis by calculating the relative direct and indirect effects of the trial (see reference *109* for more information).

Because the *Aedes* vector bites during the day when people are going about their usual activities, there is a high likelihood of contamination (or spillover effects) between clusters. For example, people who live within a cluster assigned to receive the intervention may spend a considerable amount of their day at risk of infection in untreated areas. Conversely, those living in untreated areas may travel to treated areas during the day. It is possible to more accurately estimate the effectiveness of dengue vector control by taking into account human movement and estimating individual-level time under coverage. By incorporating individual-level time under coverage in analyses, the maximum possible effect (as well as the average predicted effects) of the intervention can be estimated (Fig. 14).

Lambrechts et al (*107*) suggest an additional method for measuring the effect of dengue vector control. This involves viral sequence analysis or phylogenetic analysis. At baseline, many lineages of various serotypes would be expected to be circulating. Once an effective intervention is implemented that reduces transmission, a decrease in viral genetic diversity across serotypes would be expected, along with an increase in the average dispersion distances travelled by dengue virus into the intervention area. Although some viruses will continue to be imported into intervention areas by human-mediated dispersal, these viruses will not persist locally, thus reducing the strong spatial clustering that is typically reported in dengue virus phylogenies (Fig. 15).

.../...

Fig. 15. Schematic representing how a vector control intervention might change patterns of viral genetic diversity. Assuming that many lineages of various dengue virus serotypes (coloured circles) co-circulate before the intervention, a reduction in local dengue virus transmission is expected to result in a reduction in viral genetic diversity in the intervention area and a relative increase in the average dispersion distance (*107*).

2.6 Step 6: blinding

Blinding of trial, healthcare providers (or outcome assessors) and researchers to the intervention received by participants can reduce two important forms of bias: performance bias and detection bias. Performance bias occurs when there are systematic differences in the care received by participants in the intervention and control arms. This can be due to differences between the study arms in terms of treatment-seeking, the use of personal protective measures by the participant or to differences in the care received from clinicians. Detection bias occurs when there are systematic differences in how outcomes are assessed between participants in the intervention and control arms. Studies are said to be single blind when the participants are unaware of which treatment they have been assigned, double blind when the study participant and the investigator are both blinded, and triple blind when the study participants, investigator, laboratory staff and those analysing the data are blinded.

Although blinding is highly recommended, it is not always possible. For example, it was not possible to blind study participants in an RCT assessing the efficacy of using house screening to protect against malaria versus no house screening (*93*). The study found that children living in screened homes were less likely to use bed nets than children residing in homes that were unscreened, which may reflect a belief among householders that screening was a substitute for bed nets. However, the effect of performance bias in this study was minimized because bed net use was carefully recorded, and its effect could be adjusted for in the statistical analyses. Alternatively, a study in which there originally was blinding may become unblinded during the study. For example, some participants in an RCT of topical repellents became aware that the placebo lotion they were allocated was not providing protection against mosquito bites and this led to the withdrawal of all households in one village (*55*). This kind of response from participants can introduced attrition bias – that is, systematic differences between those individuals or communities that withdraw from the study versus those that continue. However, it is important to note that even if the participants become unblinded, it is still possible to maintain the blinding of those assessing the outcome and, thus, prevent detection bias.

When possible, participants and investigators should both be blinded to participants' allocation to reduce two important types of bias: performance bias and detection bias.

2.7 Step 7: implement the intervention

In efficacy trials, vector control interventions should be optimally implemented, with attention focused on quality control, high coverage and participants' compliance. Unless these parameters are measured, it is impossible to know whether an observed lack of effect is due to a poorly conducted study, low coverage or compliance, or a lack of efficacy of the vector control method. Quality control checks should be put in place to ensure that interventions, such as IRS, are implemented optimally (that is, the insecticides are used correctly and coverage includes all assigned structures). This can be achieved through accurate record-keeping, random spot checks and supervision (*112,113*). When possible, it is best to collaborate with national control programmes and use their staff to carry out the intervention, following adequate training, with appropriate supervision and oversight from the study's investigators. The implementation team, provided with high-quality training and supervision, will gain valuable experience in carrying out the intervention, which will prove important if the deployment is expanded within the country.

It is important to ensure that participants adhere to the intervention being tested. Efficacy studies usually employ specific community engagement techniques, such as behaviour change communication or information-education-communication strategies, to encourage optimal uptake and use of an intervention when participants' adherence is required. For example, a study by Picado et al. of using LLINs to protect against visceral leishmaniasis evaluated the community's preference for different types of LLINs (for example, in terms of fabric or colour) prior to procuring the nets (*42*). During village meetings, participants were encouraged to use the LLINs properly (for example, there were discussions about how to use them and how often to wash them), and this was reinforced through the distribution of behaviour change communication and information-education-communication materials, such as pictorial diagrams in the local language. Additionally, quarterly house-to-house visits promoted the regular and correct use of LLINs (*42*).

Adherence to an intervention can be measured by using methods such as questionnaires or diaries. However, there is also the potential to introduce bias. For example, courtesy bias may be introduced, whereby there is a tendency for participants to give favourable answers, such as affirming high compliance with an intervention, out of courtesy to the investigator. In some cases, innovative methods need to be identified to assess compliance. For example, an RCT of using topical repellents against malaria measured compliance through self-reporting of use, by estimating the proportion of lotion used from returned bottles, and by conducting "sniff checks", whereby trial staff visited villages at dusk and smelled the arms of participants to check whether the lotion had been applied (*55*). Another example is the use of motion sensors to estimate the use of LLINs by householders in Côte d'Ivoire (*114*).

> Efficacy studies typically use methods such as training and behaviour change communication to ensure high coverage and adherence to an intervention. The coverage and adherence achieved during a study should be measured.

2.8 Step 8: determine how to measure the outcome

The outcome refers to a parameter that the study sets out to measure. Outcomes should be defined in advance of the study being conducted and reflect the question the study sets out to answer. The primary outcome is the outcome of greatest importance, and a secondary outcome is typically an additional effect of an intervention that is of lesser importance or is less optimal for assessing the question asked by the study. It is important to use objective and well-standardized epidemiological and entomological outcomes and to look carefully at the sources of data.

2.8.1 Choosing sources of epidemiological data

In efficacy studies, morbidity data are typically collected using either passive case detection (that is, detection of a case triggered by potential participants attending a healthcare provider because they are sick), active case detection (cases detected by study staff who visit participants and screen them for disease) or during cross-sectional surveys (in which study staff test all or a sample of participants at a fixed time point, typically looking for infection rather than clinical disease).

In some situations, passive case detection can be implemented by using routine data from the health information system. However, the quality of these systems varies substantially between countries and areas and, therefore, the use of routine clinic data is not recommended without making a prior assessment of both case management and reporting practices (for example, the proportion of suspected cases receiving a diagnostic test). More typically, passive case detection is actually what is known as enhanced passive case detection, in which quality control measures are put in place to increase the robustness of the data, such as increasing the supervision or training of health centre staff. This is because statistical power declines as the sensitivity of surveillance decreases. Threats to surveillance sensitivity may include: the failure to identify and report cases, decreased surveillance specificity, or incorrectly reporting a non-case as a case (for example, reporting all children with fever as having malaria), or a combination of these. A good option is to station dedicated, well-trained and well-supervised study nurses in health centres to ensure that patients are tested at a high and consistent rate and to record data accurately and completely. Passive case detection can make it difficult to distinguish between new cases and cases with repeated episodes or who appear more than once owing to treatment failure or non-adherence. Furthermore, passive case detection can considerably reduce the observed intervention effect if the residence of cases is misallocated between clusters of different study arms. This type of misclassification results in a substantial increase in the probability of a type II error.

Active case detection using trained fieldworkers ensures that cases are less likely to be missed. In efficacy studies, nurses or community health workers visit and test cohort members regularly to record the incidence of infection (or disease) in each cluster. Using active case detection in cohorts in each cluster provides high-quality outcome data. Given the importance of having an accurate outcome measure with each incident case assigned to the correct cluster, trials should ensure they have sufficient resources to follow cohorts unless reliable routine reporting systems are in place. Measuring prevalence using cross-sectional surveys requires special effort, which adds to the cost and complexity of the study.

Health and Demographic Surveillance System sites, in which the population is regularly monitored, can also be used to evaluate interventions. However, there are relatively few of these sites and they are probably not representative of the remainder of a country.

2.8.2 Choosing epidemiological outcomes

It is important to use objective and well-standardized epidemiological outcomes, especially when studies do not use blinding. Studies should use standardized, WHO-recommended case definitions for epidemiological outcomes, with parasitological diagnosis or molecular verification (*43–47*). Doing this ensures that the measures used are objective and robust, and the efficacy of the intervention can be compared more easily across different studies.

Mortality is a rare outcome and, therefore, is used less commonly than morbidity in studies because a study would need a very large sample size to find a difference between treatments using mortality as an outcome. For many VBDs, attributing a death to the disease can be difficult, particularly where postmortems are uncommon and individuals may die outside of a hospital setting. However, if there is a good surveillance system, deaths from any cause can be identified with considerable reliability.

2.8.3 Choosing entomological outcomes

In addition to primary epidemiological data, trials must also generate robust, representative entomological data to evaluate the claimed entomological effect and to finalize the TPP for the new product class. These data are also needed to establish reliable correlations between entomological and epidemiological outcomes for a given vector control tool.

Entomological data should be collected in a standardized fashion across study arms and sites, and over time. WHO recommends using automated sampling tools (for example, the United States Centers for Disease Control and Prevention's light trap, sticky traps, or other traps or targets) that do not depend on fieldworkers to collect specimens (such as, human landing catches, aspiration of resting adults, and larval surveys). If fieldworkers collect samples to measure outcomes, collectors should be rotated between sampling sites to reduce the possibility of bias. A number of other techniques can help avoid bias in measurements of entomological outcomes, including separating the field teams that are implementing the intervention from those that are conducting surveillance (see, for example, reference *115*). It is also important to consider the characteristics of the vector caught by each particular trapping method. For example, young tsetse flies are likely to be overrepresented in human bait catches and because older flies are more likely to be infected, the data on infection rate might be affected by this sampling method (*116*). All ecological sampling devices are associated with measurement error because they collect only a sample of the total vector population. However, the findings will not be biased as long as measurement error is consistent across study arms.

Epidemiological and entomological outcome measures should be robust and objective. Diagnostic confirmation of epidemiological outcomes is crucial.

2.8.4 Selecting sites for entomological monitoring

Sampling sites for entomological surveys are often chosen purposively, based on where high vector densities are likely, for example, sites close to suspected larval habitats or for *Triatoma* surveys, houses with unplastered walls or built of wood (*117–119*), or houses closest to mosquito breeding sites. However, these surveys do not measure the average community exposure to infection, and there is the potential for introducing sampling bias if sites are not selected in a representative way across intervention and control sites. Therefore, WHO recommends that sampling sites for entomological surveys be selected randomly.

Generally, it is not feasible to survey every household for entomological outcomes and, therefore, a representative number of households should be selected at random to represent the unit of analysis. For smaller-scale studies (for example, those using the village as the unit of analysis), simple random sampling may be performed to obtain the requisite number of households. For larger-scale studies (those using the district as the unit of analysis), it may be logistically difficult to survey enough houses over a large area. In this case, three to four sentinel sites (that is, villages) per geographical unit could be selected and households could then be randomly selected from within these sites.

If there is likely to be substantial variation within subpopulations, it is also possible to separate the sampling frame into strata and sample from each stratum independently. For example, Joshi and colleagues stratified dwellings into two groups (houses occupied by humans alone and houses occupied by humans and animals) before using simple random sampling to select dwellings in which to measure sandfly density (*120*).

For entomological outcomes, it is generally better to sample regularly in the same set of randomly selected houses or sites throughout a transmission season. This will provide both an estimate of the exposure to vectors experienced by the community and a description of the seasonality of transmission. Repeat sampling from a new random selection of houses or sites at each sampling time point (for example, monthly) may also be considered, but this will take longer to carry out because field staff will need to find new households, provide information about the study and obtain consent.

It is important that sampling sites for entomological surveillance are selected randomly from within the study area of the epidemiological trial.

2.8.5 Duration of follow-up

Many VBDs, such as malaria, can be highly seasonal, but others, such as *Aedes*-borne diseases, are highly variable in space and between years (*109*). Therefore, WHO recommends that studies collect outcome data for at least two transmission seasons or calendar years (if transmission is perennial) following implementation of the intervention. Morbidity data do not need to be collected prior to the intervention for sufficiently powered randomized studies. However, it is good practice to collect data about the characteristics of the participants in the different study arms. For crossover studies, there should be at least one transmission season or calendar year before and after crossover and for the washout period.

For entomological outcomes, follow-up periods need to be sufficiently long and repeat measurements need to be taken to gain a picture of transmission in the area (see, for example, references *121,122*). This is because there is likely to be large variation in vector density between sampling sites and across different sampling periods (night to night, week to week, or over a transmission season) due to environmental factors, such as rainfall. Again, the general principle is that it is important to sample from many clusters, rather than few (see **section 2.4**). At a minimum, there should be at least 6 to 10 households or trap nights per site, as vector density can be highly variable over even short distances. Designs in which entomological sampling is conducted once during the follow-up period are less likely to give reliable results due to the inherent variability in vector populations even when the number of sampling units is high. Therefore, entomological outcomes should be compared as a time series over the course of a season rather than as a simple before-and-after measurement.

Follow-up lasting beyond 2 years may be necessary to assess an intervention's persistence. The duration of follow-up should be long enough to assess the persistence of the intervention (for example, typically 3 years for a new LLIN, according to WHO's requirements (*123*) for the efficacy of LLINs), with capacity to determine efficacy during several follow-up intervals, such as each year of a 3-year trial.

In most cases, studies should measure the outcome for at least two transmission seasons or calendar years following implementation of the intervention, however, this duration may vary depending on the expected residual effect of the intervention.

2.8.6 Loss to follow-up

Some studies suffer from drop out, also called loss to follow-up. This can result in attrition bias if there are systematic differences between study arms in the characteristics of those lost to follow-up. This is a common problem in cohort studies and RCTs that follow participants for extended periods.

If attrition rates are high, this may threaten the validity of a study. There are differing ideas about what level of drop out is acceptable, and these depend on the extent to which data are missing at random or not at random (*124–126*). In general, loss to follow-up of 5% or less is usually of little concern, 5–20% may be a source of bias, and more than 20% loss to follow-up indicates that bias is a concern (*126*). Studies should try to maintain high levels of follow-up when possible. Reports of trials should be clear about levels of loss to follow-up. Loss to follow-up should be reported by study arm (for example, using the CONSORT flow diagram; **Fig.16**) (*16, 62*), as should the baseline characteristics of those who remained in the trial versus those who were lost to follow-up, so that any systematic differences between the groups are made clear.

2.9 Step 9: write the protocol and register the study

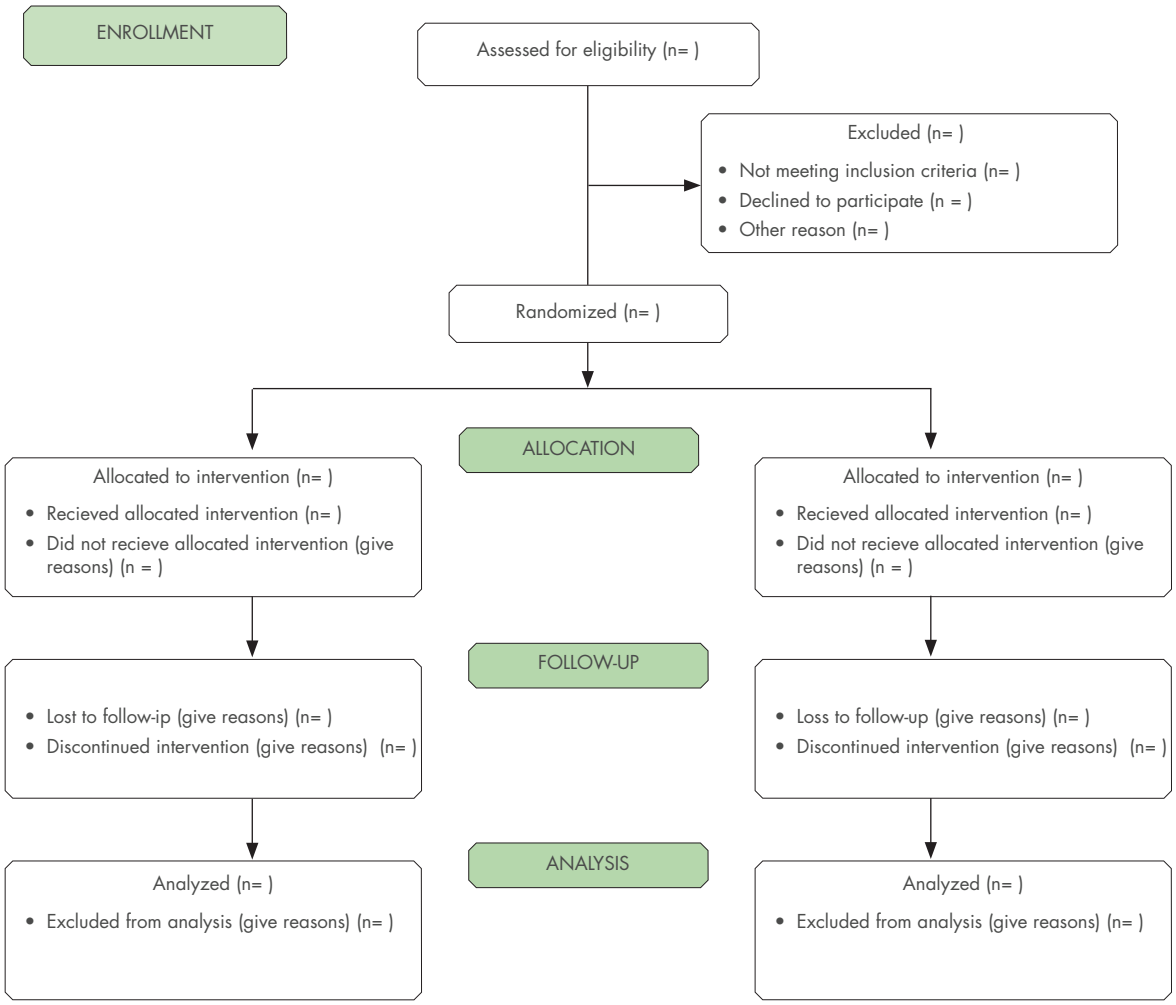
The primary goal of this manual is to guide product trials, with the aim of demonstrating public health value for global policy-development purposes. Therefore, VCAG offers guidance to researchers and product developers, and encourages them to engage early with the group and to develop a concept note for VCAG to review prior to developing the full design of the study and protocol.

Once the study's design has been determined, investigators should write the study protocol, including a description of the sample size calculation and a brief analytical plan. Guidance on what information should be included in the protocol is given in **Annex 4**. This guidance is taken from the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines, with some additions and amendments to ensure the guidelines are applicable to vector control trials (*128, 129*). Alongside the protocol, investigators should also write the information sheets for participants and the forms for informed consent. Standard operating procedures should be produced that outline the study's methods in detail. These can be used for staff training and to provide an aide memoire for field staff, and they will also help to ensure that procedures are standardized.

The trial protocol should be made publicly available, either on the Internet or by publishing it in a journal (for example, in Trials, http://www.trialsjournal.com). This will provide feedback on the protocol through peer review, make the aims of the study explicit to prevent data dredging and post hoc revisions of study aims, reduce the duplication of research effort and make publication bias less likely.

If a vector control study recruits human participants, then the study should be registered with a clinical trials registry, such as ClinicalTrials.gov (http://www.clinicaltrials.gov) or ISRCTN (http://www.isrctn.com/). There may be specific requirements for registering trials imposed by the organization funding the study, the national regulatory authority or the journal that the researchers intend to publish in.

Fig 16 . CONSORT (Consolidated Standards of Reporting Trials) flow diagram (*127*)



2.10 Step 10: ethical review and study oversight

Studies involving interventions against VBDs have unique ethical challenges due to the potential to expose individuals to infected vectors and the need to balance the public health and research benefits of these studies with the potential risk to individuals. Ethical review is an important part of the study procedure, and strong oversight of studies is needed to ensure that trials are conducted according to best practices.

2.10.1 Ethical review

Studies involving humans or that raise ethical issues (e.g., studies involving animals or in which entomological measurements are being taken) must undergo review by an ethics committee. Ethical review helps researchers think through the ethical issues and how to approach them. The risks and benefits for participants and researchers are considered. Ethics committees review research on humans according to principles laid out in the World Medical Association's Declaration of Helsinki – ethical principles for medical research involving human subjects, which was updated in 2013 (*130*), and the Council for International Organizations of Medical Sciences' International ethical guidelines for biomedical research involving human subjects, from 2002 (*131*). The ethics committee should review the study's protocol, as well as the information for participants and the forms for informed consent. Ethical approval should be received before the study starts, and all participants should give informed consent to participate before any study procedures are undertaken.

2.10.2 Study oversight

Studies should have a trial steering committee that is responsible for providing overall supervision of the trial and ensuring that the trial is conducted according to Good Clinical Practice guidelines and standard operating procedures, which should be written before the study starts. In particular, the steering committee should monitor the progress of the trial, adherence to the protocol and participants' safety, and should also consider new information; the steering committee reports formally to the study's sponsor. Membership of the steering committee should be limited, and it generally includes an independent chairperson who is not involved directly in the trial other than as a member of the committee, at least two other independent members, plus the chief investigator and one or two principal investigators, with additional experts if required. The steering committee should meet before the trial starts to approve the final protocol before it goes for ethical review; generally, the steering committee meets annually thereafter.

An additional committee should be established for studies in humans, and that is the data and safety monitoring board or, alternatively, the data monitoring committee. This committee is established by the sponsor to regularly assess the progress of the trial, the safety data and the critical efficacy end points, and to recommend to the sponsor whether to continue, modify or stop a trial. The data monitoring committee is the only body involved in the trial that has access to unblinded (unmasked) comparative data. Members should work completely independently of the chief investigator, steering committee, sponsor, funding organization and host institution.

2.11 Step 11: randomization and study conduct

Randomization is a simple procedure and is highly recommended for all study designs. Study designs that include randomization are critical to demonstrate the public health value of new tools, strategies and approaches for vector control.

2.11.1 The two-step randomization process

Randomization consists of two interrelated steps: sequence generation and allocation concealment. Sequence generation should involve a random component, such as tossing a coin or, more usually, referring to a list of random numbers (computer generated or otherwise). Allocation concealment shields those involved in a trial from knowing what the next treatment allocation will be. "If the investigator can foresee the assignment, selection bias can be introduced, for example by the investigator altering the assignment of individuals or villages to comparison groups based on their perceived risk or other factors. Allocation concealment should not be confused with blinding, and it is particularly important to consider in studies that continually recruit participants and randomization happens on enrolment. For most CRTs for vector control interventions the issue of concealing allocation is less of a concern because all clusters are randomized at the same time; however, appropriate concealment should still be considered in the study design.

In best practice, an independent person performs the randomization to assure that it is not subverted (intentionally or unintentionally) by investigators and to provide independent verification that the investigators remained unaware of the allocation sequence in advance.

Reduces the risk of confounding, if there are sufficient experimental units randomised.

2.11.2 Randomization for cluster-randomized trials

The simplest form of randomization is simple (unrestricted) randomization. In this type of randomization, no constraints are applied when generating the random sequence. If a large number of experimental units is being randomized, then simple randomization will ensure there is balance among confounding factors in the different study arms. In a CRT, however, the number of clusters may not be large enough to ensure balance when using simple randomization. Therefore, three different methods are used to obtain balance between study arms in CRTs: matching, stratified randomization and restricted randomization (*15, 132*).

Matching is when clusters are grouped such that only one cluster in each group is assigned to each study arm. For example, when a study has two arms, matched cluster pairs are formed, or when a study has three arms, matched cluster triplets are formed. When clusters are matched, investigators should perform a matched analysis of the data in which comparisons are made between the intervention and control communities within matched pairs or triplets. Matching is an extreme form of stratification. Typically, groups are matched on the baseline values of the end point of interest or on a surrogate variable that is expected to be correlated with the end point. This could be characteristics of the cluster, for example altitude or geographical location, or cluster-level summaries of individual characteristics, such as socioeconomic status.

In stratified randomization, larger groups of clusters are formed, known as strata, that are similar with respect to the outcome of interest. Clusters within each stratum are then randomly allocated to the treatment arms. Stratification has several advantages over matching. For example, it is possible to look for variations in the intervention effect among strata because there is replication within each stratum, and it is possible to use regression methods to adjust for individual and cluster-level covariates.

Matching and stratification may not achieve an adequate balance among arms when there are several variables that need to be accounted for. In such a case, a suitable approach is restricted randomization. First, a set of conditions is defined for randomization, for example, the difference in baseline prevalence between the arms cannot be more than 10% or clusters within 5 km of a health facility must be equally distributed between the arms. These conditions cannot be so restrictive that only a very small number of allocations can

meet them. Second, a large number of random allocations of the clusters are generated. Third, allocations that do not meet the predefined criteria are discarded. Last, one of the remaining allocations is randomly chosen.

2.11.3 Data capture

Questionnaires are typically used to collect information from participants. Of most relevance to this guidance are quantitative surveys in which numerical or categorical data are collected by asking the same questions of multiple members of the study population. However, it is also possible to use qualitative surveys to investigate the beliefs, attitudes and practices of members of a study population. Questionnaires are used to standardize the way questions are asked and responses recorded, whether on paper or electronically on a mobile device, and analysed later. One example of software available for programming mobile devices is OpenDataKit (https://opendatakit.org/).

For quantitative surveys, most commonly fieldworkers question participants face to face and record their responses. This allows illiterate populations to be surveyed, but this can be time consuming. In literate populations, participants may self-administer the questionnaire. As well as asking questions of participants, it is also possible for the fieldworker to observe and record data, for example, after inspecting water tanks for larvicides or sleeping spaces for the use of LLINs. Participants can be visited once (for a simple cross-sectional survey) or, more likely, on several occasions, for example, to collect information regularly about child morbidity from caregivers during weekly or fortnightly interviews, for which the first interview might be more extensive, with a shorter list of questions asked at each subsequent visit.

It is important to have a good understanding of the community to determine how questions should be phrased, otherwise accurate information may not be elicited. Questionnaires are not simple to design or administer, so researchers should work with someone experienced in designing and administering them or use existing questionnaires that have been validated in similar contexts, for example Demographic and Health Surveys or national censuses. Excellent guidance on designing and piloting studies, selecting and training interviewers, and on administering questionnaires is given in Field trials of health interventions: a toolbox (*133*). Other methods of collecting data include using diaries, in which participants record, for example, when they use a particular vector control intervention.

As well as data from questionnaires, it is likely that samples will be taken for laboratory analysis. Good practice should be followed in labelling and storing samples to ensure that they can be connected to specific participants, and that the household and laboratory data sets can be linked.

2.12 Step 12: data management and statistical analysis

Good data management and appropriate statistical analyses are critical considerations for study design. Key considerations include plans for statistical analysis, data management procedures and outcome measures, and their relation to techniques for statistical analysis.

Even for epidemiologically savvy researchers, it is a good idea to consult with a statistician regarding the sample size calculations, analysis plan and statistical analysis.

2.12.1 Analysis plan

The study protocol should include a brief description of the proposed analysis. However, it is important that a more detailed analytical plan is written before data collection in the field ends and definitely before any data analysis begins. The analysis plan provides a transparent description of the analyses that will be performed to evaluate the study's hypotheses. It also details how the results will be presented and reported, and it explains how problems in the data will be addressed, for example, if data are missing or only partial data are available. Sufficient detail should be provided such that the analysis can be repeated by any competent analyst. When analysing data, patterns often become apparent that require additional analysis that was not pre-specified in the analysis plan. Therefore, these analyses are data driven and should be distinguished from the pre-specified findings, as well as being noted and explained. The items an analysis plan should include are outlined in **Box 6**.

Box 6. Items that should be included in a statistical analysis plan (adapted from reference *136*)

Plans for statistical analyses should contain the following:

- a description of the study's data sources, linkage methods, the intended study population and the study's design, with discussions of strengths and weaknesses;
- effect measures and statistical models used to address each primary and secondary objective, including –
    - justification of the statistical techniques that will be used;
    - information about the distributions assumed for the outcome measures (e.g., normal distributions for biochemical measures, and logistic, Poisson or negative binomials for counts);
    - information about how effect sizes will be presented;
    - methods for addressing confounding in statistical models, including information about the selection of confounders and sensitivity analyses for residual confounding;
    - methods for dealing with clustering and repeated measures;
    - how missing data will be handled, including how it will be reported, methods of imputation and sensitivity analyses for handling missing data;
- formal definitions for any outcomes, for example, cutaneous leishmaniasis will be defined as a person showing clinical signs (skin lesions) who also has parasitological confirmation of the diagnosis using smear or culture from the skin lesion;
- formal definitions of any other variables, for example, the use of an LLIN by a child in the study will be determined by confirming with the caregiver that the child slept under the LLIN at the time of the household survey;
- a description of follow-up and censoring, for example, if study participants travel from the study area and are lost to follow-up or if time-bound censoring is planned for malaria episodes (for example, for 28 days following a previous episode);
- whether the analysis will be by intention to treat or per protocol, or both;
- details on interim analyses, if they will be performed;
- the software package to be used for the statistical analysis

2.12.1.1 Intention-to-treat and per-protocol analyses

Clinical trials of biomedical interventions often use an established two-stage analysis. In an intention-to-treat (ITT) analysis, data from all participants who were randomized to an intervention are used irrespective of whether (i) they received the intervention to which they were allocated and (ii) the participant complied with the protocol. A per-protocol analysis excludes from the analysis (i) those who do not receive the intervention or control to which they were allocated and (ii) individuals who do not comply with the protocol. The primary analysis is usually performed according to ITT, and the secondary analysis is performed per protocol. Because ITT combines two issues – the effectiveness of the intervention and compliance as implemented in the study – it may reduce information (*134*). However, a per-protocol analysis may give an inflated impression of an intervention's efficacy if compliance was low. Often both methods are used, for example, as shown by a study of community mobilization for dengue prevention in which the primary analysis was by ITT and the secondary analysis was per protocol (*135*).

2.12.2 Data managment

If paper forms are used to collect data, then data should be double entered – that is, two data clerks should independently input the data from all of the forms, and any discrepancies should be flagged and resolved by a third person. Data can be captured electronically by using, for example, mobile phones or tablet computers. These devices can be programmed to use logical checks to prevent errors when fieldworkers are inputting data.

2.12.3 Statistical analysis: outcome measures and their relation to techniques for statistical analysis

A trained statistician should be consulted before work starts on the study. The statistical analyses should follow the pre-specified analysis plan (**section 2.12.1**). In addition to the numbers of participants and drop-out rates as specified in the CONSORT flow diagram (**Fig. 16**), the first part of the statistical analysis should comprise for both public health and entomological outcomes, tabulations of the average values and distributions at baseline and in each arm of the trial, and potentially confounding factors, such as age, sex and socioeconomic status. In general, significance tests should not be conducted on baseline data.

Irrespective of the type of study, the incidence of disease (or infection) is typically computed as rates of the number of events in each arm, divided by the sum of the time at risk for the people studied. The rate of disease (or infection) in the intervention group can be compared with that in the control group to obtain a relative measure of effect: the rate ratio. A rate ratio of 1 indicates that there is no difference in the outcome between the control and intervention arms. From these ratio measures of effect, the the protective efficacy (PE) of the intervention can be calculated. This is the percentage reduction in morbidity or mortality in those individuals receiving the vector control intervention compared with that in the control group. It can be calculated as:

$$PE = \left(1 - \frac{\text{rate of disease or infection in the intervention group}}{\text{rate of disease or infection in the control group}}\right) \times 100$$

Similarly, the prevalence of infection (or disease) is measured as the proportion of individuals at risk who are diagnosed with the condition. The PE for such outcomes is measured as $1 - p1/p0$, where $p1$ is the proportion in the intervention arm and $p0$ the proportion in the control arm. There is also an absolute measure of effect, typically the risk difference, which is the risk of the event of interest in the intervention group minus the risk of the event in the control group.

In addition to estimates of efficacy, the analysis must provide confidence intervals for these measures. These interval estimates should account for the different types of correlation in the data. In the case of CRTs, this applies especially to the correlation between different observations in the same cluster, for example, in the household or geographical area. There might also be correlations in the data when the same people are tested during multiple survey rounds.

Statistical methods should be selected based on the distribution and type of data collected, such as, for proportions, random effects logistic regression models; for counts of numbers or events, negative binomial regression or random effects Poisson models; and non-parametric approaches. When the data are in the form of time to event, survival analysis techniques, including frailty terms, may be used. When the data form a time series, a variety of methods are available that account for serial correlations.

Statistical testing of the null hypothesis that the efficacy (or effectiveness) is zero should be carried out using methods such as likelihood ratio tests or permutation tests, depending on the methods used to calculate the confidence intervals. In large studies with a sufficient number of individuals and clusters, the different arms should be balanced for factors such as age, sex, socioeconomic status and entomological exposures, and the initial analysis should not be adjusted for covariates. Statistical models with additional terms for such factors should be used to adjust for covariates in the analysis and give a more accurate estimate of the PE, which may be relevant for understanding the generalizability of the results to other settings.

2.13 Step 13: reporting and disseminating the study

2.13.1 Study reporting

Studies should be reported according to the recognized standards that aim to alleviate problems arising from inadequate reporting, and many journals mandate reporting according to such standards. The CONSORT Statement outlines reporting standards for RCTs (*16, 127*). The Statement includes a checklist of items that should be included in the report and a flow diagram that charts the progress of participants through the trial, from enrolment to allocation, follow-up and the analysis of data (**Fig. 16**). Extensions to the original CONSORT Statement apply to other study designs and interventions, and these have their own specific checklists: the most relevant to vector control are those for CRTs (*62*) and non-inferiority and equivalence trials (*137*). Guidelines are also available that should be followed for the reporting of observational studies, as well as checklists for cross-sectional, case–control and cohort studies (for example, at http://www.strobe-statement.org) (*138*).

Studies should always be reported using standardized requirements for reporting, such as the CONSORT statement.

Standardized forms should be used to report entomological data. Raw entomological data should also be provided, either in the supplementary materials or in an online repository, to aid the synthesis of entomological data from different studies. Data from multiple phase III field trials are often summarized using meta-analysis to calculate a summary effect measure. However, this process is often not possible and valuable data are wasted unless entomological data are reported in full.

2.13.2 Research dissemination

A plan should be developed to disseminate the research. Dissemination should go beyond peer-reviewed publication of the study's findings and presentations at conferences attended by the investigators. First,

findings from the study should be reported to the study's steering committee, the ethics review committees and the organization that funded the research. Second, it is good ethical practice to present findings to the participants and communities in which the study was conducted, for example, through community meetings. Third, regular and social media can be used to engage with the general public, for example, by producing a press release. Last, a study's impact can be increased by engaging with those who may use the research, policymakers and other stakeholders, for example, by conducting meetings with the national vector control programme, WHO and its regional offices, and nongovernmental organizations, or by producing policy briefs or reports.

3. Conclusions

This document summarizes recommendations generated by the WHO Vector Control Advisory Group on how to design and conduct phase III epidemiological field trials to determine the efficacy of vector control tools. Acknowledging that certain study designs are more robust than others, a hierarchy – or ranking – of study designs for evaluating the efficacy of vector control interventions is presented along with key steps to design and conduct such trials for vector control tools. Hierarchies of study designs should be considered along with broader evaluation of the quality of available evidence, such as through the GRADE approach which takes into account factors in addition to study design.

New vector control tools are needed urgently to mitigate threats, such as emerging arboviral pathogens and insecticide resistance in vectors, and to reduce the burden of vector-borne diseases. Yet the general lack of rigorous vector control studies is an obstacle to timely evaluation, uptake and use of new tools, technologies and approaches for vector control. The intention is that this framework for the design and execution of such studies will help in strengthening the evidence base for vector control interventions. This will support the core function of WHO VCAG in assessing the public health value of such interventions and will also assist in the development of evidence based policies and uptake and use of effective new tools to reduce the morbidity and mortality due to vector-borne diseases.

References

1. Watson M. The prevention of malaria in the Federated Malay States. Liverpool: John Murray; 1921.
2. Takken W, Snellen WB, Verhave JP, Knols BGJ, Atmosoedjono S. Environmental measures for malaria control in Indonesia – an historical review on species sanitation. Wageningen: Wageningen Agricultural University Papers; 1990.
3. Soper FL, Wilson DB. *Anopheles gambiae* in Brazil 1930 to 1940. New York (NY): Rockefeller Foundation; 1943.
4. Nájera JA. Malaria control: achievements, problems and strategies. Geneva: World Health Organization; 1999.
5. Nájera JA, González-Silva M, Alonso PL. Some lessons for the future from the Global Malaria Eradication Programme (1955–1969). PLoS Med. 2011;8:e1000412. doi:10.1371/journal.pmed.1000412.
6. Sékétéli A, Adeoye G, Eyamba A, Nnoruka E, Drameh P, Amazigo UV et al. The achievements and challenges of the African Programme for Onchocerciasis Control (APOC). Ann Trop Med Parasitol. 2002;96:S15–28.
7. Alexander B, Maroli M. Control of phlebotomine sandflies. Med Vet Entomol. 2003;17:1–18.
8. Schofield CJ, Dias JC. The Southern Cone Initiative against Chagas disease. Adv Parasitol. 1999;42:1–27.
9. Camargo S. History of *Aedes aegypti* eradication in the Americas. Bull World Health Organ. 1967;36:602–3.
10. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526:207–11. doi:10.1038/nature15535.
11. Indoor residual spraying – use of indoor residual spraying for scaling up global malaria control and elimination [WHO Position Statement]. Geneva: World Health Organization; 2006.
12. Insecticide-treated mosquito nets: a WHO position statement. Geneva: World Health Organization; 2007.
13. WHO recommendations for achieving universal coverage with long-lasting insecticidal nets in malaria control (September 2013 – revised March 2014). Geneva: World Health Organization; 2014 (http://www.who.int/malaria/publications/atoz/who_recommendations_universal_coverage_llins.pdf?ua=1).
14. Wilson AL, Boelaert M, Kleinschmidt I, Pinder M, Scott TW, Tusting L et al. Evidence-based vector control? Improving the quality of vector control trials. Trends Parasitol. 2015;31:380–90. doi:10.1016/j.pt.2015.04.015.
15. Hayes RJ, Moulton LH. Cluster randomised rrials. Boca Raton (FL): Chapman andHall/CRC; 2009.
16. Schulz KF, Altman DG, Moher D, CONSORT Group: CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. PLoS Med. 2010;7:e1000251. doi:10.1371/journal.pmed.1000251.
17. Cook T, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. New York: Houghton Mifflin; 1979.
18. Porta M. A dictionary of epidemiology, 6th edition. New York: Oxford University Press; 2014.
19. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions, version 5.1.0 [updated March 2011]. London (UK): The Cochrane Collaboration; 2011 (http://handbook.cochrane.org, accessed 5 November 2017).
20. Hernán MA, Robins JM. Causal inference. Boca Raton (FL): Chapman and Hall/CRC; 2017 [forthcoming]. (https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/, accessed 5 November 2017).
21. Sachs J, Malaney P. The economic and social burden of malaria. Nature. 2002;415:680–5.
22. Alvar J, Yactayo S, Bern C. Leishmaniasis and poverty. Trends Parasitol. 2006;22:552–7. doi:10.1016/j.pt.2006.09.004.
23. Murray CJL, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384:1005–70. doi:10.1016/S0140-6736(14)60844-8.

24. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380:2197–223. doi:10.1016/S0140-6736(12)61689-4.

25. Le Prince JA, Orenstein AJ. Mosquito control in Panama; the eradication of malaria and yellow fever in Cuba and Panama. New York/London: Putnam; 1916.

26. Webber RH. Eradication of *Wuchereria bancrofti* infection through vector control. Trans R Soc Trop Med Hyg. 1979;73:722–4.

27. Iyengar MOT, De Rook H, Van Dijk WJOM. Interruption of transmission of Anopheles-borne filariasis by indoor residual spraying in Netherlands New Guinea. Trop Geogr Med. 1959;11:287–90.

28. Hashimoto K, Schofield CJ. Elimination of *Rhodnius prolixus* in Central America. Parasit Vectors. 2012;5:45. doi:10.1186/1756-3305-5-45.

29. Handbook for integrated vector management. Geneva: World Health Organization; 2012 (WHO/HTM/NTD/VEM/2012.3; http://apps.who.int/iris/bitstream/10665/44768/1/9789241502801_eng.pdf).

30. D'Souza BJ, Newman RD. Strengthening the policy setting process for global malaria control and elimination. Malar J. 2012;11:28. doi:10.1186/1475-2875-11-28.

31. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J et al. GRADE guidelines: 1. Introduction–GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011, 64:383-394. doi:10.1016/j.jclinepi.2010.04.026.

32. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6. doi:10.1136/bmj.39489.470347.AD.

33. WHO handbook for guideline development, 2nd edition. Geneva: World Health Organization; 2014 (http://apps.who.int/medicinedocs/documents/s22083en/s22083en.pdf).

34. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383:101–4. doi:10.1016/S0140-6736(13)62329-6.

35. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166–75. doi:10.1016/S0140-6736(13)62227-8.

36. Vontas J, Moore S, Kleinschmidt I, Ranson H, Lindsay S, Lengeler C et al. Framework for rapid assessment and adoption of new vector control tools. Trends Parasitol. 2014;30:191–204. doi:10.1016/j.pt.2014.02.005.

37. Pinder M, Jawara M, Jarju LB, Kandeh B, Jeffries D, Lluberas MF et al. To assess whether indoor residual spraying can provide additional protection against clinical malaria over current best practice of long-lasting insecticidal mosquito nets in The Gambia: study protocol for a two-armed cluster-randomised study. Trials. 2011;12:e147. doi:10.1186/1745-6215-12-147.

38. Pinder M, Jawara M, Jarju LBS, Salami K, Jeffries D, Bojang K 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. 2014;S0140–6736:61007–61002. doi:10.1016/S0140-6736(14)61007-2.

39. Picado A, Singh SP, Rijal S, Sundar S, Ostyn B, Chappuis F et al. Longlasting insecticidal nets for prevention of Leishmania donovani infection in India and Nepal: paired cluster randomised trial. BMJ. 2010;341:c6760. doi:10.1136/bmj.c6760.

40. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106. doi:10.1186/1745-6215-12-106.

41. Morrison AC, Minnick SL, Rocha C, Forshey BM, Stoddard ST, Getis A et al. Epidemiology of dengue virus in Iquitos, Peru 1999 to 2005: interepidemic and epidemic patterns of transmission. PLoS Negl Trop Dis. 2010;4:e670. doi:10.1371/journal.pntd.0000670.

42. Picado A, Ostyn B, Rijal S, Sundar S, Singh SP, Chappuis F et al. Long-lasting insecticidal nets to prevent visceral leishmaniasis in the Indian subcontinent; methodological lessons learned from a cluster randomised controlled trial. PLoS Negl Trop Dis. 2015;9:e0003597. doi:10.1371/journal.pntd.0003597.

43. Control of the leishmaniases: report of a meeting of the WHO Expert Committee on the control of leishmaniases, Geneva, 22–26 March 2010. Geneva: World Health Organization; 2010 (WHO Technical Report Series No. 949; http://apps.who.int/iris/bitstream/10665/44412/1/WHO_TRS_949_eng.pdf).

44. Dengue: guidelines for diagnosis, treatment, prevention and control: new edition. Geneva: World Health Organization; 2009.

45. Control of Chagas disease: second report of the WHO Expert Committee. Geneva: World Health Organization; 2002 (WHO Technical Report Series, No. 905; http://apps.who.int/iris/bitstream/10665/42443/1/WHO_TRS_905.pdf).

46. Control and surveillance of human African trypanosomiasis: report of a WHO Expert Committee. Geneva: World Health Organization; 2013 (WHO Technical Report Series No. 984; http://apps.who.int/iris/bitstream/10665/95732/1/9789241209847_eng.pdf).

47. Universal access to malaria diagnostic testing – an operational manual. Geneva: World Health Organization; 2011 (http://apps.who.int/iris/bitstream/10665/44657/1/9789241502092_eng.pdf).

48. Wilson AL, Chen-Hussey V, Logan JG, Lindsay SW. Are topical insect repellents effective against malaria in endemic populations? A systematic review and meta-analysis. Malar J. 2014, 13:446. do:10.1186/1475-2875-13-446.

49. Bowman L, Runge-Ranzinger S, McCall P. Assessing the relationship between vector indices and dengue transmission: a systematic review of the evidence. PLoS Negl Trop Dis. 2014;8:e2848. doi:10.1371/journal.pntd.0002848.

50. Focks D. A review of entomological sampling methods and indicators for dengue vectors. Geneva: UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases; 2004 (http://apps.who.int/iris/bitstream/10665/68575/1/TDR_IDE_DEN_03.1.pdf).

51. Focks DA, Brenner RJ, Hayes J, Daniels E. Transmission thresholds for dengue in terms of Aedes aegypti pupae per person with discussion of their utility in source reduction efforts. Am J Trop Med Hyg. 2000;62:11–18.

52. Focks D, Chadee DD. Pupal survey: an epidemiologically significant surveillance method for *Aedes aegypti*: an example using data from Trinidad. Am J Trop Med Hyg. 1997;56:159–67.

53. Focks D, Alexander N. Multicountry study of Aedes aegypti pupal productivity survey methodology: findings and recommendations. Geneva: World Health Organization; 2006 (https://extranet.who.int/iris/restricted/bitstream/10665/69354/1/TDR_IRM_DEN_06.1_eng.pdf).

54. Snow RW, Rowan KM, Greenwood BM. A trial of permethrin-treated bed nets in the prevention of malaria in Gambian children. Trans R Soc Trop Med Hyg. 1987;81:563–7.

55. Chen-Hussey V, Carneiro I, Keomanila H, Gray R, Bannavong S, Phanalasy S, Lindsay SW. Can topical insect repellents reduce malaria? A cluster-randomised controlled trial of the insect repellent N,N-diethyl-m-toluamide (DEET) in Lao PDR. PLoS ONE. 2013;8:e70664. doi:10.1371/journal.pone.0070664.

56. McGready R, Simpson JA, Htway M, White NJ, Nosten F, Lindsay SW. A double-blind randomized therapeutic trial of insect repellents for the prevention of malaria in pregnancy. Trans R Soc Trop Med Hyg. 2001;95:137–8.

57. Nguyen HT, Tien TV, Tien NC, Ninh TU, Hoa NT. The effect of Olyset net screen to control the vector of dengue fever in Viet Nam. Dengue Bull. 1996;20:87–92 (http://apps.who.int/iris/bitstream/10665/147874/1/dbv20p87.pdf).

58. Vittal M, Limaye LS. Field village scale trial of use of repellent in malaria control. Indian J Med Sci. 1984;38:201–3.

59. Noazin S, Shirzadi MR, Kermanizadeh A, Yaghoobi-Ershadi MR, Sharifi I. Effect of large-scale installation of deltamethrin-impregnated screens and curtains in Bam, a major focus of anthroponotic cutaneous leishmaniasis in Iran. Trans R Soc Trop Med Hyg. 2013;107:444–50. doi:10.1093/trstmh/trt036.

60. Tan AW, Loke SR, Benjamin S, Lee HL, Chooi KH, Sofian-Azirun M. Spray application of Bacillus thuringiensis israelensis (Bti strain AM65-52) against *Aedes aegypti* (L.) and *Ae. albopictus* Skuse populations and impact on dengue transmission in a dengue endemic residential site in Malaysia. Southeast Asian J Trop Med Public Health. 2012;43:296–310.

61. Varnell SP, Murray DM, Baker WL. An evaluation of analysis options for the one-group-per-condition design. Can any of the alternatives overcome the problems inherent in this design? Eval Rev. 2001;25:440–53. doi:10.1177/0193841X0102500402.

62. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 Statement: extension to cluster randomised trials. BMJ. 2012;345:e5661. doi:10.1136/bmj.e5661.

63. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Canberra: Australian Government National Health and Medical Research Council; 2009 (http://citeseerx.ist.psu.edu/viewdodownload;jsessionid=AEFFDA62A5245D6D07F060B56789ED5A?doi=10.1.1.177.4984&rep=rep1&type=pdf).

64. OCEBM levels of evidence. In: Centre for Evidence-Based Medicine [website]. Oxford: Centre for Evidence-Based Medicine; 2017 (http://www.cebm.net/index.aspx?o=5653, accessed 5 November 2017).

65. Corbett MS, Higgins JPT, Woolacott NF. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. Res Synth Methods. 2013;5:79–85. doi:10.1002/jrsm.1090.

66. Hamel MJ, Otieno P, Bayoh N, Kariuki S, Were V, Marwanga D et al. The combination of indoor residual spraying and insecticide-treated nets provides added protection against malaria compared with insecticide-treated nets alone. Am J Trop Med Hyg. 2011;85:1080–6. doi:10.4269/ajtmh.2011.10-0684.

67. Phillips-Howard PA, Nahlen BL, Kolczak MS, Hightower AW, ter Kuile FO, Alaii JA et al. Efficacy of permethrin-treated bed nets in the prevention of mortality in young children in an area of high perennial malaria transmission in western Kenya. Am J Trop Med Hyg. 2003;68:23–9.

68. Magesa SM, Wilkes TJ, Mnzava AE, Njunwa KJ, Myamba J, Kivuyo MD et al. Trial of pyrethroid impregnated bednets in an area of Tanzania holoendemic for malaria. Part 2: Effects on the malaria vector population. Acta Trop. 1991;49:97–108. doi:10.1016/0001-706X(91)90057-Q.

69. Pinder M, Conteh L, Jeffries D, Jones C, Knudsen J, Kandeh B et al. The RooPfs study to assess whether improved housing provides additional protection against clinical malaria over current best practice in The Gambia: study protocol for a randomized controlled study and ancillary studies. Trials. 2016;17:275. doi:10.1186/s13063-016-1400-7.

70. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391. doi:10.1136/bmj.h391.

71. Hargreaves JR, Copas AJ, Beard E, Osrin D, Lewis JJ, Davey C et al. Five questions to consider before conducting a stepped wedge trial. Trials. 2015;16:350. doi:10.1186/s13063-015-0841-8.

72. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials. 2015;16:352. doi:10.1186/s13063-015-0842-7.

73. Silkey M, Homan T, Maire N, Hiscox A, Mukabana R, Takken W et al. Design of trials for interrupting the transmission of endemic pathogens. Trials. 2016;17:278. doi:10.1186/s13063-016-1378-1.

74. Homan T, Hiscox A, Mweresa CK, Masiga D, Mukabana WR, Oria P et al. The effect of mass mosquito trapping on malaria transmission and disease burden (SolarMal): a stepped-wedge cluster-randomised trial. Lancet. 2016;388:1193–201. doi:10.1016/S0140-6736(16)30445-7.

75. Tiono AB, Pinder M, N'Fale S, Faragher B, Smith T, Silkey M et al. The AvecNet Trial to assess whether addition of pyriproxyfen, an insect juvenile hormone mimic, to long-lasting insecticidal mosquito nets provides additional protection against clinical malaria over current best practice in an area with pyrethroid-resistant vectors in rural Burkina Faso: study protocol for a randomised controlled trial. Trials. 2015;16:113. doi:10.1186/s13063-015-0606-4.

76. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clinical Trials. 2007;28:182–91. doi:10.1016/j.cct.2006.05.007.

77. Wolbers M, Kleinschmidt I, Simmons CP, Donnelly CA. Considerations in the design of clinical trials to test novel entomological approaches to dengue control. PLoS Negl Trop Dis.. 2012;6:e1937. doi:10.1371/journal.pntd.0001937.

78. Kramer RA ML, Senkoro K, Lesser A, Shayo EH, Paul CJ, Miranda ML. A randomized longitudinal factorial design to assess malaria vector control and disease management interventions in rural Tanzania. Int J Environ Res Public Health. 2014;11:5317–332. doi:10.3390/ijerph110505317.

79. Torr SJ, Hargrove JW, Vale GA. Towards a rational policy for dealing with tsetse. Trends Parasitol. 2005;21:537–41. doi:10.1016/j.pt.2005.08.021.

80. Walshe DP GP, Abdel-Hameed Adeel AA, Pyke GH, Burkot T. Larvivorous fish for preventing malaria transmission. Cochrane Database Syst Rev. 2013:CD008090. doi:10.1002/14651858.CD008090.pub2.

81. Majambere S, Pinder M, Fillinger U, Ameh D, Conway DJ, Green C et al. Is mosquito larval source management appropriate for reducing malaria in areas of extensive flooding in The Gambia? A cross-over intervention trial. Am J Trop Med Hyg. 2010;82:176–84. doi:10.4269/ajtmh.2010.09-0373.

82. Cohen D, Green M, Block C, Slepon R, Ambar R, Wasserman SS et al. Reduction of transmission of shigellosis by control of houseflies (Musca domestica). Lancet. 1991;337:993–7. doi:10.1016/0140-6736(91)92657-N.

83. Njie M, Dilger E, Lindsay SW, Kirby MJ. Importance of eaves to house entry by anopheline, but not culicine, mosquitoes. J Med Entomol. 2009;46:505–10.

84. Rowland M, Durrani N, Kenward M, Mohammed N, Urahman H, Hewitt S. Control of malaria in Pakistan by applying deltamethrin insecticide to cattle: a community-randomised trial. Lancet;357:1837–41. doi:10.1016/S0140-6736(00)04955-2.

85. Raphael K. Recall bias: a proposal for assessment and control. Int J Epidemiol. 1987;16:167–70.

86. Kesteman T, Randrianarivelojosia M, Raharimanga V, Randrianasolo L, Piola P, Rogier C. Effectiveness of malaria control interventions in Madagascar: a nationwide case-control survey. Malaria J. 2016;15:83. doi:10.1186/s12936-016-1132.

87. Mathanga DP, Mwandama DA, Bauleni A, Chisaka J, Shah MP, Landman KZ et al. The effectiveness of long-lasting, insecticide-treated nets in a setting of pyrethroid resistance: a case–control study among febrile children 6 to 59 months of age in Machinga District, Malawi. Malar J. 2015;14:457. doi:10.1186/s12936-015-0961-3.

88. De Serres G, Skowronski DM, Wu XW, Ambrose CS. The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. Euro Surveill. 2013;18:20585.

89. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health. 2003;57:527–9. doi:10.1136/jech.57.7.527.

90. Kirkwood BR, Sterne JAC. Essential medical statistics, 2nd edition. Oxford: Blackwell Science; 2003.

91. Suresh KP, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci. 2012;5:7–13.

92. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28:319–26. doi:10.1093/ije/28.2.319.

93. Kirby MJ, Ameh D, Bottomley C, Green C, Jawara M, Milligan PJ et al. Effect of two different house screening interventions on exposure to malaria vectors and on anaemia in children in The Gambia: a randomised controlled trial. Lancet. 2009;374:998–1009. doi:10.1016/S0140-6736(09)60871-0.

94. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004;94:423–32. PMCID:PMC1448268.

95. Hooper R, Teerenstra S, de Hoop E, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Stat Med. 2016, 35:4718–28. doi:10.1002/sim.7028.

96. Lindsay SW, Shenton FC, Snow RW, Greenwood BM. Responses of Anopheles gambiae complex mosquitoes to the use of untreated bednets in The Gambia. Med Vet Entomol. 1989;3:253–62. doi:10.1111/j.1365-2915.1989.tb00225.

97. Lindsay SW, Alonso PL, Armstrong Schellenberg JR, Hemingway J, Adiamah JH, Shenton FC et al. A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia, west Africa. 7: Impact of permethrin-impregnated bed nets on malaria vectors. Trans R Soc Trop Med Hyg. 1993;87 (Suppl. 2):45–51.

98. Quiñones ML, Lines J, Thomson MC, Jawara M, Greenwood BM. Permethrin-treated bed nets do not have a 'mass-killing effect' on village populations of Anopheles gambiae s.l. in The Gambia. Trans R Soc Trop Med Hyg. 1998;92:373–8. doi:10.1016/S0035-9203(98)91053-7.

99. Kroeger A, Lenhart A, Ochoa M, Villegas E, Levy M, Alexander N et al. Effective control of dengue vectors with curtains and water container covers treated with insecticide in Mexico and Venezuela: cluster randomised trials. BMJ. 2006;332:1247–50. doi:10.1136/bmj.332.7552.1247.

100. Maia MF, Onyango SP, Thele M, Simfukwe ET, Turner EL, Moore SJ. Do topical repellents divert mosquitoes within a community? Health equity implications of topical repellents as a mosquito bite prevention tool. PLoS ONE. 2013;8:e84875. doi:10.1371/journal.pone.0084875.

101. Moore SJ, Davies CR, Hill N, Cameron MM. Are mosquitoes diverted from repellent-using individuals to non-users? Results of a field study in Bolivia. Trop Med Int Health. 2007;12:532–9. doi:10.1111/j.1365-3156.2006.01811.

102. Maia MF, Kreppel K, Mbeyela E, Roman D, Mayagaya V, Lobo NF et al. A crossover study to evaluate the diversion of malaria vectors in a community with incomplete coverage of spatial repellents in the Kilombero Valley, Tanzania. Parasit Vectors. 2016;9:451. doi:10.1186/s13071-016-1738-4.

103. Bang YH, Sabuni IB, Tonn RJ. Integrated control of urban mosquitoes in Dar es Salaam using community sanitation supplemented by larviciding. East Afr Med J. 1975;52:578–8.

104. West PA, Protopopoff N, Wright A, Kivaju Z, Tigererwa R, Mosha FW et al. Indoor residual spraying in combination with insecticide-treated nets compared to insecticide-treated nets alone for protection against malaria: a cluster randomised trial in Tanzania. PLoS Med. 2014;11:e1001630. doi:10.1371/journal.pmed.1001630.

105. Hawley WA, Phillips-Howard PA, Ter Kuile F, Terlouw DJ, Vulule JM, Ombok M et al. Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. Am J Trop Med Hyg. 2003;68 (Suppl. 4):121–7.

106. Tsunoda T, Kawada H, Huynh TT, Luu LL, Le SH, Tran HN et al. Field trial on a novel control method for the dengue vector, *Aedes aegypti* by the systematic use of Olyset Net and pyriproxyfen in Southern Vietnam. Parasit Vectors. 2013;6:6. doi:10.1186/1756-3305-6-6.

107. Lambrechts L, Ferguson NM, Harris E, Holmes EC, McGraw EA, O'Neill SL et al. Assessing the epidemiological effect of Wolbachia for dengue control. Lancet Infect Dis. 2015;15:862–6. doi:10.1016/S1473-3099(15)00091-2.

108. Endy TP, Yoon IK, Mammen MP. Prospective cohort studies of dengue viral transmission and severity of disease. Curr Top Microbiol Immunol. 2010;338:1–13. doi:10.1007/978-3-642-02215-9_1.

109. Reiner RJ, Achee N, Barrera R, Burkot TR, Chadee DD, Devine G et al. Quantifying the epidemiological impact of vector control on dengue. PLoS Negl Trop Dis 10: e0004588. doi:10.1371/journal.pntd.0004588.

110. Rowland M, Downey G, Rab A, Freeman T, Mohammad N, Rehman H et al. DEET mosquito repellent provides personal protection against malaria: a household randomized trial in an Afghan refugee camp in Pakistan. Trop Med Int Health. 2004;9:335–42. doi:10.1111/j.1365-3156.2004.01198.

111. Sangoro O, Turner E, Simfukwe E, Miller JE, Moore SJ. 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. doi:10.1186/1475-2875-13-324.

112. Indoor residual spraying: an operational manual for IRS for malaria transmission, control and elimination, 2nd edition. Geneva: World Health Organization; 2013 (http://apps.who.int/iris/bitstream/10665/177242/1/9789241508940_eng.pdf).

113. Larval source management – a supplementary measure for malaria vector control: an operational manual. Geneva: World Health Organization; 2013 (http://apps.who.int/iris/bitstream/10665/85379/1/9789241505604_eng.pdf).

114. Koudou BG, Malone D, Hemingway J. The use of motion detectors to estimate net usage by householders, in relation to mosquito density in central Côte d'Ivoire: preliminary results. Parasit Vectors. 2014;7:96. doi:10.1186/1756-3305-7-96.

115. Fillinger U, Kannady K, William G, Vanek MJ, Dongus S, Nyika D et al. A tool box for operational mosquito larval control: preliminary results and early lessons from the Urban Malaria Control Programme in Dar es Salaam, Tanzania. Malar J. 2008;7:20. doi:10.1186/1475-2875-7-20.

116. Tsetse situation : sampling for tsetse. In: Murray M, Trail JCM, Turner DA, Wissocq Y, editors. Livestock productivity and trypanotolerance – Network Training Manual. Addis Ababa: International Livestock Centre for Africa; 1983 (https://ilriorg/InfoServ/Webpub/fulldocs/LivProd/Tochtm#TopOfPage, accessed 5 November 2917).

117. Sandoval-Ruiz CA, Guevara R, Ibáñez-Bernal S. Household risk factors associated to infestation of Triatoma dimidiata, the Chagas disease vector in Central Region of Veracruz, Mexico. Salud Pública Méx. 2014;56:213–20.

118. Fillinger U, Ndenga B, Githeko A, Lindsay SW. Integrated malaria vector control with microbial larvicides and insecticide-treated nets in western Kenya: a controlled trial. Bull World Health Organ. 2009;87:655–65. doi:10.2471/BLT.08.055632.

119. Matthews GA, Nkot PB, Tanna S, Dobson H, Bywater A. Comparison of indoor residual spraying and insecticide-treated bed nets used alone and in combination for mosquito control. Asp Appl Biol. 2010;99:165–70.

120. Joshi AB, Das ML, Akhter S, Chowdhury R, Mondal D, Kumar V et al. Chemical and environmental vector control as a contribution to the elimination of visceral leishmaniasis on the Indian subcontinent: cluster-randomized controlled trials in Bangladesh, India and Nepal. BMC Med. 2009;7:54. doi:10.1186/1741-7015-7-54.

121. Dutta P, Khan AM, Khan SA, Borah J, Sharma CK, Mahanta J. Malaria control in a forest fringe area of Assam, India: a pilot study. Trans R Soc Trop Med Hyg. 2011;105:327–32. doi:10.1016/j.trstmh.2011.02.008.

122. Karch S, Asidi N, Manzambi ZM, Salaun JJ. Efficacy of Bacillus sphaericus against the malaria vector Anopheles gambiae and other mosquitoes in swamps and rice fields in Zaire. J Am Mosq Control Assoc. 1992;8:376–80.

123. Guidelines for laboratory and field testing of long-lasting insecticidal nets. Geneva: World Health Organization; 2005 (WHO/CDS/WHOPES/GCDPP/2005.11; http://apps.who.int/iris/bitstream/10665/69007/1/WHO_CDS_WHOPES_GCDPP_2005.11.pdf).

124. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child. 2008;93:458–61. doi:10.1136/adc.2007.127316.

125. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19:751–60.

126. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ. 2002;325:652–4. doi:10.1136/bmj.325.7365.652.

127. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869. doi:10.1136/bmj.c869.

128. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586. doi:10.1136/bmj.e7586.

129. Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jeric K et al. SPIRIT 2013 Statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158:200–7. doi:10.7326/0003-4819-158-3-201302050-00583.

130. WMA Declaration of Helsinki – ethical principles for medical research involving human subjects. Ferney-Voltaire: World Medical Association; 2013 (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, accessed 5 November 2017).

131. International ethical guidelines for biomedical research involving human subjects. Geneva: Council for International Organizations of Medical Sciences, World Health Organization; 2002 (http://www.fercap-sidcer.org/publications/pdf/201202/FERCAP-18-CIOMS%20Ethical%20Guidelines.pdf).

132. Ivers NM, Halperin IJ, Barnsley J, Grimshaw JM, Shah BR, Tu K, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. Trials. 2012;13:120. doi:10.1186/1745-6215-13-120.

133. Smith PG, Morrow R, Ross D. Field trials of health interventions: a toolbox. Oxford: Oxford University Press; 2015 (http://www.oapen.org/search?identifier=569923;keyword=field%20trials%20of%20health%20interventions, accessed 5 November 2017).

134. Feinman RD. Intention-to-treat. What is the question? Nutr Metab (Lond). 2009;6:1. doi:10.1186/1743-7075-6-1.

135. Andersson N, Nava-Aguilera E, Arosteguí J, Morales-Perez A, Suazo-Laguna H, Legorreta-Soberanis J et al. Evidence based community mobilization for dengue prevention in Nicaragua and Mexico (Camino Verde, the Green Way): cluster randomized controlled trial. BMJ. 2015;351:h3267. doi:10.1136/bmj.h3267.

136. The statistical analysis plan. In: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP): Guide on Methodological Standards in Pharmacoepidemiology (Revision 4). London: European Medicines Agency; 2010 (EMA/95098/2010,).

137. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012;308:2594–604. doi:10.1001/jama.2012.87802.

138. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–7. doi:10.1016/S0140-6736(07)61602-X.

139. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401–6. doi:10.1016/j.jclinepi.2010.07.015.

140. Guidance: specific questions that need answering when considering the design of clinical trials. London: NHS Health Research Authority; 2014 (http://www.hra.nhs.uk/documents/2014/05/guidance-questions-considerations-clinical-trials.pdf).

141. Clark T, Davies H, Mansmann U. Five questions that need answering when considering the design of clinical trials. Trials. 2014;15:286. doi:10.1186/1745-6215-15-286.

Annexes

Annex 1. Suggested further reading

General trial design and conduct	Smith PG, Morrow R, Ross D. Field trials of health interventions: a toolbox, 3rd edition. Oxford: Oxford University Press; 2015 (http://www.oapen.org/search?identifier=569923;keyword=field%20trials%20of%20health%20interventions, accessed 5 November 2017) Matthews JNS. Introduction to randomized controlled clinical trials, 2nd edition. Boca Raton (FL): Chapman and Hall/CRC; 2006. Domanski MJ, McKinlay S. Successful randomised trials: a handbook for the 21st century. Philadelphia: Lippincott Williams & Wilkins; 2009. ICH harmonised tripartite guideline: guideline for good clinical practice E6(R1). Geneva: International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1996 (https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf).
Protocol writing	Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jeric K et al. SPIRIT 2013 Statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158:200–7. doi:10.7326/0003-4819-158-3-201302050-00583. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586. doi:10.1136/bmj.e7586.
Randomization	Vickers AJ. How to randomize. J Soc Integr Oncol. 2006;4:194–8. PMCID:PMC2596474.
General statistical analysis	Kirkwood BR, Sterne JAC. Essential medical statistics, 2nd edition. Oxford: Blackwell Science; 2003. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991. IICH harmonised tripartite guideline: statistical principles for clinical trials E9. Geneva: International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1998 (http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf).
Generalized linear mixed models	Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol. 2009;24:127–35. doi:10.1016/j.tree.2008.10.008.
Time series and interrupted time series (quasi-experimental designs)	Cook T, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. New York: Houghton Mifflin; 1979. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin; 2002.
Step-wedge designs	Hargreaves JR, Copas AJ, Beard E, Osrin D, Lewis JJ, Davey C et al. Five questions to consider before conducting a stepped wedge trial. Trials. 2015;16:350. doi:10.1186/s13063-015-0841-8. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials. 2015;16:352. doi:10.1186/s13063-015-0842-7. Prost A, Binik A, Abubakar I, Roy A, De Allegri M, Mouchoux C et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. Trials. 2015;16:351. doi:10.1186/s13063-015-0837-4. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391. doi:10.1136/bmj.h391.
Crossover trials	Jones B, Kenward MG. Design and analysis of cross-over trials, 2nd edition. London: Chapman and Hall; 2003. Senn S. Cross-over trials in clinical research, 2nd edition. Chichester: Wiley; 2002.

Sample size calculations	Lwanga SK, Lemeshow S. Sample size determination in health studies : a practical manual. Geneva: World Health Organization; 1991 (http://apps.who.int/iris/handle/10665/40062, accessed 5 November 2017). Chow S-C, Wang H, Shao J. Sample size calculations in clinical research. Boca Raton (FL): Chapman and Hall/CRC; 2007. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28:319–26. doi:10.1093/ije/28.2.319. Johnson PCD, Barry SJE, Ferguson HM, Müller P. Power analysis for generalized linear mixed models in ecology and evolution. Methods Ecol Evol. 2015;6:133–42. doi:10.1111/2041-210X.12306.
Sample size calculations for step-wedge designs	Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. J Clin Epidemiol. 2013;66:752–8. doi:10.1016/j.jclinepi.2013.01.009. Hooper R, Teerenstra S, de Hoop E, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Stat Med. 2016;35:4718–28. doi:10.1002/sim.7028. Hemming K, Girling A. A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials. Stata J. 2014;14:363–80. Baio G, Copas AJ, Ambler G, Hargreaves JR, Beard E, Omar RZ. Sample size calculation for a stepped wedge trial. Trials. 2015;16:354. doi:10.1186/s13063-015-0840-9. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials. 2007;28:182–91. doi:10.1016/j.cct.2006.05.007.
Cluster-randomized controlled trials	Hayes RJ, Moulton LH. Cluster randomised trials. Boca Raton (FL): Chapman and Hall/CRC; 2009.
Dengue trial design	Lambrechts L, Ferguson NM, Harris E, Holmes EC, McGraw EA, O'Neill SL et al. Assessing the epidemiological effect of Wolbachia for dengue control. Lancet Infect. Dis. 2015:15;862–6. doi:10.1016/S1473-3099(15)00091-2. Wolbers M, Kleinschmidt I, Simmons CP, Donnelly CA. Considerations in the design of clinical trials to test novel entomological approaches to dengue control. PLoS Negl Trop Dis. 2012;6:e1937. doi:10.1371/journal.pntd.0001937. Reiner RJ, Achee N, Barrera R, Burkot TR, Chadee DD, Devine G, et al. Quantifying the epidemiological impact of vector control on dengue. PLoS Negl Trop Dis. 2016;10:e0004588. doi:10.1371/journal.pntd.0004588.
Restricted randomization for cluster-randomized trials	Ivers NM, Halperin IJ, Barnsley J, Grimshaw JM, Shah BR, Tu K et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. Trials. 2012;13:120. doi:10.1186/1745-6215-13-120.
Study reporting	Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. PLoS Med. 2010;7:e1000251. doi:10.1371/journal.pmed.1000251. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869. doi:10.1136/bmj.c869. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012;308:2594–604. doi:10.1001/jama.2012.87802. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–7. doi:10.1016/S0140-6736(07)61602-X.
Research dissemination	Implementation research toolkit. Geneva: World Health Organization; 2014 (http://www.who.int/tdr/publications/topics/TDR-brochure.pdf).

Annex 2. Summary of the GRADE approach to rating the quality of evidence (adapted from reference 139)

Study design	Initial rating of quality of evidence
Randomized controlled trial	High (4+)
Observational study	Low (2+)

Downgraded if	Number of points downgraded by
Risk of bias Bias refers to an overestimate or underestimate of the effect of an intervention. Features of the study's design and how the study is conducted, such as a lack of blinding, can introduce bias.	−1 if serious −2 if very serious
Inconsistency Studies will have slightly different results due to random error and chance. However, inconsistency is present when results from different studies are heterogeneous even after exploration of factors that might explain the differences.	−1 if serious −2 if very serious
Indirectness Indirectness occurs when the evidence from a study differs from the evidence of interest in terms of participants, the intervention, the comparator or the outcomes, or a combination of these.	−1 if serious −2 if very serious
Imprecision This refers to (i) whether the result is statistically significant, (ii) whether the intervention would be considered to have public health value if the actual result was the upper or lower limit of the 95% confidence interval of the effect estimate, and (iii) the power of the studies and the overall meta-analysis.	−1 if serious −2 if very serious
Publication bias This refers to bias introduced in a meta-analysis because studies that show a benefit of the intervention are more likely to be published than studies that do not.	−1 if likely −2 if very likely

Upgraded if	Number of points upgraded by
Effect size The effect size refers to the magnitude of difference between the treatment and control groups.	+1 if large +2 if very large
Dose response This refers to an increasing or decreasing trend in the outcome in relation to increasing levels of exposure.	+1 if evidence of a dose response gradient
Plausible residual confounding Confounders are factors associated with both the exposure and the outcome that are not on the causal pathway. Evidence can be upgraded if plausible residual confounding would further support inferences regarding the treatment effect, that is, there is negative confounding.	+1 if would reduce a demonstrated effect +1 if would suggest a spurious effect if no effect was observed

Final GRADE assessment
High (4+)
Moderate (3+)
Low (2+)
Very low (1+)

GRADE, Grading of Recommendations Assessment, Development and Evaluation.

Annex 3. Items to include in a sample size calculation (adapted from references *140, 141*)

Item	Comment
Explain what the study aims to show	What is the research question?
Describe the design of the study	
State clearly the primary outcome measure(s)	
State the test procedure on which the sample size is based	Researchers need to specify the statistical test used to calculate the sample size NB: The way the sample size is calculated is determined by the way the data will be analysed. Researchers should begin by thinking of the analysis that will ultimately be performed on the primary outcome measure to ensure that the sample size is calculated appropriately.
State the allocation ratio	The allocation ratio is the ratio of the number of participants allocated to different treatment groups, for example, having equal numbers in both the treatment and control groups is expressed as 1:1; twice as many patients in the treatment group as in the control group is expressed as a ratio of 2:1.
If the study aims to show superiority, state and justify the difference sought between the two arms	In a superiority trial, the objective is to demonstrate that the test treatment is better than the comparator. The research protocol should clearly describe the treatment difference sought, which should be clinically relevant and realistic. This may also be referred to as the minimum clinically important difference – that is, the difference one would not like to miss.
If the study aims to demonstrate non-inferiority or equivalence, state and justify the acceptance margin	Researchers should give the acceptance margin that represents the largest difference that is clinically acceptable to patients.
Report all parameters used in the sample size calculation	Depending on the design of the study, researchers will need to mention one or more of the following. • Continuous data: This refers to the variance in each treatment group. Variance is often assumed to be equal in each group. Effect size (difference in mean between control and intervention). • Binary data: This includes data such as the proportion of participants in the control group for whom the treatment was a success versus a failure, who are alive versus dead, with and without symptoms. Effect size (difference in proportion between control and intervention). • Survival data (time to event): This is the rate or mean or median survival time in the control group. Effect size (difference in rate between control and intervention).
Reporting a study with time-to-event outcome measures	For these studies, the number of events is important, and a sufficient number of participants should be recruited to ensure a sufficient number of events. The number of participants depends on the accrual period (that is, the period during which participants are recruited and enrolled) and the duration of the study (that is, the accrual period plus the follow-up period).
Reporting a group sequential design	If the study has a group sequential design, then the number of interim analyses (also known as interim looks), stopping rules and the survival time assumption (whether exponential survival time or proportional hazards) should be reported. Group sequential designs involve multiple looks at accumulating data. These designs should account for multiple testing.
Reporting a factorial design	A factorial design should have sufficient power to test for interaction. This should be reported in the sample size determination.
Reporting a cluster design	Researchers should report the estimated coefficient of variation (k) or intracluster correlation and state how it was estimated or from where it was obtained.

Item	Comment
Explain the rationale for the parameters used in the sample size calculation	A summary should outline how the assumptions used in the sample size calculation were chosen and why they are considered plausible for the planned study. It may be useful to take into account data from other studies when choosing parameters, bearing in mind that studies may have important differences, such as the population evaluated.
Describe any procedures used to re-estimate the sample size during the study	If there is uncertainty about the assumptions used in the sample size calculation, it may be prudent to check their validity using interim data. If researchers will be reviewing the sample size during their study, they should include a description of the re-estimation procedure in the protocol.
Report the type I error	A type I error is the incorrect rejection of a true null hypothesis (that is, a false positive). Specify whether the type I error is one- or two-sided. Often, one-sided tests are not justifiable.
Report the type II error	A type II error is the failure to reject a false null hypothesis (that is, a false negative).
If the study has multiple end points, describe any adjustments made for multiple testing (multiplicity), interim analyses or multiple study arms	Adjusting for multiplicity may be required if there are multiple outcome measures, comparisons of multiple treatment arms, and multiple looks at accumulating data during interim monitoring (for example, as in group sequential studies). Such multiple comparisons increase the risk of a type I error. The strategy for controlling the type I error rate should be described. If researchers do not intend to adjust for multiple comparisons or looks, or both, then this should also be reported, together with a brief explanation of why adjustment is not necessary.
State the number of patients or events required for the analysis	The number of patients required for the analysis is the number obtained from the sample size calculation.
Explain the allowance (if any) for drop outs (loss to follow-up)	Researchers should state the proportion of study participants expected to be lost to follow-up.
State the total number of patients that will be enrolled	This is the total sample size required – that is, the total number of patients to be enrolled after adjusting for drop outs.

Annex 4. Items that should be included in a clinical trial protocol and related documents (adapted from the SPIRIT 2013 checklist, references *128, 129*)

Section and item	Item No.	Description
ADMINISTRATIVE INFORMATION		
Title	1	Descriptive title identifying the study's design, population, interventions and, if applicable, trial acronym
Trial registration	2a	Trial identifier and registry name; if not yet registered, name the intended registry
	2a	All items from WHO's Trial Registration Data Set (http://www.who.int/ictrp/network/trds/en/)
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material and other support
Roles and responsibilities	5a	Names, affiliations and roles of those who contributed to the protocol
	5b	Name and contact information of the trial's sponsor
	5c	Role of the study's sponsor and organization that is funding the study, if any, in designing the study; collecting, managing, analysing and interpreting the data; writing the report; and in the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities
	5d	Composition, roles and responsibilities of the study team, the trial steering committee, the data and safety monitoring board (or data monitoring committee) and other individuals or groups overseeing the trial, if applicable (see item 21a for additional information about the data monitoring committee)
INTRODUCTION		
Background and rationale	6a	Description of the research question and justification for undertaking the trial, including a summary of relevant studies (published and unpublished) examining the benefits and harms of each intervention
	6b	Explanation of the choice of comparators
Objectives	7	Specific objectives or hypotheses
Trial design	8	Description of the trial's design, including the type of trial (for example, parallel group, crossover, factorial), level of allocation (for example, individual or cluster) and framework (for example, superiority, noninferiority)
METHODS: PARTICIPANTS, INTERVENTIONS AND OUTCOMES		
Study setting	9	Description of study's setting(s) (such as, the country, region, latitude and longitude, VBD endemicity or seasonality, primary and secondary vectors, coverage of standard practice vector control interventions)
Eligibility criteria	10	Inclusion and exclusion criteria for participants
Interventions	11a	Interventions for each group, with sufficient detail to allow replication (for example, the dose, the manufacturer), including how and when the intervention will be administered; also explain which vector control interventions the control group will receive or use
	11b	Discussion of strategies to improve adherence to intervention protocols (for example the use of "hang up campaigns" for LLINs, behaviour change campaigns to reduce sources of infection or disease, staff training) and the procedures that will be used to monitor adherence (such as, home visits to verify LLIN use) or coverage (such as, measuring insecticide on walls)

Section and item	Item No.	Description
Outcomes	12	Primary, secondary and other outcomes, including the specific measurement variable, method of measurement, analysis metric (for example, change from baseline, final value, time to event), method of aggregation (for example, median, proportion) and time point for each outcome Be sure to explain how and where entomological outcomes will be measured, including how sampling sites will be selected
Participant timeline	13	Schedule for enrolment, implementation of interventions, assessments and visits for participants; a schematic diagram is highly recommended
Sample size	14	Estimated number of participants needed to achieve study objectives and how this was determined, including clinical and statistical assumptions supporting any sample size calculations; also, detail sample size calculations for entomological outcomes
METHODS: ASSIGNMENT OF INTERVENTIONS (FOR CONTROLLED TRIALS)		
Allocation	15	
Sequence generation	16a	The method for generating the allocation sequence (for example, computer-generated random numbers) and a list of any factors used for stratification; to reduce the predictability of a random sequence, details of any planned restriction (such as blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions
Allocation concealment mechanism	16b	The mechanism for implementing the allocation sequence (for example, using sequentially numbered, opaque, sealed envelopes) and any steps that will be taken to conceal the sequence until allocations are assigned
Implementation	16c	Who will generate the allocation sequence? Who will enrol participants? Who will assign participants to interventions?
Blinding (masking)	17	Who will be blinded after assignment to interventions (for example, participants, care providers, outcome assessors, data analysts), and how will this blinding be done?
METHODS: DATA COLLECTION, MANAGEMENT AND ANALYSIS		
Data collection	18a	Plans for assessing and collecting baseline, outcome and other trial data, including any related processes used to improve data quality (such as duplicate measurements, training of assessors) and a description of study instruments (such as questionnaires, laboratory tests), along with their reliability and validity, if known; explain where data collection forms can be found if they are not included in the protocol
	18b	Plans to promote participant retention and to complete follow-up, including a list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols
Data management	19	Plans for data entry, coding, security and storage, including any processes that will be used to improve data quality (for example, double data entry, range checks for data values); explain where details of data management procedures can be found if they ae not included in the protocol
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes; explain where other details of the statistical analysis plan can be found if they are not included in the protocol
	20b	Methods for any additional analyses (such as, subgroup and adjusted analyses)
	20c	Definition of analysis of non-adherent population (for example, as in a randomized analysis), and any statistical methods to handle missing data (such as multiple imputation)
METHODS: MONITORING		
Data monitoring	21a	Composition of data monitoring committee, summary of its role and reporting structure, statement about whether it is independent from the sponsor and whether members are free from competing interests; explain where further details about the committee's charter can be found if they are not included in the protocol; alternatively, provide an explanation about why a data monitoring committee is not needed
	21b	Description of any interim analyses and stopping guidelines, including who will have access to the interim results and make the final decision to terminate the trial

Section and item	Item No.	Description
Harms	22	Estimated number of participants needed to achieve study objectives and how this was determined, including clinical and statistical assumptions supporting any sample size calculations; also, detail sample size calculations for entomological outcomes
Auditing	23	Schedule for enrolment, implementation of interventions, assessments and visits for participants; a schematic diagram is highly recommended
ETHICS AND DISSEMINATION		
Research ethics approval	24	Plans for seeking approval from a research ethics committee or institutional review board
Protocol amendments	25	Plans for communicating important protocol modifications (such as changes to eligibility criteria, outcomes or analyses) to relevant parties (such as investigators, ethics committees, institutional review boards, participants, trial registries, journals, regulators)
Consent or assent	26a	Who will obtain informed consent or assent from potential participants or authorized surrogates, and how will this be done? (see item 32)
	26b	Additional consent provisions for collecting and using participants' data and biological specimens in ancillary studies, if applicable
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared and maintained to protect confidentiality before, during and after the trial
Declaration of interests	28	Financial and other competing interests for principal investigators, both for the overall trial and for each study site
Access to data	29	Statement of who will have access to the final trial data set and disclosure of contractual agreements that will limit such access for investigators
Ancillary and post-trial care	30	Financial and other competing interests for principal investigators, both for the overall trial and for each study site
Dissemination policy	31a	Statement of who will have access to the final trial data set and disclosure of contractual agreements that will limit such access for investigators
	31b	Financial and other competing interests for principal investigators, both for the overall trial and for each study site
	31c	Statement of who will have access to the final trial data set and disclosure of contractual agreements that will limit such access for investigators
APPENDICES		
Informed consent materials	32	Model consent form and other related documentation to be given to participants and authorized surrogates
Biological specimens	33	Plans for the collection, laboratory evaluation and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable