

---

# Data requirements and protocol for determining comparative efficacy of vector control products



# **Data requirements and protocol for determining comparative efficacy of vector control products**



# Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abbreviations</b>	<b>vi</b>
<b>Glossary</b>	<b>vii</b>
<b>Executive summary</b>	<b>xi</b>
<b>1. Background information and rationale</b>	<b>1</b>
1.1 How the comparative efficacy guidance fits with the WHO evaluation pathways and guideline development process	2
1.2 Aim of comparative efficacy assessment	2
1.3 Intervention classes affected by this guidance	2
1.4 Setting the non-inferiority margin	2
1.5 Cost and conduct of comparative efficacy studies	4
1.6 Principles of non-inferiority determination in comparative efficacy studies	4
<b>2. Conduct of comparative efficacy studies</b>	<b>5</b>
2.1 Selection of the active comparator and controls	5
2.2 Site selection	6
2.3 Study design	6
2.4 Sample size considerations	9
2.5 Conduct of comparative efficacy studies	11
<b>3. Data analysis and determination of non-inferiority</b>	<b>14</b>
3.1 Non-inferiority margin	14
3.2 Calculation and analysis of non-inferiority	14
3.3 Determination of non-inferiority	15
3.4 Data reporting	16
<b>4. Future considerations</b>	<b>17</b>
<b>References</b>	<b>18</b>
<b>Annex. Data template and data analysis code</b>	<b>20</b>



## Acknowledgements

The World Health Organization (WHO) gratefully acknowledges the individuals who contributed to the development and revision of this protocol.

Sarah Moore (Ifakara Health Institute/Swiss Tropical and Public Health Institute, United Republic of Tanzania), John Bradley (London School of Hygiene and Tropical Medicine, United Kingdom of Great Britain and Northern Ireland), Joseph Challenger (Imperial College London, United Kingdom of Great Britain and Northern Ireland), Olukayode Odufuwa (Ifakara Health Institute/Swiss Tropical and Public Health Institute, United Republic of Tanzania), Peter Smith (London School of Hygiene and Tropical Medicine, United Kingdom of Great Britain and Northern Ireland), and Immo Kleinschmidt (London School of Hygiene and Tropical Medicine, United Kingdom of Great Britain and Northern Ireland) contributed to the writing and revision of the protocol.

The following WHO staff also contributed to and revised the current protocol: Lauren Carrington, Seth Irish and Jan Kolaczinski (WHO Global Malaria Programme, Switzerland), and Dominic Schuler and Geraldine Foster (WHO Prequalification Unit - Vector Control Products Assessment Team, Switzerland).

## Abbreviations

AI	active ingredient
CI	confidence interval
FIC	first-in-class
IRS	indoor residual spraying
ITN	insecticide-treated net
LS	Latin square
OR	odds ratio
PBO	piperonyl butoxide
SIC	second-in-class
WHO	World Health Organization



## Glossary

<b>active comparator</b>	The active comparator is the reference product to which the candidate product is compared in comparative efficacy studies. Ideally, this should be the first-in-class product, but there may be situations in which a second-in-class product provides a suitable alternative.
<b>blood-feeding inhibition</b>	Blood-feeding inhibition is the proportional reduction of blood feeding. For assessment of insecticide-treated nets and indoor residual spraying products, blood-feeding inhibition is assessed in experimental huts and is calculated relative to untreated controls. For other interventions, this end-point may be measured using other methods. For insecticide-treated net or indoor residual spraying products, this end-point is not used for decision-making, but investigators are required to report their results to the World Health Organization (WHO).
<b>blood-feeding rate</b>	Blood-feeding rate is calculated as the proportion of blood-fed female mosquitoes relative to the total number of female mosquitoes collected.
<b>candidate product</b>	A candidate product is a new product that is being evaluated for its comparative efficacy against an active comparator product in an established intervention class, which enables WHO to determine whether an existing recommendation applies, whether it needs to be extended or whether a new recommendation needs to be formulated.
<b>comparative efficacy assessment</b>	Comparative efficacy assessments in the context of vector control evaluations are entomological studies used to determine the relative entomological performance of candidate products against an active comparator. This direct comparison of products in the same study avoids the issues of data comparability (differences in point estimates) introduced by testing respective products at different sites and at different times. Using data from comparative efficacy studies, a non-inferiority analysis seeks to determine whether the entomological impact of a candidate product is not worse than that of an active comparator by establishing that mortality lies within a predefined acceptable range (i.e. above the non-inferiority margin). See also <i>active comparator</i> , <i>candidate product</i> , <i>control</i> .
<b>control</b>	A negative control is a product that is as similar as possible to the candidate product but does not have the active ingredient. A negative control is used to monitor the quality of the evaluation by ensuring that the observed mortality (or other entomological effect under investigation) is due to the active ingredient(s) and not due to poor execution of the study (e.g. induced mortality from poor handling of mosquitoes). A positive control in this context could be either the standard comparator (see below) or the active comparator (see above).
<b>current standard of care</b>	The current standard of care is the type of vector control product predominantly used by the national malaria control programme in the country where the study will be implemented or where the intervention is expected to be deployed.

<b>entomological effect</b>	Entomological effect refers to a product's effect on a disease vector, such as reduction of vector survival, biting rates, fertility, or human–vector contact, or changes to the vector's susceptibility to infection or transmission. Products with different biochemical modes of action may have similar entomological effects on target insects; for example, indoor residual spraying formulations with pyrethroids and carbamates differ in their chemical modes of action, yet are considered to have a similar entomological effect, in that they both kill mosquitoes.
<b>fertility</b>	Fertility is defined here as the proportion of blood-fed females that develop viable eggs. Fertility is measured in one of two ways among blood-fed mosquitoes that remain alive long enough to complete egg development. First, it can be assessed by dissecting mosquitoes to look for viable (Stage V) eggs at a set time, usually 72 hours after the collection of blood-fed females from an experimental hut. Second, it can also be measured by counting the proportion of females that lay eggs.
<b>first-in-class</b>	First-in-class refers to the first product in an intervention class that generated the epidemiological data to inform the development of one or more WHO recommendations, in turn establishing that intervention class. The first-in-class product should ideally be used as the active comparator in subsequent comparative efficacy studies.
<b>intervention class</b>	<p>An intervention class is defined as a group of interventions with a similar entomological effect. For a new intervention class to be established, one or more WHO recommendations need to be developed based on evaluation of a first-in-class product. Such evaluation consists of a minimum of two independent, well conducted studies with epidemiological end-points. Over time, a number of recommendations may be developed that are linked to the same intervention class but cover different groups of products, e.g. pyrethroid-piperonyl butoxide nets and pyrethroid-chlorfenapyr nets.</p> <p><i>Note: An intervention class is disease-specific, meaning that for interventions with different target diseases, epidemiological impact needs to be demonstrated for each vector-borne disease (or groups of diseases, e.g. in the case of Aedes-borne arboviruses) to develop one or more disease-specific WHO recommendations via the guideline development process and hence establish the class.</i></p>
<b>mortality</b>	Mortality is a standard end-point used in entomological studies, describing a mosquito that cannot stand or fly in a coordinated manner or that shows no movement. Mortality of mosquitoes is usually measured at 24 hours after exposure to an intervention, but may be measured at alternative or additional timepoints, depending on the mode of action of the active ingredient in the product under evaluation.

<p><b>non-inferiority analysis</b></p>	<p>A non-inferiority analysis is a type of analysis conducted on comparative efficacy data. Non-inferiority is shown when the impact of a candidate vector control intervention is no worse than that of the active comparator by a prespecified amount. This amount is known as the non-inferiority margin. The null hypothesis in non-inferiority analysis is that the candidate product is worse than the active comparator by more than the non-inferiority margin. The alternative hypothesis is that the candidate product is no worse than the active comparator, i.e. that the lower bound of the observed range of efficacy (the 95% confidence interval) falls above the non-inferiority margin.</p>
<p><b>non-inferiority margin</b></p>	<p>The non-inferiority margin is the largest allowable difference between the candidate product and the active comparator in terms of the estimates of the chosen end-points (e.g. mortality, fertility). The non-inferiority margin (called “delta”) is predefined for a given comparison. The entire 95% confidence interval of the odds ratio of the end-point for the candidate product must be above (for mortality) or below (for fertility) the non-inferiority margin. For non-inferiority analyses of vector control interventions assessed by WHO, this margin has been set by means of expert consensus to what is considered a tolerable limit from a public health standpoint – a fixed 7% difference, assessed by odds ratios.</p>
<p><b>odds ratio</b></p>	<p>The odds ratio is a measure of how the odds of an outcome differ for two groups. The odds of an outcome is defined by one of two equations: <math>p/(1 - p)</math>, where <math>p</math> is the risk of the outcome, or <math>P/(100 - P)</math>, where <math>P</math> is the percentage risk of the outcome. In the context of the primary end-points of the analyses outlined in this document, the odds ratio is calculated from the odds of a mosquito dying or being sterilized (or other end-point) through exposure to the candidate product versus the odds of the same outcome with the active comparator product. An odds ratio of 1 indicates no difference between the active comparator and the candidate product, an odds ratio <math>&gt; 1</math> indicates that the outcome is more likely with the candidate product, and an odds ratio <math>&lt; 1</math> indicates that the outcome is more likely with the active comparator product.</p>
<p><b>primary end-point</b></p>	<p>The primary end-point is the main outcome to be evaluated, upon which the comparative efficacy study is powered. In the context of comparative efficacy studies for vector control products, the primary end-point is used to make the ultimate decision regarding the non-inferiority of a product and thus whether it will be considered as covered by an existing WHO recommendation. The determination of the primary end-point will depend on the mode of action of the active ingredient in the candidate product.</p>
<p><b>public health value</b></p>	<p>Public health value, also known as disease impact, is the proven protective efficacy of an intervention to reduce or prevent infection and/or disease in humans.</p>
<p><b>residual efficacy</b></p>	<p>The residual efficacy of a product is the duration for which the entomological effect of the product remains above a predefined threshold following application/deployment.</p>

<b>second-in-class</b>	Second-in-class refers to products that are assigned to an established intervention class that already has one or more WHO recommendations associated with it, based on evidence generated by a first-in-class product. For a second-in-class product to be considered as covered by the same recommendation, it is required to demonstrate non-inferiority to an appropriate active comparator (ideally the first-in-class product) in comparative efficacy studies. If the second-in-class product differs in terms of its use pattern or active ingredient, WHO may need to develop a new recommendation or consider extending an existing recommendation via the guideline development process.
<b>secondary end-point</b>	Secondary end-points are outcomes of interest measured in a study, other than the primary end-point. In the context of the analyses outlined in this document, secondary end-points remain important in terms of understanding entomological modes of action and how a product performs, but they are not considered in the decision to determine the non-inferiority of a product, and it is not necessary to consider them in sample-size calculations. Blood-feeding inhibition, for example, is a secondary end-point for evaluating indoor residual spraying and insecticide-treated net products.
<b>standard comparator</b>	A standard comparator is a product that is the current standard of care in the country where the study is being conducted or where the intervention is expected to be deployed, and that belongs to another/older class. For example, in a comparative efficacy study of a candidate pyrethroid-piperonyl butoxide net, the active comparator would be the first-in-class pyrethroid-piperonyl butoxide net, while the standard comparator would be a pyrethroid-only net.
<b>study</b>	In the context of the vector control product evaluations outlined in this document, a study is an individually powered evaluation of a product that aims at generating data to enable non-inferiority analysis of a second-in-class product compared to a first-in-class product. WHO currently requires two well conducted, sufficiently powered, independent studies for a product to be covered by a recommendation associated with an existing intervention class. Where feasible, studies should provide geographical diversity and be conducted in at least two different regions (i.e. East, Central or West Africa). The terms “study” and “trial” are often used interchangeably.
<b>superiority study</b>	A superiority study is one type of comparative efficacy assessment. In contrast to a non-inferiority study, a superiority study seeks to demonstrate that the candidate product is better than the comparator by a predetermined margin. The null hypothesis in superiority studies is that the new treatment is not better than the active comparator, and the study is powered to reject this hypothesis if the candidate product is superior by a specified amount.

## Executive summary

The World Health Organization (WHO) evaluates vector control products with the aim of providing assurance to its Member States that interventions meet two sets of criteria. They must have public health value, namely proven efficacy to reduce or prevent infection and/or disease in humans, and they must meet quality, safety and entomological efficacy standards.

Assessments of the corresponding information for these criteria are conducted under separate procedures overseen by the responsible department. Assessment of quality, safety and efficacy is part of the Prequalification programme and applies at the level of individual products. This process is managed by the WHO Prequalification Unit - Vector Control Products Assessment Team.

Public health value is assessed at the level of an intervention class by means of a first-in-class product generating data that demonstrate epidemiological impact against one or more target diseases. Based on these data a WHO recommendation will be formulated and – in the case of malaria – be communicated via the WHO guidelines for malaria. Over time, the recommendation will cover a group of interventions that share similar characteristics and entomological modes of action, but will be increasingly diverse due to ongoing innovation. An intervention class for which evidence of public health value has yet to be generated or assessed is considered “tentative”. Interventions in a tentative intervention class must follow the new intervention pathway to generate evidence of public health value, which is assessed by the WHO Vector Control Advisory Group. Positive determination of the public health value of a new intervention by the Vector Control Advisory Group triggers the WHO guideline development process, which culminates in the publication of one or more new WHO recommendations. The publication of the recommendation(s) then establishes the (previously tentative) intervention class. All subsequently developed products submitted for assessment and assigned to an established intervention class and an appropriate recommendation under this class are not required to directly demonstrate public health value, provided that they are found to be non-inferior to the first-in-class product (or an appropriate substitute) with respect to an entomological end-point that provides the best correlate of protection. To demonstrate non-inferiority, manufacturers must generate comparative efficacy data on the entomological impact of the new product relative to the first-in-class product. In the case of insecticide-treated net and indoor residual spraying products, these data are generated by means of experimental hut studies.

WHO requires these entomological data as *indirect* evidence of public health value to ensure that a second-in-class product demonstrates similar impact to the first-in-class product that generated the epidemiological data and for which a recommendation is in place. In a comparative entomological efficacy assessment, a product needs to demonstrate:

- non-inferiority to the first-in-class product (active comparator) on the primary end-point(s); and
- superiority over the control or current standard of care (standard comparator) on the primary end-point(s), if applicable.

A fixed absolute difference of 7% is used to calculate the odds ratios used for determination of non-inferiority. Evaluation of the non-inferiority of malaria vector control products enables WHO to determine whether an existing recommendation applies, whether it needs to be extended or whether a new recommendation needs

to be formulated. The aim of the comparative efficacy analysis of products is to provide a relatively easy and cost-effective means of determining the entomological performance of products against a comparator, using data generated through studies already required as part of product evaluation for WHO prequalification. By including a common comparator of known efficacy in all studies, comparative analysis avoids the difficulties of data comparability otherwise introduced by testing different products separately at different sites and at different times. In addition to validating whether an existing WHO recommendation applies to a new product, whether it needs to be extended or whether a new recommendation needs to be developed, results from non-inferiority analyses may inform procurement decisions and/or product selection by WHO Member States and their implementing partners under increasingly resource-constrained conditions.

# 1. Background information and rationale

Since 1 January 2017, the World Health Organization (WHO) has implemented a new process for evaluating vector control products (1). The process seeks to provide enhanced assurance regarding product safety, quality and efficacy (both entomological and epidemiological) to better meet the needs of WHO Member States. The assessment of individual products for their quality, safety and entomological efficacy is overseen by the WHO Prequalification Unit – Vector Control Products Assessment Team. The WHO technical departments, namely the Global Malaria Programme and the Department of Control of Neglected Tropical Diseases, review epidemiological data to assess the public health value of new vector control interventions, which in turn inform the development of WHO recommendations through the guideline development process (2). Assessment of public health value is the mandate of the WHO Vector Control Advisory Group, while the WHO guideline development process is supported by specific guideline development groups and is overseen by the Guidelines Review Committee (3).

The evaluation process for vector control products has continued to evolve to incorporate implementation experience and conform to the guideline development process. As part of these efforts, the WHO Global Malaria Programme and Department of Control of Neglected Tropical Diseases, with the support of the Vector Control Advisory Group, reviewed and reduced the overall number of intervention classes. With fewer intervention classes that are broader in scope, the number of epidemiological trials to inform WHO recommendations was reduced. In doing so, however, the potential product diversity within a class increased considerably, raising the question as to whether products grouped within a specific class perform similarly to the first-in-class (FIC) product that established the intervention class and whether the WHO recommendation(s) that was/ were originally developed based on data for the FIC product continue(s) to be applicable to the increasingly diverse range of products in that class.

The need to validate whether WHO recommendations are applicable, or whether they need to be amended or complemented by supplemental recommendations, was recognized by WHO and its advisory groups as early as 2017. Based on technical consultation, WHO embarked on a process to explore the use of comparative entomological efficacy as a correlate of protection (4). A notice of intent to this effect was published by WHO in 2018 (5), followed by a study protocol in 2019 (6). WHO then evaluated the practicality of the process, assessing the actual data generated by studies comparing mosquito nets treated with a pyrethroid insecticide only and those treated with a pyrethroid and the synergist piperonyl butoxide (PBO) (7). For indoor residual spraying (IRS), comparative efficacy data were used to expand the relevant WHO recommendation for IRS to include neonicotinoid insecticides in 2017 (8) and to broflanilide in 2023 (9), and the need for comparative data was explicitly referenced in the associated preferred product characteristics in 2022 (10).

Based on these encouraging practical experiences and in the context of an ever-increasing range of vector control products seeking to enter the market, the WHO Malaria Policy Advisory Group advised WHO to implement comparative efficacy assessments (using a method called “non-inferiority assessment”) broadly across all established intervention classes (11,12). In 2023, the Malaria Policy Advisory Group reiterated its earlier guidance that comparative efficacy assessments of entomological data are required for all products other than the FIC product that generated the epidemiological data used to establish an intervention class (13). In line with this guidance, the Global Malaria Programme and the WHO Guidelines Review Committee discussed how entomological data should be used. Revised guidance now incorporates methodological recommendations from the 2021 (7) and 2023 (9) technical consultations and establishes comparative efficacy assessment as part of the WHO vector control evaluation process.

## 1.1 How the comparative efficacy guidance fits with the WHO evaluation pathways and guideline development process

To assess the public health value of a new vector control intervention, at least two well conducted, independently powered studies with epidemiological end-points are needed. These epidemiological impact data are, along with other eligible information, used to develop one or more WHO recommendations for products within that class. The publication of the WHO recommendation(s) then formally establishes the intervention class, and subsequent products submitted for evaluation falling within that same class (called second-in-class (SIC) products) do not need to conduct epidemiological studies. They do, however, need to demonstrate that they are non-inferior to the FIC product (or a suitable alternative active comparator) with respect to entomological efficacy to be considered as covered by the applicable recommendation(s).

## 1.2 Aim of comparative efficacy assessment

Entomological comparative efficacy assessments are conducted to validate the applicability of WHO recommendations developed by means of epidemiological studies, to provide an evidence base to extend existing recommendations, or to inform the development of new recommendations under the same intervention class.

## 1.3 Intervention classes affected by this guidance

WHO processes require that comparative assessments of entomological data be conducted for all malaria vector control products other than FIC products. The only exception is for pyrethroid-only insecticide-treated nets (ITNs), which do not require evidence of non-inferiority, as they are in the process of being replaced by next-generation ITNs. To date, this method has also not been adopted to assess products for use in NTD vector control.

The latest edition of this document provides specific guidance on comparative efficacy assessments for IRS and ITNs and the associated non-inferiority analyses. The concepts are, however, broadly applicable to the comparative evaluation of all other vector control interventions and should be adopted by manufacturers in consultation with WHO for the evaluation of other products. To support this mainstreaming of comparative efficacy assessments, WHO will work with manufacturers and collaborators that are currently conducting epidemiological and entomological evaluations of new vector control products to evolve the information contained in this guidance and accommodate specific details on comparative efficacy evaluation methods and end-points other than those for IRS and ITNs.

## 1.4 Setting the non-inferiority margin

Setting the non-inferiority margin is an essential prerequisite of non-inferiority analysis. It requires a balance between ensuring that products demonstrating non-inferiority have a similar impact to that of the FIC product (for which epidemiological outcomes have been measured) and minimizing the risk of excluding well performing products from being covered by a WHO recommendation.

As outlined in the 2019 version of the protocol, the primary end-point used for non-inferiority assessments of ITN and IRS products was mortality, and the non-inferiority



margin was originally set at an odds ratio (OR) of 0.7 (7). However, it was subsequently determined that this OR would require unrealistically large sample sizes for a candidate product to demonstrate non-inferiority to a well performing active comparator. To avoid high-performing products being delayed or prevented from entering the market, the topic of appropriate non-inferiority margins was revisited during a technical consultation in 2023 (9). It was decided that, although the initial decision to express the non-inferiority margin as an OR was appropriate, the value of the OR should vary depending on the level of mortality achieved by the active comparator. The minimum mortality permissible for the candidate product was thus deemed acceptable if it was not more than a fixed absolute difference of 7% below that of the active comparator product. How this 7% absolute difference converts into an OR depends on the mortality induced by the active comparator product; examples are given in Table 1 (9).

**Table 1. The non-inferiority margin based on a fixed difference of 7% between the mortality induced by the candidate product and mortality induced by the active comparator product**

Active comparator		Candidate product		Corresponding OR for a 7% non-inferiority margin <sup>b</sup>
Point estimate (%)	Odds <sup>a</sup>	Acceptable lower bound (%) of CI with a 7% absolute difference	Odds <sup>a</sup>	
95	19.00	88	7.33	0.39
90	9.00	83	4.88	0.54
80	4.00	73	2.70	0.68
70	2.33	63	1.70	0.73
60	1.50	53	1.13	0.75
50	1.00	43	0.75	0.75
40	0.67	33	0.49	0.74
30	0.43	23	0.30	0.70

<sup>a</sup> If the percentage mortality is p, then the odds of mortality = p/(100 – p)  
<sup>b</sup> The OR is calculated by dividing the odds in the second column by the odds in the first (e.g. 7.33/19.00 = 0.39)

The interpretation of these ORs is that, to demonstrate non-inferiority, the lower limit of the 95% confidence interval (CI) for the OR comparing the candidate product to the active comparator product cannot fall below the non-inferiority margin. For example, if mortality in the active comparator product is 90%, the lower limit of the 95% CI for the OR cannot fall below 0.54.

For the outcomes of blood feeding and fertility, it was decided that the same 7% absolute difference should be used. However, for these outcomes, the candidate product should have a blood-feeding or fertility rate that is at most 7% *higher* than that of the active comparator product. Examples of how to convert these differences into ORs are given in Table 2.

**Table 2. The non-inferiority margin based on a fixed difference of 7% between blood feeding or fertility for the active comparator and the candidate product**

Active comparator		Candidate product		Corresponding OR for a 7% non-inferiority margin <sup>b</sup>
Point estimate (%)	Odds <sup>a</sup>	Acceptable lower bound (%) of CI with a 7% absolute difference	Odds <sup>a</sup>	
5	0.05	12	0.14	2.59
10	0.11	17	0.20	1.84
20	0.25	27	0.37	1.48
30	0.43	37	0.59	1.37
40	0.67	47	0.89	1.33
50	1.00	57	1.33	1.33
60	1.50	67	2.03	1.35
70	2.33	77	3.34	1.43

<sup>a</sup> If the percentage of blood feeding/fertility is p, then the odds of blood feeding/fertility =  $p/(100 - p)$

<sup>b</sup> The OR is calculated by dividing the odds in the second column by the odds in the first (e.g.  $0.14/0.05 = 2.59$ )

## 1.5 Cost and conduct of comparative efficacy studies

To avoid the need to conduct additional studies to generate comparative efficacy data, and thus to minimize the cost and time associated with data generation, the study design for IRS and ITN products was developed to ensure alignment with standard experimental hut studies routinely conducted to generate data for the prequalification dossier. The same principles will be applied to develop guidance for comparative efficacy data generation for interventions other than ITNs and IRS, with detailed methods to be developed in consultation with manufacturers and researchers.

## 1.6 Principles of non-inferiority determination in comparative efficacy studies

WHO's principles for comparative efficacy assessments are as follows:

- Non-inferiority is determined through a comparative assessment drawing on indirect (entomological) evidence to provide a certain level of reassurance of the likely public health value of all products within an intervention class, other than the FIC product (or another suitable active comparator) that directly generated the epidemiological data to demonstrate such impact.
- No additional studies beyond those needed to generate data for the WHO prequalification dossier should be required to enable non-inferiority analysis.
- The results of the comparative efficacy assessment cannot be used as a label claim.
- Comparative efficacy assessments are intended to support the decision-making processes underpinning the sourcing of vector control products by WHO Member States.
- Non-inferiority is not used as a measure of product quality; it is used only to assess the entomological efficacy of the product in the context of WHO's living guidelines and the recommendations contained in them.

## 2. Conduct of comparative efficacy studies

Comparative efficacy studies require the assignment of an active comparator (more detail provided in section 2.1) to which the candidate will be compared, and selection of at least two geographically separate study sites (section 2.2). The study should be conducted using a standardized experimental design (section 2.3) with adequate replication (section 2.4). A standard study procedure has been developed for experimental hut studies (section 2.5), with intervention-specific points relating to ITNs (section 2.5.1) and IRS (section 2.5.2) outlined for consideration. Adherence to these procedures will enhance the likelihood that the non-inferiority of the candidate product to the active comparator may be reliably inferred using a standardized analysis (sections 3.1, 3.2, 3.3), reporting and presentation (section 3.4).

### 2.1 Selection of the active comparator and controls

The active comparator used in comparative efficacy studies should ideally be the FIC product that demonstrated public health value. This is irrespective of the number of other products subsequently covered by the same WHO recommendation and grouped within the same intervention class.

It is recognized that investigators may experience difficulties in obtaining the FIC product for evaluation in comparative efficacy studies or, in the case of IRS, that the diversity of products under an existing recommendation requires alternative choices to be made. In these cases, the Global Malaria Programme should be consulted during the planning phase of the studies to assist with the sourcing of the product or to determine a suitable alternative by means of the decision sequence below (gmp-vcr@who.int).

There are four possibilities for assigning an active comparator product, listed below in order of priority:

1. the FIC product of that intervention class;
2. any SIC product for which epidemiological evidence of public health benefit is available from randomized controlled trials;
3. any SIC product that has shown superiority to the FIC product in entomological studies with respect to the primary end-point; and
4. in the event that no SIC product has shown superiority to the FIC product and the FIC product is unavailable, the most appropriate product among the SIC products.

If more than one product is available in the same intervention class after the prioritization above, the study investigators should prioritize selection of the active comparator product based on similar active ingredients (AIs), product design, method of production and insecticide delivery method (e.g. incorporated vs coated for ITNs). Applicants are requested to inform WHO of their selection of an active comparator before starting the study; they may need to provide the rationale for their choice. The choice of a specific active comparator will, in all cases, need to be justified in the study report, especially if there is no direct evidence that the chosen active comparator has public health value.

Preferably, the active comparator should be obtained from the manufacturer to ensure that it fully conforms to its specification, which could result in inconsistent performance across studies and, in turn, undermine data quality. A certificate of analysis of any active comparator should be supplied with the study report to assure the quality at time of manufacture of the product used in the comparative efficacy study.

In addition to the active comparator, all studies are required to include a negative control arm, i.e. an untreated net for ITNs or water spray for IRS, to verify that the conduct of the study is of sufficient quality and to estimate the natural mortality during mosquito holding or, where relevant, changes in blood feeding that are induced by the candidate product. As per standard procedures for entomological investigations for vector control, studies in which the overall mortality in the control arm (over the duration of the study) is > 10% at 24-hour holding, or > 20% for longer holding times, are not considered acceptable and will need to be repeated. Investigations into the cause of excessive mortality should be conducted at the investigators' discretion.

When testing ITNs with dual AIs, an appropriate standard comparator (e.g. pyrethroid-only ITN) should be included in the study design.

## 2.2 Site selection

A minimum of two independently powered studies must be conducted in two sites with differing vector populations and/or resistance status, such as one in West or Central Africa and one in East Africa.

Sites suitable for undertaking the comparative assessment studies need to meet multiple criteria, as listed below:

- There should be a sufficient number of huts of the same design to enable all arms of the experiment to be run simultaneously.
- Quality-assured (Good Laboratory Practice-compliant) test facilities are required for vector control product evaluation.
- Mosquito species composition and the insecticide resistance status of each major vector species at the study site must be characterized. This should include molecular characterization of the resistance mechanisms, and characterization of metabolic mechanisms for products designed to counteract this type of resistance mechanism (14).

If *Anopheles* species complexes are present, data should be collected and analysed to determine the dominant species. However, a pooled analysis of species complexes may also be performed, if justified, as long as the insecticide resistance status of each subspecies is evaluated and reported.

## 2.3 Study design

The primary end-point to be measured to assess non-inferiority will be determined by the type of intervention being tested. Manufacturers/investigators will be informed of the required end-points to evaluate at the time WHO provides a response to a manufacturer's determination of pathway request. For a product to be considered covered by the recommendation in question, the product must show that it is non-inferior when compared to the active comparator with respect to the primary end-point.

Secondary end-points will not be used directly by WHO to make decisions on product non-inferiority. Nevertheless, non-inferiority analysis is strongly encouraged for relevant secondary end-points such as blood-feeding inhibition. WHO will inform manufacturers/investigators of the minimum secondary end-points to be measured. This information will enable a comprehensive understanding of the broader mode of action of a product, but will not be used for decision-making.

### **2.3.1 ITNs**

For ITNs, both unwashed and washed (20 times) nets should be used in the study. In the primary analysis, a single pooled estimate of efficacy combining the washed and unwashed nets is generated for each net to give an estimate of overall product performance over its lifetime in the field. Combining the two arms (washed and unwashed nets) also increases replication in the analysis and consequently the precision of the estimates. Three analyses are conducted for ITNs:

- primary analysis of the primary end-point: all data (unwashed and washed for ITNs), with wash status included as a fixed covariate in the primary analysis;
- secondary analysis of the primary end-point for unwashed ITNs; and
- secondary analysis of the primary end-point for washed (20 times) ITNs.

The non-inferiority determination is based on the results of the analysis of the primary end-point. For ITNs, the primary end-point to be assessed will depend on the AI(s) of the net. ITNs for which the primary mode of action is to kill mosquitoes have the primary end-point of mortality and a secondary end-point of blood feeding. ITNs for which the primary mode of action is to reduce mosquito fertility have a primary end-point of fertility reduction and secondary end-points of mortality and blood feeding. Table 3 shows the recommended holding times for different insecticide classes, based on their different modes of action. The estimated effect of the intervention, e.g. percentage mortality and 95% CIs for each arm, and the combined results for washed and unwashed nets should be reported in addition to the OR and its 95% CI, derived from the regression model (section 3.2). Results of all analyses must be presented together in forest plots (e.g. Fig. 1), subdivided by species where relevant. A recommended sample script for analysis is available (see Annex).

### **2.3.2 IRS**

Similarly, for IRS, the analysis should be conducted over the full duration of the observed product efficacy (i.e. residual efficacy) and include data from all substrates tested. Substrates may include concrete, mud or wood, for example. As all IRS treatments are expected to provide residual efficacy for at least three months, an analysis at this timepoint should also be done. Two analyses are conducted for IRS:

- primary analysis of the primary end-point: all substrates at the longest duration of efficacy of the candidate product, with substrate included as a fixed covariate in the primary analysis; and
- secondary analysis of the primary end-point for each IRS substrate at three months and at the longest duration of efficacy.

The non-inferiority determination is based on the results of the analysis of the primary end-point only. Currently, all IRS formulations are expected to be assessed on the basis of mosquito mortality. Table 3 shows the recommended holding times for different insecticide classes, based on their different modes of action. The estimated effect of the intervention, e.g. percentage mortality and 95% CIs for each arm, and the separate and combined results for the substrates should be reported in addition to the OR and its 95% CI, derived from the regression model (section 3.2). Results of all analyses must be presented together in forest plots (e.g. Fig. 1), subdivided by species where relevant. An example of an analysis script is available (see Annex).

**Table 3. List of non-inferiority end-points and when to measure them, according to the chemical mode of action (applicable to ITNs and IRS only)**

Insecticide class	Chemical mode of action	Example chemistry	Note	Primary end-point	Holding time before measurement	Additional end-points	Quality assurance bioassay
Sodium channel modulators (pyrethroids)	Nerve action	Pyrethroids Pyrethroids with synergist PBO	Not applicable for pyrethroid-only ITNs	Mortality	24 hours	Proportion blood-fed	Cone test
Acetylcholinesterase (AChE) inhibitors (carbamates, organophosphates)	Nerve action	Pirimiphos-methyl Bendiocarb		Mortality	24 hours	Proportion blood-fed	Cone test
Nicotinic acetylcholine receptor competitive modulators (neonicotinoids)	Nerve action	Clothianidin	AI is slow-acting so longer holding times are needed	Mortality	72 hours	Proportion blood-fed	Cone test
Uncouplers of oxidative phosphorylation via disruption of the proton gradient (pyrroles)	Energy metabolism	Chlorfenapyr	Requires insects to be metabolically active during testing and AI is slow-acting so longer holding times are needed	Mortality	72 hours	Proportion blood-fed	Tunnel test/TBD
GABA-gated chloride channel allosteric modulators (meta-diamides and isoxazolines)	Nerve action	Broflanilide	AI is slow-acting so longer holding times are needed	Mortality	72 hours	Proportion blood-fed	Cone test
Juvenile hormone mimics (pyriproxyfen)	Growth regulation	Pyriproxyfen	Inhibits the development of viable eggs/larvae	Fertility	72 hours	Proportion dead, Proportion blood-fed	Cone test

## 2.4 Sample size considerations

Non-inferiority analysis requires the OR to be measured precisely enough to enable clear classification of products. If studies have inadequate statistical power to justify their results, WHO will be unable to consider them in its assessment. As such, it is imperative that sample size and study power be considered in advance of the study in order to avoid, to the extent possible, requests for additional studies.

Sample size is estimated as the number of replicates required to measure the point estimate of the primary end-point precisely enough that the 95% CI has a high probability of enabling a determination of non-inferiority. WHO assessments of non-inferiority should use a margin of 7% of the absolute difference in measures of the primary end-point between treatment arms, converted into an OR as in Tables 1 and 2. Sample size calculations should be presented in reports alongside study results as justification for the conclusions drawn. The requisite sample size will depend on the length of the study, number of huts per treatment arm, the absolute entomological efficacy of the intervention tested and other variability inherent in the study.

The required sample size for a study can be calculated through simulation. The primary analysis makes the assumption that the average impact of the intervention over its lifetime can be used as a single end-point, thus simplifying power calculations. For experimental hut studies, variability includes the number of mosquitoes collected per hut per day, as well as the differences between huts, between sleepers and between observations. These factors will vary by setting and over time, so it is important to use data from recent hut studies at that site to estimate study power and thus calculate the required sample size for each treatment group. In the case of hut studies, this sample size will be the number of hut-nights per product, and the duration of the study will be influenced by the number of huts available to complete these observations.

The study power measures the probability of the study demonstrating the non-inferiority of the candidate product to the active comparator if the true efficacy of the new product is not worse than that of the active comparator. Study power can be estimated by simulating the anticipated efficacy of the active comparator product and the candidate product, and assuming a 7% difference between them to determine the percentage of runs that correctly classify the candidate as non-inferior. Comparative efficacy assessments for ITNs should also consider superiority to the standard comparator as part of their sample size calculations. The study should be designed to have a power of at least 80% (i.e. a type two error rate  $\beta \leq 20\%$ ), meaning that the probability of demonstrating non-inferiority to the active comparator and superiority to the standard comparator (if applicable) is at least 80%.

Even if recent data have been used to parameterize the power calculations, the number of hut-nights needed to obtain the required study power can be re-estimated during the experiment, such as after one full rotation of the Latin square (LS) (e.g. for a 7x7 LS, this would be after 49 experimental nights). While it is legitimate to re-estimate sample size based on values of parameters observed in the study, it is not legitimate to increase the sample size based on non-inferiority not being achieved and to continue to run the experiment until an analysis shows non-inferiority. To do so would increase the type one error rate. Plans to re-estimate sample sizes should be specified in advance of the study, in either the protocol or the statistical analysis plan. Any amendment to the original sample size should be documented in the protocol and statistical analysis plan, and also in reports or scientific publications describing the study.

The steps to conduct sample size calculations for comparative efficacy assessments, using the example of mortality, are as follows:

1. Calculate:
  - a. expected mosquito mortality observed for the active comparator;
  - b. expected mosquito mortality observed for the standard comparator (if applicable);
  - c. average number of mosquitoes expected per hut and the variability in mosquito density per day; and
  - d. variability due to hut, sleeper and day.
2. Define:
  - a. number of huts that will be used;
  - b. days per week the experiment is to be run (e.g. seven days for a 7x7 design, followed by a break before the next treatment is allocated in the case of ITN studies); and
  - c. number of LS rotations for the experimental hut studies (e.g. for a 7x7 design, one full rotation using seven huts (one hut per treatment) is a minimum of 49 study nights).
3. Use the information derived in steps 1–2, together with the defined non-inferiority margin based on a 7% absolute difference in mosquito mortality between the active comparator and the candidate product (Tables 1 and 2), to simulate theoretical experimental hut study results for all treatments (assuming that the percentage mortality follows a binomial distribution). To estimate study power, the true mortality of the candidate product (i.e. the underlying actual probability that a mosquito will die) should be the same as that of the FIC product, i.e. the candidate product is truly no worse than the active comparator.
4. Fit the logistic regression model outlined in section 3.1 to simulated data and determine whether non-inferiority to the active comparator and superiority to the standard comparator (if applicable) have been shown.
5. Repeat the simulation (steps 3–4) 1000 times and record the percentage of times non-inferiority to the active comparator and superiority to the standard comparator (if applicable) are both demonstrated. This is considered the study power.
6. Repeat steps 3–5, adjusting the number of replicates used (i.e. increasing the number of huts for each study arm or the number of rotations according to the experimental capacity available) until the desired power of  $\geq 80\%$  has been achieved.



## 2.5 Conduct of comparative efficacy studies

To date, comparative efficacy data in the area of vector control have been generated for ITN and IRS products only, using experimental hut studies as the established evaluation method. This section outlines the approach for experimental hut studies in detail. Future versions of this document will be expanded to cover evaluation methods suitable for assessing other types of vector control interventions that cannot be tested in experimental huts, once these methods have been established.

The methods to be used for conducting experimental hut studies on ITNs are described in the implementation guidance provided by the WHO Prequalification Unit – Vector Control Products Assessment Team (17); therefore, only those recommendations that are not included in that guidance document are provided here. These same principles and approaches will be used in the development of guidelines and implementation guidance related to IRS.

### 2.5.1 ITN evaluation

Currently, ITNs must demonstrate non-inferiority to an appropriate active comparator using the combined data for unwashed and washed (20 times) ITNs. This washing effect aims to reflect the change in efficacy of a net over its usable lifetime (2–3 years), and thus the primary analysis is based on the combined data, rather than on data from either individual condition.

For all ITN products, inclusion of a pyrethroid-only ITN in the study is required to determine any killing effect of a candidate net that deploys additional or alternative AIs over that of pyrethroid-only ITNs. The pyrethroid-only net is classified as the standard comparator within the study and must be a WHO-prequalified product. Ideally, the standard comparator should have the same pyrethroid insecticide as the candidate net.

To ensure a well-designed study, the minimum set of treatment arms needed to evaluate a candidate dual AI net in an area of pyrethroid resistance includes:

1. candidate ITN unwashed (candidate product)
2. candidate ITN washed 20 times (candidate product)
3. active comparator unwashed (see guidance provided in section 2.1)
4. active comparator washed 20 times (see guidance provided in section 2.1)
5. standard comparator unwashed (pyrethroid-only ITN)
6. standard comparator washed 20 times (pyrethroid-only ITN)
7. untreated net unwashed (negative control).

This would, at minimum, result in a 49-night study, based on a single full LS rotation with seven huts used. More than one candidate product may be evaluated in any given experimental hut study. Study power and sample size must be reconsidered as appropriate.

End-points to be evaluated are specifically indicated in section 2.3.1; to note, the end-points differ according to the primary mode(s) of action of the AI(s) of the net in question. Both primary and secondary end-point analyses must be reported using the combined data from unwashed and washed nets, along with the individual treatment comparisons.

### 2.5.2 IRS evaluation

IRS products need to be assessed for their efficacy on different substrates, which should reflect the building material substrates to which they will be applied under intended deployment conditions. Substrates could be wood, mud or concrete, or other wall materials. To be considered covered by an existing WHO recommendation, the candidate IRS product must demonstrate non-inferiority to an appropriate active comparator with respect to the mortality end-point, using the combined data from multiple substrates tested within a study.

WHO requires IRS products to provide a minimum residual efficacy of at least three months, and this should be considered the minimum acceptable period of efficacy for any new candidate product. For products with longer residual efficacy, mortality in cone bioassays should be measured monthly for the candidate product until it drops below 80% for two consecutive months. If the cone bioassay is not appropriate, another method may be proposed during the initial consultation with WHO. The minimum estimated duration of efficacy should be considered in calculating the sample size and estimating the number of replicates needed for the study.

Products must demonstrate non-inferiority to a suitable active comparator, which in the case of IRS may include products of different insecticide classes, i.e. with different chemical modes of action, provided that mortality is measured for all products at the same holding times. For IRS, the active comparator should be a WHO-prequalified IRS product (other than a pyrethroid insecticide); it should be selected based on the standard of care in the geographical region where the experimental hut study is being conducted or where the product is intended for deployment. Data must be reported for a 24-hour holding time, or longer holding times for products with slower modes of action, if required, provided that the control mortality is acceptable (< 20%) (see Table 3). The following study arms are required at minimum for each substrate being evaluated:

1. water (negative control) for substrate 1
2. water (negative control) for substrate 2
3. active comparator (see guidance provided in section 2.1) for substrate 1
4. active comparator for substrate 2
5. candidate IRS product for substrate 1
6. candidate IRS product for substrate 2.

The material used for the walls of the structures to be sprayed (e.g. mud, concrete, wood, etc.) will affect the performance of the product. Therefore, the selection of the substrate should be justified based on the common housing materials in the region where the product is to be used.

Because IRS treatments cannot be rotated between huts, the use of at least four huts per treatment/substrate arm is recommended to overcome the heterogeneity between huts, taking between-hut variability into account. Data quality is improved by further increasing the number of huts per arm, and this should be considered in the power analysis for the study. To this end, study arms, and the number of replicates per arm, should be maximized depending on hut availability for the selected study site. If the number of huts in an experimental hut study site is a limiting factor, the number of negative control huts can be reduced to one per substrate type (resulting in a minimum of 18 huts needed for two substrates).

### **2.5.2.1 Preparation of experimental huts**

IRS is applied to the walls of the hut in accordance with the manufacturer's instructions for use. The ceiling and doors are left unsprayed to avoid confounding the efficacy of the IRS on different substrates, e.g. mud walls and thatch roof. The ceiling and doors of the huts in IRS studies should be covered with a material that reduces mosquito resting (e.g. stretched plastic) so as to maximize the likelihood of mosquitoes resting on the treated wall surfaces.

The quality of spraying in hut studies is an essential prerequisite for any comparative efficacy evaluation or study for WHO prequalification. Spray application must be within  $\pm 50\%$  of the label-recommended target dose for the IRS product (18), as determined through filter paper analysis. Optimal spraying is achieved by employing well trained personnel (19), using calibrated compression sprayers with control flow valves and carefully calculating the concentration of insecticide in the tank prior to spraying. Gravimetric verification of the spray dose is recommended based on i) the calculation of the hut surface area sprayed; ii) the weighing of decompressed spray tanks before and after spraying; and iii) the estimated quantity of solution applied per metre of surface in the huts.

### **2.5.2.1 Experimental hut procedure for IRS**

Hut studies should follow existing WHO guidance (16). Treatments should be allocated randomly to the experimental huts. Sleepers should enter and leave the huts at predefined times each evening and morning. Similar to the studies with ITNs, the sleepers should be rotated through the huts each night. In the exceptional cases in which cows are used as bait animals, appropriate justification for not using humans must be provided in the report, as comparative efficacy studies are designed to replace studies with epidemiological end-points. Nevertheless, similar to studies with humans, use of cattle as baits would require daily rotation between huts.

### 3. Data analysis and determination of non-inferiority

#### 3.1 Non-inferiority margin

As mentioned in section 1.4, a 7% absolute difference in efficacy between the active comparator and the candidate product has been used to set the non-inferiority margin, which is expressed as an OR and varies depending on the effect of the active comparator (see Tables 1 and 2).

#### 3.2 Calculation and analysis of non-inferiority

To assess whether one product is non-inferior to another, a prespecified model is applied to the data, which accounts for the effect of various experimental factors. The model also provides the OR to compare the active comparator and candidate product. For both ITNs and IRS products, this calculation is performed without disaggregating by wash status or substrate.

To assess non-inferiority for ITNs, a logistic model for mortality (or fertility) should be used, with the following fixed effects:

- treatment (referring to the product)
- hut
- sleeper
- number of washes (which will be either 0 or 20)
- day of collection.

To assess non-inferiority for IRS, a logistic model for mortality should be used, with the following fixed effects:

- treatment (referring to the product)
- hut
- sleeper
- substrate
- day of collection.

All covariates should be categorical fixed effects and the active comparator should be used as the reference intervention (intercept). For experimental hut studies, covariates include treatment, hut, sleeper, day of collection, and whether the net was washed or not as fixed effects, because these factors are sources of systematic variability that are accounted for in the experimental design.

Once the OR and 95% CI of the candidate product relative the active comparator has been calculated, this is plotted on a graph with the active comparator mean on the y-axis and the OR on the x-axis. For mortality, ORs above 1 indicate better candidate performance compared to the active comparator, whereas ORs below 1 indicate poorer candidate performance. The opposite is true for fertility and blood feeding.

The second step is to determine non-inferiority. Non-inferiority of a product can only be determined by comparing the bounds of the 95% CI of the OR to the non-inferiority margin. To do so, the variable non-inferiority margin should be plotted on the same graph (see Annex). Since higher mortality indicates a better product, the candidate product will be determined to be non-inferior in terms of mosquito mortality if the lower bound of the 95% CI estimate falls entirely above the non-inferiority margin (Table 1). It will be determined to be superior if the lower bound of the 95% CI estimate falls entirely above 1.

When the primary end-point is the proportion of fertile mosquitoes (or the proportion that are blood-fed), more efficacious products will result in relatively lower ORs (i.e. below 1). Therefore, the candidate product will be determined to be non-inferior if the upper bound of the 95% CI estimate falls entirely below the non-inferiority margin (Table 2). It will be determined to be superior if the upper bound of the 95% CI estimate falls entirely below 1.

### 3.3 Determination of non-inferiority

A candidate product must show non-inferiority to the active comparator with respect to the primary end-point to be considered as covered by the WHO recommendation that is applicable to the active comparator.

For products for which the primary end-point is mortality, the candidate product is deemed non-inferior if the following criteria are met:

- The lower bound of the 95% CI estimate of the OR of the candidate product to the active comparator is greater than the OR corresponding to a 7% non-inferiority margin (see Table 1).

In addition, for ITNs, the following applies:

- The candidate product is classified as superior to the standard comparator at the 5% significance level (i.e.  $P < 0.05$ ). The choice of whether the candidate product should be compared to a control or to the current standard of care will depend on the product and should be justified, following consultation with the Global Malaria Programme. The statistical model to assess superiority to the standard comparator is the same as the one described in section 3.2 to assess non-inferiority to the active comparator.

For products for which the primary end-point is fertility (e.g. for sterilizing ITNs), the following criteria apply:

- The primary end-point is mosquito fertility and the upper bound of the 95% CI estimate of the OR of the candidate product to the active comparator is lower than the OR corresponding to a 7% non-inferiority margin (see Table 2).

In both instances, these data need to be consistent across at least two studies, performed in different geographical and epidemiological environments. If either of the two studies fails to meet these criteria, a third study should be conducted.

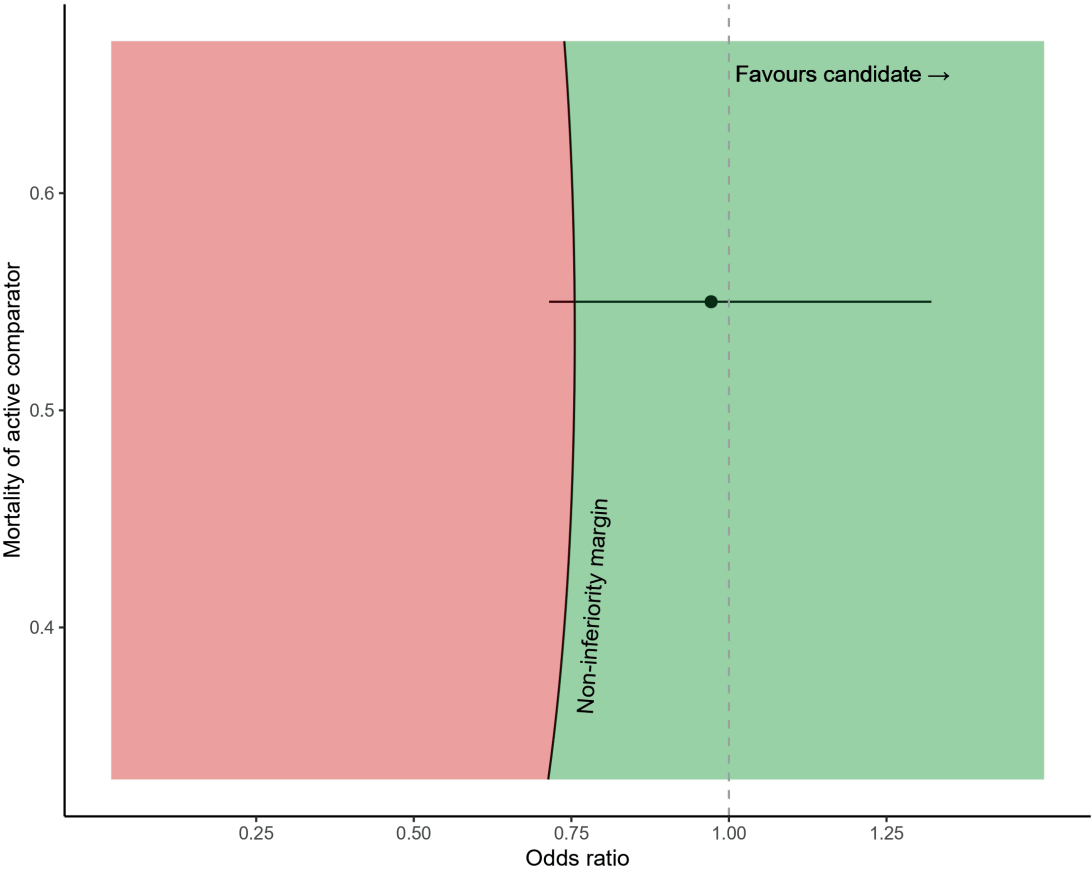
In addition to the criteria for non-inferiority, candidate products also need to show superiority over the standard comparator product. Studies in which the candidate product fails to show benefit over the negative control/standard comparator (depending on the product claim) will need to be repeated in a third study.

No more than three studies should be conducted in total, and non-inferiority needs to be demonstrated in at least two distinct locations. Products will be considered as having failed to demonstrate non-inferiority if they show inferiority or inconclusive results in any two of the maximum three comparative efficacy studies. These products will be required to undergo further product development to enhance their performance, or they will need to provide epidemiological impact data for assessment of their public health value.

### 3.4 Data reporting

The data will need to be reported to WHO as recommended in the guidance of the WHO Prequalification Unit - Vector Control Products Assessment Team (17). Figures must be produced to present the ORs and non-inferiority margins graphically. These studies can be plotted on a graph with the OR on the x-axis and mortality in the active comparator arm on the y-axis. An example is provided in Fig. 1, with supporting guidance on how to produce the plots found in the Annex.

**Fig. 1. A hypothetical example of a non-inferiority analysis of mortality, showing the calculated ORs using the fixed difference non-inferiority margin of 7%. In this case, this product would not be considered non-inferior because the lower bound of the 95% CI of the OR falls below the non-inferiority margin.**



## 4. Future considerations

It should be noted that the comparative efficacy data generated in experimental hut studies to assess the non-inferiority of ITN and IRS products provide limited insight into the bioefficacy and/or durability of products over time under field conditions. The performance of vector control products should therefore be monitored under field conditions to assess performance over time.

Techniques other than experimental huts may be equally suitable for non-inferiority assessments and may offer certain advantages. Experimental hut studies with free-flying mosquitoes are currently the accepted method for evaluating vector control products designed to be used indoors. However, infrastructure requirements mean that these tests can currently only be carried out in a small number of sites, mainly in Africa. Consequently, products can only be evaluated against a limited number of mosquito vector populations. Furthermore, the reliance of experimental hut studies on having a sufficient number of local free-flying mosquitoes means that study duration is affected by the season and level of routine local mosquito control. Resources permitting, other potential alternative testing methods that are approved by the WHO Prequalification Unit - Vector Control Products Assessment Team, e.g. the Ifakara ambient chamber test may be used for non-inferiority studies alongside experimental huts in order to investigate whether these other methods present suitable alternatives for generating non-inferiority data.

For other intervention classes, guidance will be developed to support appropriate testing of non-inferiority in due course. WHO encourages active engagement of manufacturers and researchers in this process. Any enquiries should be directed to: [gmp-vcr@who.int](mailto:gmp-vcr@who.int).

## References

1. WHO handbook for guideline development, second edition. Geneva: World Health Organization; 2014 (<https://iris.who.int/handle/10665/145714>, accessed 10 October 2023).
2. Norms, standards and processes underpinning development of WHO recommendations on vector control. Geneva: World Health Organization; 2020 (<https://iris.who.int/handle/10665/338030>, accessed 10 October 2023).
3. Guidelines Review Committee [website]. In: Groups. Geneva: World Health Organization (<https://www.who.int/groups/guidelines-review-committee>, accessed 10 October 2023).
4. WHO malaria policy advisory committee meeting: meeting report, October 2017. Geneva: World Health Organization; 2017 (<https://iris.who.int/handle/10665/259396>, accessed 10 October 2023).
5. Notice of intent to introduce non-inferiority studies as a component of the vector control evaluation process. Geneva: World Health Organization; 2018.
6. Data requirements and protocol for determining non-inferiority of insecticide-treated net and indoor residual spraying products within an established WHO policy class. Geneva: World Health Organization; 2019 (<https://iris.who.int/handle/10665/276039>, accessed 10 October 2023).
7. Technical consultation on determining non-inferiority of vector control products within an established class: report of a virtual meeting, 31 August–2 September 2021. Geneva: World Health Organization; 2021 (<https://iris.who.int/handle/10665/349446>, accessed 10 October 2023).
8. Data requirements and methods to support the evaluation of new vector control products. Geneva: World Health Organization; 2017 (WHO/HTM/GMP/2017.14; <https://iris.who.int/handle/10665/259536>, accessed 10 October 2023).
9. Technical consultation to assess comparative efficacy of vector control products: meeting report, 5 and 9 June 2023. Geneva: World Health Organization; 2023 (<https://iris.who.int/handle/10665/372841>, accessed 10 October 2023).
10. Indoor residual surface treatments for malaria transmission control in areas with insecticide-resistant mosquito populations: preferred product characteristics. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/356901>, accessed 10 October 2023).
11. WHO malaria policy advisory group (MPAG) meeting: meeting report, October 2021. Geneva: World Health Organization; 2021 (<https://iris.who.int/handle/10665/348220>, accessed 10 October 2023).
12. WHO malaria policy advisory group (MPAG) meeting: meeting report, October 2022. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/364715>, accessed 10 October 2023).
13. WHO malaria policy advisory group (MPAG) meeting report, 18–20 April 2023. Geneva: World Health Organization; 2023 (<https://iris.who.int/handle/10665/368391>, accessed 10 October 2023).
14. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/356964>, accessed 10 October 2023).



15. World Health Organization, WHO Pesticide Evaluation Scheme. Guidelines for laboratory and field testing of long-lasting insecticidal nets. Geneva: World Health Organization; 2013 (WHO/HTM/NTD/WHOPES/2013.1; <https://iris.who.int/handle/10665/80270>, accessed 10 October 2023).
16. Guidelines for testing mosquito adulticides for indoor residual spraying and treatment of mosquito nets. Geneva: World Health Organization; 2006 (WHO/CDS/NTD/WHOPES/GCDPP/2006.3; <https://iris.who.int/handle/10665/69296>, accessed 10 October 2023).
17. Implementation guidance (module 5). Semi-field methods for ITNs: experimental hut and IACT studies. Geneva: World Health Organization; 2023 ([https://extranet.who.int/prequal/sites/default/files/document\\_files/IG\\_Semi-field\\_studies\\_Nov\\_2023\\_0.pdf](https://extranet.who.int/prequal/sites/default/files/document_files/IG_Semi-field_studies_Nov_2023_0.pdf), accessed 8 March 2024).
18. Report of the twentieth WHOPES working group meeting, WHO/HQ, Geneva, 20–24 March 2017. Geneva: World Health Organization; 2017 (WHO/HTM/NTD/WHOPES/2017.04; <https://iris.who.int/handle/10665/258921>, accessed 10 October 2023).
19. Indoor residual spraying: an operational manual for indoor residual spraying (IRS) for malaria transmission control and elimination, second edition. Geneva: World Health Organization; 2015 (<https://iris.who.int/handle/10665/177242>, accessed 10 October 2023).

## Annex. Data template and data analysis code

A template for data entry of experimental hut data, files related to the calculation of sample size, analysis of data, and plotting of non-inferiority figures, and a tutorial designed to provide guidance on carrying out non-inferiority assessments for insecticide-treated products in experimental hut studies can be accessed here: [https://github.com/JDChallenger/WHO\\_NI\\_Tutorial](https://github.com/JDChallenger/WHO_NI_Tutorial).

The methodology used here matches that used by the WHO Global Malaria Programme for assessing the comparative efficacy of vector control products (1).

The data set used in the tutorial to illustrate the methodology is a synthetic data set (that is, one generated by computer simulation), rather than one from a real-world experimental hut study. The treatment arms are different types of bed nets, but the same methodology can be used for IRS. The treatment arms in the data set are named for their role in the non-inferiority assessment. The treatment arms are: an untreated control net, a standard comparator (unwashed and washed), an active comparator (unwashed and washed), and a candidate net (unwashed and washed). Therefore, there are seven treatment arms in total.

Finally, a note on the choice of the non-inferiority margin used here: the efficacies of the candidate net and the active comparator (for mosquito mortality or blood-feeding inhibition) are compared by constructing an OR and a CI. The 95% CI should then be compared to the non-inferiority margin. For mosquito mortality, the entire 95% CI must lie **above** the non-inferiority margin for the candidate product to be non-inferior to the active comparator. If this is not the case, the candidate is said to be “not non-inferior” to the active comparator. By contrast, for blood feeding, the entire 95% CI must lie **below** the non-inferiority margin for non-inferiority to be achieved.

In this work, the non-inferiority margin is set so that the candidate net efficacy is no more than 7% lower (in absolute terms) than that of the active comparator. Therefore, when assessing mosquito mortality, the non-inferiority margin is determined so that the mosquito mortality measured for the candidate net is no more than 7% lower than that measured for the active comparator in order for non-inferiority to be achieved. This means that the OR for the non-inferiority margin will vary from study to study and must be calculated for each assessment.

## Reference

1. Technical consultation to assess comparative efficacy of vector control products: meeting report, 5 and 9 June 2023. Geneva: World Health Organization; 2023 (<http://iris.who.int/handle/10665/372841>, accessed 10 October 2023).



For further information please contact:

**Global Malaria Programme**  
**World Health Organization**

20, avenue Appia  
CH-1211 Geneva 27  
Switzerland

