



Methods for evaluating efficacy of vector control products: Landscaping consultation

August 2021

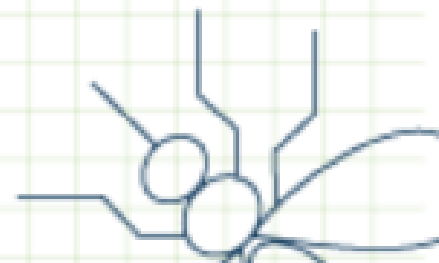
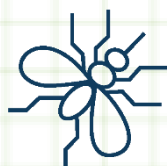
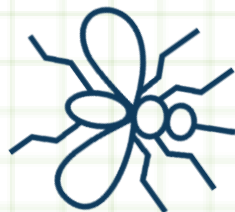


Table of contents

Table of contents	2
Acronym List	2
Executive Summary	3
Aims and approach	6
Feedback from Partners	6
Experience in evaluating product efficacy	6
Suitability of existing evaluation methods	7
Relevance of data requirements	13
Suitability of endpoints and stated thresholds for defining efficacy	13
Where responsibility for method development should sit	16
Experience in working with external testing facilities	18
Key priority areas for improvement or development	20
References	22

Acronym List

AI	Active Ingredient (insecticide or synergist)
CIPAC	Collaborative International Pesticides Analytical Council
CRO	Contract Research Organisation
EHT	Experimental Hut Trial
EPA	Environment Protection Agency
GLP	Good Laboratory Practice
IRS	Indoor Residual Spray
ITN	Insecticide Treated Net
JMPS	Joint Meeting on Pesticide Specifications
KD	Knock down, typically scored 1 hour post-exposure
MoA	Mode of Action
MR4	Reference to the MR4 Methods in <i>Anopheles</i> Research (from BEI Resources)
NGO	Non-Governmental Organisation
PQ	Vector Control Product Prequalification team of the WHO
RCT	Randomised Control Trial
WHO	World Health Organization
WHOPES	WHO Pesticide Evaluation Scheme, precursor to PQ

Executive Summary

Experience in evaluating product efficacy

- Generally, there is a feeling that WHO PQ/VCT is an improvement on WHOPEs, however there needs to be:
 - Clearer information on data requirements, particularly for new product classes
 - Guidance on acceptable test methods, or how 'novel' test methods will be validated
 - Better information distribution regarding the new PQ process to testing facilities

Suitability of existing evaluation methods

General

- Methods were developed for pyrethroids and their suitability for non-pyrethroids, and for combination products, needs to be determined, or methods adapted when required
- The development of standardised guidelines which need to be suitable for the whole range of different products is challenging, and the following is proposed as a better model (similar to the EPA):
 - The manufacturer develops a method for evaluating their product
 - This is externally reviewed for the robustness of the data and to check it supports the product claims; a judgement of the test's suitability is made on that basis.
 - If new data is generated afterwards which supports additional claims they can be added later.
- To evaluate the full impact of a product on disease transmission it is important to measure more endpoints than the standard mortality and blood-feeding inhibition, as well to look at sub-lethal effects.
- There is massive variability in the data generated by laboratory bioassays, difficulty in comparing data generated by different laboratory bioassays, and uncertainty over how to translate laboratory results to field performance.
- The requirement for multiple RCTs for a first in class chemistry or product creates additional cost and delay in getting a product to market, and stifles innovation.
 - There is not a clear alternative to this evaluation pathway.
 - Finding smaller-scale proxies for these trials could be a good alternative.
 - There needs to be an incentive to being first in class e.g., market exclusivity, concession payments from second in class products.
- The easiest route to market may be to make simple product claims against susceptible mosquitoes, although most products aim to target resistant populations.
- It can be a challenge to maintain susceptible laboratory colonies, although there are sufficient guidelines (MR4).

Nets

- There are no methods available to demonstrate the efficacy of nets which are non-insecticide treated.
- There is insufficient standardisation of the washing process for ITNs.
- There is no method available for monitoring durability of PBO nets.
 - After this consultation, work conducted by the I2I team has established by stakeholder consensus a methodology for this (Lissenden et al., 2021) which is in the process of being validated.

IRS

- In general, methods for evaluating IRS products are well established and effective.
- The main challenge with IRS efficacy testing methods is related to lack of consistency in application rate and divergence from the target dose rate.

Larvicides

- Efficacy testing guidelines were developed using old insecticides and need to be updated to evaluate slow-release products, biological agents, and growth regulators.

Repellents & space sprays

- There is no formal product class or guidelines for evaluation of spatial repellents, and products cannot be listed until a policy has been developed for their recommendation.
- The guidelines for *space sprays* are suitable, but could include additions for measuring wind speed and direction.
- The data collected for *topical repellents* make it easy to support product claims and leave no room for misinterpretation.
- WHO PQ listing is much less important for *dengue control tools* which tend to be sold commercially.

Resistance testing

- Coordination of groups working on resistance monitoring for new AIs would be invaluable.
- Further research into the impact of using MERO as an adjuvant for some compounds in the new WHO resistance monitoring manual is warranted.

Relevance of data requirements

- It is unclear how much (species/strain/location) data is sufficient to demonstrate efficacy in all relevant sites.
- Gaining PQ listing is critical for ITNs, but even after listing there should be an option to keep adding new data to the dossier, refining product claims as more data is generated.
- There are different requirements for PQ listing and for country registration which complicates evaluation.
- There is a need to consider IRM in product evaluation.

Suitability of endpoints and stated thresholds for defining efficacy

- The 80% mortality threshold for efficacy was established based on pyrethroids, and thresholds developed historically for pyrethroids may not be relevant for new products or chemistries.
- It makes more sense for thresholds to be set to match the specific methodologies, and for entomological endpoints to be chosen to fit the product. ‘Guard rails’ may be more useful than binary pass/fail criteria, to minimise the risk of losing valuable products.

“Predefining metrics for success is a recipe for failure, and you will lose chemistries which do not fit the current methodologies.”

- On the other hand, there is a confidence which comes with using established methods to apply an initial set of tests across all products in a class as a ‘bare minimum’ evaluation, and then potentially applying additional methods which are product-specific.
- WHO guidelines are not established to consider non-standard endpoints.

Where responsibility for method development should sit

- Opinions on who is responsible for method development were divided.
 - Several interviewees argued it was the manufacturers’ responsibility as they know the product/chemistry/MoA best.
 - Others saw a role for PQ, research institutions and end users.
 - It was suggested by several interviewees that it would be useful to have an equivalent body to CIPAC for entomological methods.

Experience of working with testing facilities

- A list of available testing facilities, and the testing services that they can offer, beyond the few listed on the PQ list, would be useful, as it is not always obvious where to go for local evaluation of products.

Aims and approach

This consultation aimed to collect information on companies' experiences with vector control product evaluation and the methods used to generate data for this purpose. This report summarises the discussions and main areas where interviewees felt improvement was needed.

Twenty-nine companies were invited to participate, based on a contact list of the I2I industry group provided by Robert Sloss, I2I's Industry Group Chair. Thirteen companies contributed (12 through interview, 1 through written response), five companies expressed interest but did not contribute, and eleven did not respond. Interviews were conducted via Microsoft Teams by two individuals (Rosemary Lees, Natalie Lissenden, or both), who used a list of pre-provided questions (Annex 1) to lead the discussions. Calls were recorded to assist with producing summary notes for each interview. Pertinent quotes are included to support the conclusions, but these have been anonymised and specific contributors cannot be identified in this report.

Feedback from Partners

Experience in evaluating product efficacy

- Where products vary from the standard tools containing synthetic insecticides there is a need for manufacturers to work with procurers, in-country regulators, and end users to educate them on how to effectively use and evaluate the product.
- Where products fall outside listed WHO classes, in-country regulators generally perform product evaluation, either with their own methodologies or in partnership with manufacturers.
- Generally, there is a feeling that PQ is an improvement on WHOPES. Not all GLP sites are familiar with the change from WHOPES to PQ, and so in some cases capacity building must be conducted by the manufacturer.
- Interviewees were optimistic that PQ, with some oversight from VCAG, will accept data from novel methods if they are deemed appropriate for the product under evaluation, meaning product evaluation does not need to be constrained by existing methods. Some companies are already adapting methods for internal product testing and expecting external facilities to follow provided protocols even if they differ somewhat from the methods they are used to.
- For products which target arboviruses, such as Dengue, less significance is placed on PQ listings. Historically, companies were able to sell products while awaiting WHOPES approval (which was a slow process), however this will no longer be the case under PQ.

Suitability of existing evaluation methods

“The test needs to be useful, not just make us feel better about our products.”

General points

- The mindset around product evaluation is very process-orientated, but to stimulate development of novel tools more flexibility is needed. With the EPA the process is that the manufacturer develops a method for evaluating their product, then the EPA review the robustness of the data and check it supports the product claims, how well replicated data is etc. and makes a judgement on that basis. If new data is generated afterwards which supports additional claims they can be added later. This could be a more straightforward model to use than the development of standardised guidelines which need to be suitable for the whole range of different products.
- We need to consider what the purpose of a given method is. If a company sets out to show its product works this will introduce a bias – for example, extending the exposure time might give greater mortality but may not make it a more informative or useful test.
- To evaluate the full impact of a product on disease transmission it is important to measure more endpoints than the standard mortality and blood-feeding inhibition, and as well look at sub-lethal effects. As examples, adults who are not killed by a net may still have reduced longevity or feeding inhibition, larvae which are not killed by an IGR larvicide may have reduced adult fecundity, or adults who are deterred by a space spray may be inhibited from biting. The effect of insecticide exposure on parasite development in mosquitoes is also relevant to its impact on transmission. The challenge is that the more endpoints that are evaluated the larger and more expensive evaluation becomes. It is also more complicated to regulate products which have multiple effects.
- Where products have different activity profiles, for example speed of kill, it is important to interpret data carefully when comparing products. Using different endpoints may affect which product is deemed to be more effective.
- Although products will mostly be targeting pyrethroid resistant mosquitoes, thresholds are set using susceptible strains, and so may not be suitable. For example, longer exposure times may be required for resistant strains. The easiest route to market may therefore be to make simple product claims against susceptible mosquitoes.
- On the other hand, it is a challenge to maintain susceptible laboratory colonies, required to allow the demonstration of reduced resistance ratios and benefits of new products, without them becoming hybridised and losing susceptibility, leading to delays, disruptions and additional costs. The MR4 Methods in *Anopheles* Research (from BEI Resources) provide guidance on maintaining susceptible laboratory colonies but it could be better disseminated. A second issue with susceptible strains is that most available susceptible populations are *An. gambiae*, whereas other *Anopheles* species might be more relevant depending on the target deployment site.
- There is too much focus on the insecticide itself, when the inert ingredients used in formulations may also have toxicity or synergistic effects, both in measuring product efficacy and in monitoring for resistance, which is currently done with the insecticide alone.

- There is a challenge in measuring the biopotency of combination products, and it is not clear whether insecticides should be evaluated separately or together.

Laboratory testing

- Bioassay variability leads to the requirement for many replicates to collect robust data, with the associated time and cost, and this is a big issue. Results can vary a lot between labs conducting the same study.
- There is poor correlation between lab and field results, with products failing in the lab but seeming to be effective in the field, leading to questions over the usefulness or quality of lab tests, especially where there are standardisation and quality issues.
- The New Nets Project and ESSENTIALS should help to link lab and modelling results to field results, helping us to collect more meaningful data about product efficacy.

Field testing

- The requirement for multiple RCTs for a first in class chemistry or product creates additional cost and delay in getting this product to market. Where the product has been shown to kill mosquitoes as effectively as products which are already listed, it is not clear that RCTs provide any additional information. Although data can be collected about durability and side effects, proxies could be developed for these, making smaller village-scale trials sufficient. The requirement for RCTs is more reasonable where the endpoint is novel, or where a link between entomological effect and epidemiological endpoint needs to be demonstrated.
- There are factors in epidemiological settings which are unrelated to the product but which might affect efficacy, e.g. human or mosquito behavioural change, but with high enough coverage these factors can be overcome by a product that effectively kills mosquitoes.
- There seems to have been a delay in making changes to the requirement for 3 RCTs because of the lack of a clear alternative evaluation pathway, and because of a fear of ‘Me Too’ products getting through the process without sufficient comparison to listed products.
- A specific challenge with new product types/methods of evaluation exists when the regulatory group changes over time, meaning that manufacturers need to repeat explanations, leading to additional delays.
- A WHOPES listing used to grant a producer a monopoly and protection from competitors. Now, there is reluctance to bring a first in class product/new product type to PQ, as the manufacturer must bear the burden of demonstrating epidemiological efficacy through RCTs. Consequently, there is a strong incentive to being the second in class, and this stifles innovation. There needs to be some way to credit innovators, perhaps providing an incentive to being first in class, such as market exclusivity, or somehow making it a community effort. One way would be to divide the cost of demonstrating efficacy of a product class between the first in class who makes the initial outlay and those who join the market later who pay some compensation in return for a listing. This is already done in some areas of regulatory data, with companies paying to use relevant safety data generated by another company, for example. This would not, however, remove the huge barrier to entry, meaning that large companies would still be able to maintain a monopoly.
- There are questions over the classification of new product classes: “How close is the product to an accepted paradigm, and how far do you have to go to prove your product is the same as the current paradigm?”. Where there are fixed guidelines from PQ, manufacturers will work to meet those

requirements as it is less risky, and there is no incentive for them to move away from the paradigm. Similarly, manufacturers know best how to evaluate their product but may be reluctant to take a risk to advocate for a particular new method, but there is a benefit to everyone to having better methods, and so a collaboration between those that can contribute would be valuable.

- It is easier to demonstrate a new product type using a pyrethroid, since the impact of the tool can first be demonstrated, and then the additional benefit of an alternative insecticide can be demonstrated separately. It is more complicated to try and explain and demonstrate the two together, but the inclusion of these intermediate steps may disincentivise innovation.

ITNs – General Points

- There is currently a requirement for a new dossier to be submitted if the construction, or knitting pattern, of an ITN is changed, adding a 2–3-year delay to bringing in a change which may have a clear benefit, for example to durability. Relevant measures of strength (bursting, snag, flammability) and storage stability are significant for the efficacy of an ITN, but other physical specifications such as denier are not critical and do not need to be specified and changes to these should not require a new dossier submission. The tight specifications limit the incentive to innovate.
- The insecticide used to produce an ITN needs to have been assessed by JMPS, which adds 2-3 years of delay to bringing a net to market, yet processes developed by JMPS for hazardous technical materials are not appropriate for ITNs. When assessing hazardous technical material impurities above 1g/kg or 0.1% need to be evaluated for possible toxicology, whereas the insecticide itself may only make up <1% of the net, and any impurities would be present at an even lower concentration.
- There is a fear of push back from manufacturers of products which have already been through WHOPES or PQ assessment to making changes such as removing the need for ITNs to go through JMPS assessment, even where there is no technical reason not to streamline a process.
- Novel methods to improve blood-feeding inhibition or kill mosquitoes without the use of insecticides are being developed, but there are no methods available to demonstrate their efficacy. To demonstrate the additional benefits would require demonstration of epidemiological (blood-feeding inhibition, reduced transmission) and not entomological (KD, mortality) endpoints. Some testing methodologies may not be appropriate for novel types of net, for example making holes in nets (which is a standard procedure) may completely negate the added benefit and not be a useful test, and it may not be clear what a suitable comparator net would be.

ITNs – Specific methods or products

- Cone tests and bottle bioassays, for example, were developed for pyrethroids, and may not be suitable for other chemistries, but they are recommended by PQ and WHO and so are the accepted methodologies.
- Cone tests are the easiest and cheapest way to assess bioefficacy so are widely used and accepted, but this doesn't mean that the test is adequate, effective, or that the results are consistent. Even with susceptible mosquitoes there are behavioural differences between cohorts, and repellency affects results, so that data vary between testing sites and between strains. This leads to questions from the WHO about the quality of the product, and it is difficult to distinguish true quality control issues with the net from inherent noise in the data.
- A specific issue with the cone test is that mosquitoes may rest on the cone or on the plug, an issue which could be minimised by replacing the cotton wool plug with a surface mosquitoes are less able to rest on.

- For new product types, adaptations to the cone test might make it a better predictor of net efficacy, compared to a tunnel test or experimental hut trial (EHT), and since PQ has been flexible in accepting adaptations to accommodate slower acting insecticides there is optimism that adapted methods might be accepted.
- Tunnel tests are more consistent, and better replicate wild mosquito behaviour, though there are still differences observed between strains. Several interviewees expressed interest in I-ACT chambers as an additional or alternative testing methodology, and an example of the more sophisticated behavioural assays currently under development.
- There is a lack of understanding for how to convert results from one assay type to another e.g. equivalency of results from tunnel tests and EHTs. Some test methodologies may be more suitable for some product types than others. For example, non-repellent chemistries favour longer durations of exposure on surfaces that might deliver better mortality; even between different EHT types results can differ, due to the difference in treated surface area: volume ratio, which is particularly critical in non-repellent products.
- There is no standardisation of the washing process for ITNs: the methods are different at different stages of evaluation and QC, and there is a paucity of detail in some of the guidelines. There is a CIPAC validated method for washing nets for physicochemical tests, so perhaps the other stages could be brought into line with this. The methods are not necessarily field relevant, so that a wash retention index reported in a product dossier may not match longevity in field conditions. The requirement for 20 washes may not correspond to how nets are used and washed, and handling is different between regions, though some standard is needed for consistency. There is a need to harmonise the end points measured during wash resistance testing.
- There is no method available for monitoring durability of PBO nets, for which some specific challenges exist. For example, PBO is a liquid and may weaken the fabric over time and be lost quickly from the net. With these nets now being widely adopted it would be valuable to gain more understanding of how well they last and the way that PBO is lost over time and with use.

IRS – General points

- In general, methods for measuring IRS are well established and effective with well standardised testing protocols, particularly now there is a recognition of the need for a 72-hour holding period to measure mortality for some AIs.
- There is only one IRS product class, though in practice NMCP managers differentiate between products based on MoA.
- There may be an effect of temperature on the efficacy of insecticides used in IRS, for example pyrethroids are more effective at lower temperatures, but different modes of action chemistries might be more effective at different temperatures.

IRS – Specific methods

- There is not a good correlation between results from analysing filter papers and cone bioassay results.
- The main challenge with IRS efficacy testing methods is related to lack of consistency in application rate and variation from the target dose rate. Issues with calibration of spray equipment and human error often leads to a need for repeat testing. Some facilities do not have the spray equipment required for accurate application. There is a need for better application technology to provide more even distribution so that

efficacy can be measured more reliably, though sharing of best practice and learnings about suitable spray equipment would help.

- Another challenge in testing IRS is the lack of standardisation in wall surfaces, for example there is no definition of a 'standard mud'. The ongoing relevance of mud as a testing substrate is not clear, and where it is used it presents challenges of bioavailability, which might not be a general issue with the product but a problem only with mud. Better extraction methods are needed, and perhaps some standard way of characterising the mud used in IRS studies.

Larvicides

- Guidelines for efficacy testing of larvicides were developed using old insecticides and focus on granules or liquid formulations so need to be updated to consider the methods needed to evaluate slow-release products such as tablets. They were also developed for synthetic insecticides and are not suitable for biological agents and growth regulators which have a different route of entry, activity profile and effective application rates. Solubility of new insecticides can be an issue, for example. Larvicides which need to be ingested must compete with other food sources and so must be applied at high rates, and clarity is needed on whether different products should be compared at the same concentration, or each at their own recommended application rate.
- In general, manufacturers work with testing sites to provide protocols, or feedback on their protocols, to make sure they are suitable for the product under evaluation.
- No reference standard is now available from PQ, and so VectoBac is the recommended comparator, though not universally used.

Spatial Repellents

- There is no formal product class for spatial repellents, and products cannot be listed until a policy has been developed for their recommendation. There are no formal guidelines for product evaluation, and the phased testing approach currently in the guidelines is not suitable. Methods are being developed and validated by manufacturers alongside the generation of data for the dossier.
- There is a misalignment between the research community's desire to collect data on all possible endpoints and the need for a simple test of the endpoints most relevant to product efficacy in preventing transmission, though it is not known for sure what these might be.
- WHO guidelines are less important for country registration than the methods used in country to evaluate products, but these may not be suitable and there may be limited capacity to establish new testing methodology. The data required varies significantly between countries.

Topical repellents

- There are some differences between the EU and US evaluation methods, but data is based on complete protection times in each case. The definitions are clear, so it is easy to determine whether the data supports the product claim, leaving no room for misinterpretation.

Space sprays

- Interviewees felt the guidelines for space sprays were mostly suitable.

- Wind direction is routinely measured at the start of a day's testing but may change during the day. Tightening up the guidelines with a recommendation to regularly measure wind direction and speed over the course of the day would be helpful.

Dengue vector control tools

- The process for evaluating products against *Aedes* is less well defined and the guidelines are less clear. Methods for demonstrating epidemiological impact are particularly lacking for dengue control tools.
- The WHO guidance on trapping *Aedes* (WHO, 2018) was put together by Heather Ferguson with input from industry partners. It is a good document, but not well disseminated and so not everyone knows about it or is using it.
- However, WHO listing is less important for dengue control tools which tend to be sold commercially rather than bought by NGOs or governments.
- There is a challenge in applying current IRS methods developed for a malaria setting to a dengue setting, where the living standards and economies of scale are different, affecting uptake and acceptability.
- Better application technology is needed, and spray methods need to be adapted to different settings. This is an example of where technology and methods from one disease/vector/country have been repurposed for another where they do not fit or require further optimisation.

Resistance testing

- There is an opportunity now, where new chemistries are reaching the field which mosquitoes have not been exposed to previously, to review the methods used for resistance monitoring, rather than continue the pragmatic approach that has been adopted since GPIRM. There are a few different groups and individuals working in this area now, and some coordination would be valuable.
- There is a fear that some of the new diagnostic concentrations and methods for resistance monitoring recommended by the WHO will be too complicated. ('Determining discriminating concentrations of insecticides for monitoring resistance in mosquitoes: report of a multi-centre laboratory study and WHO expert consultations' and supporting SOPs have been published since the consultation, which may have allayed this concern.)
- WHO have updated guidelines for resistance monitoring, and for some chemistries these recommend the use of MERO as an adjuvant because the WHO bottle assay does not give reproducible results without it. The use of an adjuvant is not standard, and it is not clear what effect this might have on resistance monitoring. There is no historical data with pyrethroids for example, and so it is not clear how equivalent the test will be, since it changes the way the chemistry is presented, nor what concentration should be used for the best results.
- There is a standard method from the CDC for coating bottles which is well validated, but for the new WHO discriminating dose assays using the 'WHO bottle' assay the drying time has not been validated for each chemistry.
- It would be very helpful if standard treated bottles were made available, to prevent researchers having to treat their own.
- There is an issue with the quality of papers produced in the WHO centre in Penang, with very variable dosing being observed.

Relevance of data requirements

General points

- Having a quality standard to adhere to helps to produce high-quality data and reports to submit to regulators.
- Because results in a bioassay may change depending on the mosquito population used, there may be a perceived need to replicate efficacy testing in all sites. It is unclear how much data is sufficient to represent all relevant sites, and even where the requirements of PQ are met there may still be a requirement to repeat testing to open markets in additional countries or to meet the needs of specific donors.
- Gaining PQ listing is so critical for ITNs, but even after listing there should be an option to keep adding new data to the dossier, refining product claims as more data is generated. This should be seen as part of the normal life cycle of a product, and industry should be more comfortable with bringing new data forward as it is generated, from industry, academia or in country testing.
- There are different requirements for PQ listing and for country registration. For malaria control tools, some countries simply rely on a PQ listing to use a product, whereas some have their own specific requirements. For example, Ethiopia currently requires an additional RCT to be conducted in country (though this is currently under review). It was noted that I2I's work with PQ and with the African Union should help with better collaboration and improve alignment.

"The WHO only give recommendations for specific products. They state that integrated approaches are needed, but then cannot give recommendations for combination interventions using different products together."

Suitability of endpoints and stated thresholds for defining efficacy

"Isn't it more important to be effective than it is to meet guidelines?"

General points

- You do need to have some metrics by which to judge efficacy, and some threshold to define acceptable activity, but there are not many product options for public health use and so there is a risk that if criteria are too strict, we will miss some which may be effective in reducing transmission. With hard and fast requirements around thresholds you risk losing a potentially valuable product because it just falls short; guard rails are more useful than binary criteria. There were a lot of discussion about the 80% mortality threshold in a cone test, for example:
 - There was concern about this threshold being applied to other methods and endpoints - it makes more sense for thresholds must be set to match the specific methodologies.

- Several interviewees highlighted that the 80% mortality threshold for efficacy was established based on pyrethroids, and thresholds developed historically for pyrethroids may not be relevant for new products or chemistries.
- Some interviewees felt that the 80% mortality in a cone test may be too high a bar for an ITN or IRS, where epidemiological impact might be significant below this level, or that some regulatory bodies (e.g. EPA) set unrealistic thresholds which exceed the level which would achieve an epidemiologically relevant impact.
- On the other hand, another interviewee pointed out that if the true kill rate of a deployed product is only 80% this presents a challenge for resistance management – you would want to see all exposed mosquitoes being killed by a new IRS deposit or unwashed net to avoid it driving evolution of resistance.
- It was suggested that the 80% threshold in a cone test may be a reasonable measure if it's not the only measure of efficacy, for example if more behavioural assays are also performed. The provision in the guidelines that where an ITN fails in a cone test it should be further tested in a tunnel test means that failure in the cone test is not a fatal negative for efficacy testing or durability monitoring.
- Thresholds for durability established for one product type may not be valid for others. For example, if a product is cheap and logistically feasible to replace or retreat it is less critical that it lasts for a long period of time. Current products are blocked before they get to market because they do not meet the current paradigm of what a 'successful product' looks like.
- If older mosquitoes are more susceptible to a given product, then current evaluation would underestimate its impact on transmission.
- Consideration needs to be given to the fact that activity might not increase with concentration above a certain point.
- Environmental conditions affect results and need to be controlled for consistency and to ensure they are suitable for testing products containing some insecticides. Similarly, quality control with mosquitoes used for testing is needed to minimise variability of results.
- Different insecticides and formulations may have different patterns of activity which needs to be taken into account in their evaluation, for example mortality in a cone test declining but then recovering at a later time point, efficacy may be different against different ages of mosquitoes, a longer holding period may be needed to evaluate slower acting insecticides.
- WHO guidelines are not established to consider non-standard endpoints. For example, a product that just causes blood feeding inhibition might be effective in preventing transmission but would not be deemed effective.
 - It is important that the endpoints are chosen to fit the product, and that a product is not rejected just because it does not fit a strict set of fixed criteria. For example, if you know a product is not repellent you are wasting resources in measuring repellency. Measurement of mortality after a 24-hour holding period is not suitable for more slow-acting insecticides.
 - The criteria may change as we learn more about how products work. For example, the efficacy of the roof panel of an ITN is now thought to be more important than the side panels as this is where more of the mosquito activity is.

ITNs

- There was a question around the field relevance of a 3-minute exposure to an ITN in a cone test, or any test forcing exposure to a surface, and a feeling that this parameter should be informed by behavioural studies. The tunnel test was felt to be an improvement on the cone test, allowing more natural interaction with a net.
- Some of the specifications required by PQ and JMPS such as g/m² or denier are not the most relevant for efficacy.
- Mortality is accepted as the key endpoint for an ITN. Clearer guidance is needed on what effect is expected from certain AIs, for example what is the most appropriate endpoint to use to assess the sterilising effect of pyriproxyfen. When considering a pyrethroid-PBO net there is uncertainty over what thresholds of increased efficacy in pyrethroid-PBO ITNs over pyrethroid-only ITNs are required to demonstrate product efficacy, and what entomological activity of the synergist is required to increase the impact on transmission.
- It is important to build in controls to allow you to interpret the results meaningfully, for example with a PBO net a test should include a pyrethroid-only net to show the resistance level of the mosquitoes and the added benefit of the PBO.
- Focus on a p<0.05 for a significant increase in mortality is arbitrary and looking at the effect size is more meaningful.
- Combining insecticides in a product may create an additive effect or a synergistic one, and this needs to be considered when measuring efficacy if you want to understand how the product is working.

IRS

- The 24-hour observation period was developed for pyrethroids; different holding periods might be suitable for different chemistries and should be based on the expected speed of activity of the product under evaluation.
 - However, a challenge with longer holding periods is that it can be difficult to maintain negative control mortality <20%.
- Mortality may fluctuate over time, and so a residual efficacy trial should not be stopped, or a product deemed ineffective, as soon as mortality in a cone test falls below 80% in a single time point.

Where responsibility for method development should sit

- Interviewees' opinions on where responsibility lies for method development were different and are summarised by general category below.
- Briefly, 5 companies felt the responsibility for developing and validating new methods to evaluate a product should rest solely with the manufacturers, 1 felt that the responsibility is with PQ and 1 that research or academic institutions should develop methods with PQ having a role in overseeing the process, and 3 felt that a collaboration between manufacturers and either PQ, academia or some other neutral body was the most appropriate approach. The additional role of end users in evaluating products was highlighted by one interviewee.
- Currently, it is felt that manufacturers develop methods (either on their own or in consultation with academic/research institutes), which are taken to WHO for validation, and analytical methods are then approved by and disseminated by CIPAC.

The role of manufacturers

"It can be scary to move away from the existing methods that we're all comfortable with."

- Manufacturers know the products best and should be able to determine the most suitable methods to match the MoA, but some oversight of validation and recommendations from WHO is needed.
- It is the responsibility of the manufacturer to have people in house who can design test/evaluate the product claims. There needs to be a level of trust, that when a body of evidence is taken to WHO that they trust the scientific integrity. However, industry partners vary in their capacity to develop and validate methods in house, so collaboration is needed.
- Where available methods are not suitable it is the manufacturer's responsibility to give advice on how to adapt testing methods to match their product. Manufacturers may be biased but they know their product best and have the relevant information about how and where it works.
- However, leaving responsibility with the manufacturers may lead to a lack of consistency of methods across different products. If manufacturers develop methods, there is a risk of losing comparability between products, and they are likely to propose the methods that best suit their specific product. There is also a challenge that if you develop a new method then ask external labs to validate it you need to have confidence that they will be able to reproduce results.
- One factor that could influence where responsibility lies is the scope of the method: if the method is only being used to create data for one specific dossier, then responsibility may lie with the manufacturer, with the regulator having responsibility for approving these methods. But if it is a method that needs to be applied out of the scope of that dossier there is more of a need for external validation.
- Manufacturers should develop in house methods for resistance monitoring.

The role of WHO/PQ

- Some interviewees believed it to be the responsibility of WHO and others to standardise methods so that products can be compared fairly, and to ensure the quality of data generation.
- For ITNs, for example, there are well established methods which do not exist for other product types. For novel product types, the best model may be for manufacturers to propose methods to WHO for approval.
- Historically, WHO used to hold knowledge transfer meetings, involving industry partners, researchers, and other key stakeholders. This might be a good forum to disseminate methods.
- One interviewee suggested that the WHO's focus should be on human safety data and validating product claims and less on validating methods, though there is a need to standardise methods for epidemiological trials.
- Several interviewees stated that PQ is seen as being defensive and not open to discussion of their approach. It can be hard for companies to raise questions about whether WHO guidelines are appropriate, and it is not clear how they would propose a new method to PQ. It is safer just to think how to meet the current guidelines using the most reasonable available method.
- PQ works with industry but not with researchers or academia, meaning it is difficult for new methods to be developed or methods adapted to new modes of action, particularly where there is no consensus in the community.
- A wider representation of expertise in PQ might be helpful. Some interviewees suggested that I2I might provide a forum to help bring issues to PQ in a constructive way, allowing freer discussion than can sometimes be possible between industry partners in other fora or directly with PQ.
- On the other hand, some interviewees mentioned that the WHO has been open in the past to reviewing dossier data produced using new methods and approving the methods, and PQ is now more receptive to considering new methods for evaluation than they have been.

The case for a collaborative approach

"Currently there is a risk that a new product or chemistry could be lost because it doesn't have a strong enough advocate."

- Manufacturers should be involved in innovation, de-risked by working with other partners. There are natural champions, who are advocating for specific issues and driving angles, but a more collaborative approach would be beneficial.
- There are important learnings from evaluating multiple products in the same class. Manufacturers should collaborate to determine whether a method needs to be adapted due to some specific characteristics of a particular type of product, or a specific brand. Evaluation by a wider audience, with experience in applying methods is important, though there may be an issue with patented technologies.
- It is preferable to have multiple sites testing a range of products and developing and validating methods between sites and across Africa, rather than for method validation to sit in a single manufacturer's testing facility.

- A partnership could be formed between manufacturers and a central group (probably PQ) which could share previous experiences. One suggestion was the formation of a committee formed which includes some leading scientists, industry, WHO and CDC, yet small enough to function smoothly to establish protocols for evaluation of products, and to establish gold standards.
- Alternatively, manufacturers could work with a neutral not-for-profit organisation to develop and validate methods to take to PQ for approval. It was suggested by several interviewees that it would be useful to have an equivalent body to CIPAC for entomological methods, coordinating validation studies in at least 2 sites in addition to the manufacturer's evaluation, with all methods validated against several chemistries.
- There was a lot of conversation among industry about quality, and so there should be plenty of scope for collaboration to validate methods for post-market testing.

Experience in working with external testing facilities

When asked where companies performed product evaluation testing, several mainly or solely used external partner facilities, and described their experiences with academic/research institutions, African and Asian testing sites, and Ministry of Health laboratories (who perform testing for country registrations). About half of the companies interviewed use internal testing sites for development of the product and methods for its evaluation, and external testing sites for dossier preparation and product registration, and in order to demonstrate efficacy to consumers.

General points

- Commercial organisations will approach external testing facilities for the following reasons:
 1. To avoid capital investment in laboratories and equipment
 2. For specific skills or capacity not available in house for testing
 3. To access resistant mosquito colonies, which are difficult to establish
 4. To meet a regulatory/pseudo-regulatory requirements for external testing
- Several interviewees expressed no issues in access to external testing sites
- Another felt that if field sites are contacted at least a year or a season ahead of time, or longer for semi-field sites, there is no difficulty in obtaining testing slots. Other sites can be more flexible and turn around additional experiments in a 1–2-month timeframe.
- One company described how vector control stands out from agricultural and pest control insecticide development in the difficulty of accessing rapid, high throughput testing. Even though only a relatively small number of methods are required for product evaluation, which are established and somewhat standardised, it is difficult to access testing facilities who will simply follow a provided or standard protocol and deliver raw data in a timely fashion, while managing expectations on timelines. Delays in this data provision causes delays in bringing products to market, limiting the ability to iteratively develop prototypes when data is needed within a month, and making multi centre studies very inefficient and protracted. Funding and cost are generally not limiting in getting access to testing. The reasons for delays and factors limiting capacity are felt to be as follows:

1. Capacity: Third parties do not manage expectations and often dramatically oversell. Delays are introduced into product evaluation by busy schedules of testing facilities.
 2. Resources: Standard equipment required for requested testing is not always available. Mosquito numbers is often the limiting factor, in terms of seasonality of wild mosquitoes and challenges in rearing resistant mosquitoes to large numbers. It is important for facilities to be realistic in the numbers of mosquitoes they can provide for laboratory testing, and the timelines they will need to rear enough for a study.
 3. Regulatory: Navigating the regulatory requirements to ship samples to Africa is a challenge and contributes to extending timelines. Many sites need to be more familiar with the shipping and paperwork requirements, including the Convention on Biodiversity, and ensure they have the right information and required paperwork to hand.
 4. Academic politics: Many labs are linked to academic institutes and do not see themselves as CROs. Often the requirement is to receive samples (blinded where possible), run tests according to the protocol provided, and provide raw data in a timely manner. Delays are often introduced by testing sites interpreting and analysing data and providing detailed reports, where this is not required. There needs to be a distinction made between 'research work' where detailed studies are needed to understand how a product works, methods may need to be developed or adapted, and companies may need guidance from academic institutes, and 'development work' earlier in product pipelines where reliability and fast turnaround times for standard testing are required to allow iterative improvements of, for example, formulations.
 5. GLP: The requirements of GLP slow down the completion of a study, and GLP is not always a requirement. Study designs are often tailored to existing products, but for new product types or prototypes there may be no requirement for specific controls or to test against specific thresholds, which may not have been established yet. Performing non-GLP testing in a GLP site does not reduce timelines or cost significantly. Since only later stage testing needs to be done to GLP there is a demand for non-GLP testing facilities. GLP adds to the cost of a trial, but speed is more important than cost. There is good standardisation in GLP labs, but is there a level of certification which is less onerous but that more labs could achieve and would still help improve standards? GLP adds another layer of complexity to performing the required testing, though the need for quality data is clear.
- Funding for research in ITN development has increased recently, from BMGF and IVCC, for example, but the number of test facilities has not increased. So, especially now that PQ require testing to be done at GLP sites, it can be difficult to get a testing slot. As a side effect of this demand, and the GLP accreditation itself, the cost of semi-field testing has increased. Several interviewees mentioned an increased in external testing costs. Testing in Africa is more costly than sites in Asia, though it is not clear whether this is due to GLP accreditation or variation in costs between countries. A benefit of the WHOPES process was that they organised the semi-field trials. The transition to PQ has not speeded up the time to market.
 - There is a shortage of quality labs and particularly those who are GLP accredited, with many sites not being dedicated CROs but performing external testing as an 'add-on' to their own work. The trust between industry and testing sites is important, in knowing they will respect confidentiality.
 - Facilities are well equipped to conduct testing using standard WHO procedures, but more clarification is needed regarding the evaluation of new products such as dual-AI ITNs.

- The transition from WHOPES to PQ has meant that industry can be more involved in testing done by external facilities, causing the amount of work required, but also the quality of data, to increase. Study protocols need to be prepared centrally and checked to make sure they cover all the important experimental details, and are then shared with those doing the testing with a chance to discuss in detail to make sure they are followed carefully and correctly.
- One interviewee felt that communication can be difficult with field sites, and another that quality is not consistent between all sites, and some need improvement.
- It's not clear that East and West Africa sites are enough to cover the variability of malaria settings, and malaria burden is not factored into where sites are located. There is only one GLP site in East Africa, in Tanzania, and it would be good to have more options.
- A list of available testing facilities, and the testing that they are able to offer, would be useful, beyond the few listed on the PQ list, as it is not always obvious where to go for local evaluations.

Specific methods

- Sufficient general capacity exists to perform efficacy trials of space sprays, but in most sites some capacity building and training is required, which is time consuming. Not all sites have the equipment required for detailed evaluation, for example to monitor droplet size or perform detailed calibration.
- There is not the capacity to perform tunnel tests at all sites, which can hamper certain products where bite inhibition is important from being deployed in those countries.
- EHTs are the final and best test to predict efficacy in the field but are not affordable for all stages of product development, and a laboratory test is needed that predicts efficacy in an EHT. For ITNs the tunnel test is the best available method, more affordable and avoiding the need for human exposure, but there is a paucity of facilities able to perform high throughput tunnel test studies, defined by one interviewee as a study every month or two.
- There is a need for increased capacity to test larvicides and spatial repellents. The only lab in Europe which has the ability to perform 2 adjoining room tests is Arctec, which presents difficulties in registering spatial repellent products.
- Consumer products must be registered in each country where they are going to be sold. Each will have its own minimum requirements that must be met, but regulators often have a limited amount of knowledge particularly of new product types so a lot of explanation is needed each time.
- CIPAC methods are not always available.

Key priority areas for improvement or development

Infrastructure/process

- The feeling of interviewees was that the focus of product evaluation has moved from meeting specific thresholds or criteria to the building of a dossier to highlight the range of knowledge about a product and

where the gaps are. This leaves room for interpretation around thresholds and allows for further data to be included after submission to improve understanding about the product. It should not be seen as a problem if new data is generated that deviates from a listing, but rather that it's healthy to add to our understanding. As the data builds around a product, it will become clearer which endpoints provide the most relevant information, and what the minimum testing requirements are, which will help to define a more manageable minimum data set for new products in the class.

- Several companies highlighted the benefits of conducting semi-field testing in parallel with laboratory testing, rather than testing sequentially. For example, reliance on cone tests with a 3-minute exposure and an 80% mortality threshold would lead to potentially useful products being rejected, where efficacy would be seen in tunnel tests or EHTs. For straightforward products it would also be beneficial to create the package of data in parallel to shorten the timeline to market.
- There is a need for better coordination, and a neutral stakeholder may be useful to oversee method validation, similar to CIPAC but for entomological methods.
- One interviewee proposed the removal of the requirement for JMPS evaluation as part of PQ, leaving PQ to set the specifications for ITNs and allow for a more efficient product evaluation process and quicker route to market.
- Two companies mentioned that form registration reciprocity between countries, and between PQ and other regulatory bodies such as the EPA, would remove the need to repeat testing where the requirements are the same.
- One company requested that PQ provide a clear document laying out the process for application and the data requirements, with an idea of timelines.
- Capacity building in field sites was highlighted as a priority by one company.

Data requirements

- A reduction in the scale of 'Phase 3' testing was prioritised by one company, who saw it as designed just to address any differences or unresolved questions for that specific product compared to existing products which have already been demonstrated to be effective in epidemiological trials.
- One company mentioned as a priority that the definitions of success and the methods used to test them are matched to the product, rather than requiring products to fit a narrow prescriptive definition of efficacy.

New or better methods

- Maintenance of susceptible colonies for 'Phase 1' testing was a priority issue for one company.
- More thinking is needed around managing resistance; resistance management is not currently required in PQ dossiers, and PQ are not well placed to evaluate claims about sub-lethal effects or resistance management effects.
- The bottle bioassay was not developed to look at two AIs – does this preclude usefulness of this method to look at resistance to combinations? The results of DD testing might not correlate with the way resistance might develop to an actual product, and we don't know how to interpret results to inform operational decisions or improved product development.

- Development of good (not too costly) methods for quantifying mosquito numbers in the field, to demonstrate efficacy of an intervention in the field was highlighted as a priority of one company.
- More consistent spray systems were an important need highlighted by one interviewee.
- The entire process of lab testing of IRS needs to be reviewed and improved, according to one interviewee, as there are a lot of areas of uncertainty and inconsistency.

Access to external testing

- Increased reliability and speed of data generation in external testing sites is needed, to allow iterative product development, which must be high quality but does not have to be done to GLP. A list of non-GLP testing facilities would be useful.
- More facilities able to do tunnel testing, without animal hosts if necessary, are a requirement, as is reliable access to more resistant lab strains of mosquitoes and to more field populations.
- Access to a pilot scale net formulating, extruding and weaving facility was mentioned as important.

References

Lissenden, N., Armistead, J. S., Gleave, K., Irish, S. R., Martin, J. L., Messenger, L. A., Moore, S. J., Ngufor, C., Protopopoff, N., Oxborough, R., Spiers, A., & Lees, R. S. (2021). Developing Consensus Standard Operating Procedures (SOPs) to Evaluate New Types of Insecticide-Treated Nets. *Insects* 2022, Vol. 13, Page 7, 13(1), 7. <https://doi.org/10.3390/INSECTS13010007>

WHO. (2018). Efficacy-testing of traps for control of aedes spp. mosquito vectors. <https://apps.who.int/iris/handle/10665/275801>

Annex 1 – Questions provided to participants of the consultation to prepare written responses or as preparation for online conversation



Landscaping the entomological methods needs for evaluation of vector control tools: A consultation exercise for I2I

Many thanks for your agreement to take part in this landscaping exercise. We have drafted the following questions as a starting point for discussion, but feel free to share any other experiences you have had in this field or any ideas on improvements or new developments that are needed.

Your product and your experience evaluating its efficacy

- Based on your experience, what methodological issues have you encountered in the evaluation of your product(s) which, if addressed could:
 - Provide more relevant data to describe the “entomological effect” of your product?
 - Provide more relevant data to assure the quality of your product?
 - Provide more consistent data for product evaluation? (reduced variability)

Suitability of existing evaluation methods

- Considering what you know about the existing evaluation methods (e.g. WHOPES testing guidelines, PQ evaluation), how appropriate are they for your product(s)? Are there any instances where the right method is not available?
- Are there areas for improvement that would help streamline or improve product evaluation, or produce more useful or relevant data?

Data requirements - technical

- Do you think the data requirements for evaluating a product are relevant to its efficacy? e.g. is it useful to measure mortality, bite reduction, and repellency in an experimental hut trial or are some data redundant for some product types?
- Do you feel that the required endpoints are valid and, if not, what would you change/add/remove? e.g. Is 80% mortality in a cone test a suitable threshold for acceptable efficacy?

Data requirements – responsibility for method development

- Do you think it is the manufacturers’ responsibility to develop and evaluate new methods where it’s needed for a new product? If not, whose is it?
- How do you think we can best work as a community to determine the most suitable methods and most relevant data requirements?

Data generation

- Do you conduct all your efficacy testing in house, or do you also/instead rely on other labs to conduct testing?
- Do the available facilities meet all your testing requirements, or are there services you can’t get or improvements you would like to see?

Priorities for improvement

- If you could change just ONE thing relating to evaluation methods, what would it be?



Innovation to Impact
Pembroke Place
Liverpool L3 5QA UK

contact@innovation2impact.org
+44 151 702 9308