Global **Malaria** Programme

Informal consultation on methodology to distinguish reinfection from recrudescence in high malaria transmission areas

Report of a virtual meeting, 17–18 May 2021

World Health Organization

Global **Malaria** Programme

Informal consultation on methodology to distinguish reinfection from recrudescence in high malaria transmission areas

Report of a virtual meeting, 17–18 May 2021

World Health Organization

Informal consultation on methodology to distinguish reinfection from recrudescence in high malaria transmission areas: report of a virtual meeting, 17–18 May 2021

ISBN 978-92-4-003836-3 (electronic version)

ISBN 978-92-4-003837-0 (print version)

Suggested citation. Informal consultation on methodology to distinguish reinfection from recrudescence in high malaria transmission areas: report of a virtual meeting, 17–18 May 2021. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see https://www.who.int/copyright.

This publication contains the report of an informal consultation and does not necessarily represent the decisions or policies of WHO.

Contents

ABBREVIATIONS

Africa CDC	Africa Centres for Disease Control and Prevention
AL	artemether-lumefantrine
AmpSeq	amplicon sequencing
AS	artesunate
ASAQ	artesunate-amodiaquine
ASSP	artesunate+sulfadoxine-pyrimethamine
bp	base pair
CE	capillary electrophoresis
COI	complexity of infection
DP	dihydroartemisinin-piperaquine
glurp	gene of glutamate-rich protein
GMP	Global Malaria Programme
HeOME	heterozygome
IBC	image barcode
SNP	single nucleotide polymorphism
TEG DER	Technical Expert Group on Drug Efficacy and Response
TES	therapeutic efficacy study
MIP	molecular inversion probe
MMV	Medicines for Malaria Venture
MOI	multiplicity of infection
msp1	gene of merozoite surface protein 1
msp2	gene of merozoite surface protein 2
PCR	polymerase chain reaction
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
US-CDC	United States Centers for Disease Control and Prevention
WGS	whole genome sequencing
WHO	World Health Organization

EXECUTIVE SUMMARY

This data-driven meeting assessed the advantages and disadvantages of changing the way in which recurrences are differentiated as reinfection or recrudescence following the treatment of uncomplicated *Plasmodium falciparum* malaria. This has implications for the evaluation of antimalarial efficacy in therapeutic efficacy studies (TESs), as well as in regulatory trials for the development of new antimalarial drugs.

Guidance for discriminating *P. falciparum* recrudescence from reinfection was published by the World Health Organization (WHO) in 2008 *(1)*. Blood samples collected pre-treatment (day 0) and on the day of treatment failure (day X) are compared using three markers: genes for merozoite surface protein 1 (*msp1*), merozoite surface protein 2 (*msp2*) and glutamate-rich protein (*glurp*). A standard genotyping methodology is recommended, including the use of capillary electrophoresis (CE). The decision algorithm (referred to as the WHO/MMV algorithm) requires the three markers to be genotyped and analysed in sequence, starting with *msp1*, followed by *msp2* and then *glurp*, and stopping once a marker has classified the paired samples as a reinfection. In this case, if any marker indicates reinfection, the recurrence is deemed a reinfection. An alternative approach has been suggested (termed the 2/3 algorithm) whereby *msp1/msp2* are evaluated, and only in cases where these two markers are discordant, *glurp* is used as the deciding factor. In this case, even if *msp1* indicates a reinfection, if *msp2* and *glurp* indicate a recrudescence, the recurrence is deemed a recrudescence.

The consultation examined evidence around changes in the genetic markers used to determine the relatedness of initial and recurrent parasites, as well as the algorithms used to analyse these markers to classify recurrences as either a recrudescence or reinfection. In particular, the panel examined the applicability of recent advances in genotyping and analysis. The meeting focused on areas of high transmission in Africa because the high multiplicity of infection (MOI; i.e., the number of concurrent clones in an infection) and high reinfection rates in such areas complicate the discrimination of recrudescence from reinfection.

Summary conclusions

The different methodologies for genotyping and analysis used to differentiate recrudescence from reinfection all have advantages and limitations, and clearly give different results. This has important consequences because, in some cases, the difference in the number of recurrences classified as recrudescence drives the efficacy rate below 90%, which is the currently recommended threshold requiring a change in treatment policy. In addition, this may affect decisions during drug development and adoption of new treatments, where a 95% efficacy threshold is recommended.

The panel considered which methods are most likely to be closest to the 'true' values for reinfection and recrudescence. The most robust and reliable genotyping method is amplicon sequencing (AmpSeq). However, the capacity to apply this technology in Africa needs to be strengthened before this method can be adopted as a standard. This could be achieved by capacity building in African countries and/or offering deep sequencing at core facilities(s) in Africa or elsewhere, which could process samples and return data to countries for analysis – the dual aim being to ensure that drug efficacy is accurately measured and that countries retain ownership of their data. The examined data clearly indicate that *glurp* is not an ideal marker. Therefore, until AmpSeq implementation is feasible, as an interim solution, *glurp* should be replaced with alternative markers. Microsatellites with a diversity relevant to the study site location appear to be the most

feasible and reliable option. For the transition period, data from the new methods and the current *msp1/msp2/glurp* markers should be reported to enable historical comparison.

In terms of analysis, the 2/3 approach is comparable to the WHO/MMV algorithm in low to moderate transmission settings, but may overestimate recrudescence rates for artemether-lumefantrine (AL) in high transmission settings. The 2/3 method ignores data from the 'third' discordant marker, which is a reasonable strategy when *msp1* and *msp2* agree; however, ignoring the other markers is not necessary if microsatellites are used in place of *glurp*. Match-counting is simple to use and does not disregard information from any markers; however, this method may underestimate recrudescence. The panel also considered Bayesian analysis, which has been applied to TESs conducted by the United States Centers for Disease Control and Prevention (US-CDC). The main advantage of this approach is that it provides a measure of uncertainty around the results. However, validation of the model used for Bayesian analysis is needed, including when AmpSeq data are used to distinguish between reinfection and recrudescence. Furthermore, the feasibility of using Bayesian analysis at the country level needs to be carefully assessed.

It is unclear what impact a change in methodology would have on the drug efficacy thresholds that are used to establish antimalarial treatment policy in countries and to support new drug approvals. Therefore, it is anticipated that a transition period will be required to generate comparative data for the different uses of this information. This should be considered in the context of expanding expertise and capacity in Africa for next generation sequencing. Furthermore, the broader trend towards genetic analysis of infectious diseases suggests that such approaches are likely to become more widely accepted and better understood over the next few years.

Recommendations

1 As an interim solution, *msp1* and *msp2* should continue to be used, but *glurp* should be replaced with one microsatellite from the following: *Poly-α*, *Pfpk2* and *TA1*. For simplicity and reasons of practical implementation, WHO/MMV match-counting (3/3) should be maintained as the primary analysis methodology for reporting and policy change. Bayesian and 2/3 algorithms may be applied for evaluation and comparison, but not for primary reporting. These methods should be applied in both low to moderate and high transmission settings in Africa. Outside Africa, the current method (*msp1/msp2/glurp*) should still be applied.

2 For a transition period, data should be analysed and reported using both the current (*msp1/msp2/glurp*) and new (*msp1/msp2/*microsatellites) methods to enable historical comparison and to understand the implications of the new methods in terms of thresholds for treatment policy change and introduction of new antimalarial drugs. Countries are not required to use *glurp* if they have already switched microsatellite markers or *msp1/msp2/*microsatellites. Data transparency will be critical for comparative analysis and to provide a database for analytical methodology development.

3 As a medium-term (five-year) target, AmpSeq should be evaluated in parallel across Africa and outside Africa to compare it with current methods at sites and to validate whether this genotyping methodology should be adopted as the standard. A simple match-counting algorithm complemented by a Bayesian algorithm could be paired with this approach, but more comparative data are needed to inform the recommendation on the algorithm for analysing AmpSeq results.

Research needs

- Identify the most polymorphic microsatellite markers per country or region.
- Develop automated pipelines to enable TES design and interpretation in order to increase the feasibility of new methods.

- Explore alternative methods of model validation and different modelling approaches.
- Define a systematic process for validating genotyping methods and data analysis.
- Compare different genotyping and analysis methodologies using the same datasets.
- Evaluate the impact of Bayesian analysis on recrudescence rates in areas of high transmission, given the trend in increased recrudescence rates using this method, to determine whether this is an artefact of the method, whether there is some reason for higher failure rates in these areas (e.g., a high MOI may be more challenging for antimalarial drugs to clear), or whether there is emergence of true antimalarial resistance that needs rigorous confirmation.
- Evaluate the interaction between increases in the sensitivity of methods and the detection of gametocytaemia (rather than recrudescence).

1. RATIONALE

The WHO Global Malaria Programme (GMP) organized an informal consultation bringing together researchers and partners to review WHO guidance on the methodologies used to differentiate reinfection from recrudescence in *P. falciparum* antimalarial efficacy trials in sub-Saharan Africa. The methodologies used can affect the estimated drug failure rates in TESs, and thus has implications for recommendations around drugs for the treatment of *P. falciparum* malaria.

2. BACKGROUND

Evaluating antimalarial drug efficacy in uncomplicated *P. falciparum* malaria is problematic because of the occurrence of new infections, particularly in areas of high transmission. This can lead to inaccurate estimates of therapeutic efficacy rates. Methods that accurately discriminate between recurrences caused by therapeutic failure (recrudescence) and reinfection are therefore needed. However, there is no gold standard for evaluating the ability of the different methodologies to correctly classify reinfection and recrudescence.

The challenges of classifying recrudescence versus reinfection are different in low to moderate and high transmission settings. In low to moderate transmission settings, parasite diversity is often minimal. Therefore, reinfections may be misclassified as recrudescence. High transmission settings present different challenges, given the relatively higher levels of both polygenomic infections in the population and the higher risk of reinfection during the study period. High infection rates increase the possibility that a reinfection will share alleles with the initial infection, causing the reinfection to be misclassified as a recrudescence. Conversely, a recrudescence might be misclassified as a reinfection because, with an MOI >1, low-frequency alleles may not be observed for a particular locus or may be erroneously classified as a different allele due to polymerase chain reaction (PCR) errors *(2,3)*. Consequently, there are biological limitations to the accuracy of drug efficacy estimates due to adjustments for reinfection. The sensitivity of the genotyping methods, the diversity of the markers used, and the classification algorithms applied all affect the degree of uncertainty around these estimates.

There are several different genetic markers and analysis techniques that can be used to differentiate individual parasite infections. Depending on the research question, however, each has both advantages and limitations. Specific laboratory techniques may be more appropriate for given settings and transmission levels. The cost and complexity of each technique and whether it can be performed in most countries or only in specialized research laboratories are key considerations. Recommended techniques and decision-making strategies should be feasible and simple, and provide operationally relevant information for malaria control programmes.

Current guidance

The current guidance for discriminating *P. falciparum* recrudescence from reinfection was published by WHO in 2008 *(1)*. Blood samples collected pre-treatment (day 0) and on the day of treatment failure (day X) are compared using three genetic markers: *msp1*, *msp2* and *glurp*. A standard PCR genotyping methodology is recommended, including the use of CE. The decision algorithm (referred to hereafter as the WHO/MMV method) requires the three markers to be genotyped and analysed in sequence, starting with *msp1*, followed by *msp2* and then *glurp*, and stopping once a marker

has classified the paired samples as a reinfection. With this method, as soon as a new infection is identified with one marker, the overall outcome is a new infection. WHO has abandoned the sequential analysis and recommended genotyping the three markers systematically, enabling additional and comparative analyses.

Meeting of the Technical Expert Group on Drug Efficacy and Response, June 2017

This 2008 guidance was reviewed in 2017 at a meeting of the Technical Expert Group on Drug Efficacy and Response (TEG DER). The TEG DER also explored the challenges related to the use of *msp1*, *msp2* and *glurp*. Of the three markers, *glurp* was thought to be the least useful, as it may increase the proportion of true recrudescent infections classified as reinfections during clinical trials in high transmission settings. However, *glurp* could still be a valuable marker in regions of low malaria endemicity. Despite the limitations of *msp1* and *msp2*, continued use of these markers should provide the minimum essential information for correcting the data to reflect an accurate estimate of drug efficacy. While microsatellites may have their own limitations, these genotyping data were considered to provide similar results as those found using *msp1*, *msp2* and *glurp (4)*.

The TEG DER proposed a revised algorithm to distinguish between *P. falciparum* recrudescence and reinfection, termed the 2/3 algorithm. In this case, both *msp1* and *msp2* are genotyped for all samples. If *msp1* and *msp2* yield congruent results, the result is reported as the overall result of the genotyping (reinfection or recrudescence). However, if there is discordance between the *msp1* and *msp2* markers, a third marker should be genotyped. This marker could be *glurp* or another validated highly diverse gene. If this third marker supports one of the two previous results, the majority result (2/3) is reported as the overall result *(5)*. However, it was recognized that different groups and organizations use different markers, and that even when the same markers are used, genotyping and analysis methods may be inconsistent.

Informal consultation

The aim of this consultation was to collate comparative data, to give experts the opportunity to provide advice on changes needed to currently recommended methodologies, and to provide direction for the development of tools and methods for future use. Through this meeting, WHO sought to address research results generated by different methodologies for distinguishing between recrudescence and reinfection. Some of these research methodologies need further validation before they can be recommended for use in routine TESs. While waiting for the validation of new tools, the key expected outcome was an agreement on an interim common position on the genetic markers and analysis algorithm to be used by both countries and researchers to differentiate recrudescence from reinfection.

3. INTRODUCTION AND DECLARATIONS OF INTEREST

The list of participants is provided in Annex 1. All invitees attended the meeting, except for Daouda Ndiaye who sent apologies. Organizations invited as observers were The Global Fund to Fight AIDS, Tuberculosis and Malaria, Bill & Melinda Gates Foundation, Medicines for Malaria Venture (MMV), and the United States Agency for International Development (USAID). The meeting agenda is provided in Annex 2.

Only members of the expert panel and GMP Secretariat attended and took part in the final discussion and formulation of key recommendations. All 11 members of the expert panel submitted Declarations of Interest, which were assessed by the GMP Secretariat at WHO. Of the 11 experts on the panel, six declared no interest and five reported

interests that were summarized in a report, which is available upon request. Based on this assessment, all the experts were able to fully participate in the discussion and deliberation of the consultation. Discussions were conducted under a confidentiality agreement signed by all participants.

4. OBJECTIVES

The primary objective was to discuss the most appropriate genotyping and analysis methods for reporting recrudescence and reinfection rates following antimalarial treatment of *P. falciparum* malaria in sub-Saharan Africa, with special focus on high transmission regions.

Specific objectives

- Review available data and assess the advantages and disadvantages of:
  - changing the markers used to differentiate recrudescence from reinfection;
  - changing the algorithms used to classify recrudescence and reinfection.
- Assess transmission settings where a change in the current methodology could improve the precision of classifying recurrent *P. falciparum* as recrudescence or reinfection.
- Discuss alternative tools for future use and suggest research needed to validate these tools.

5. PROCESS AND PRESENTATION

GMP convened this consultation to review the evidence and advise WHO on the most appropriate methods to differentiate recrudescence from reinfection in sub-Saharan Africa, with special focus on high transmission regions.

Background documents

In preparation for the meeting, WHO collected relevant publications on the topic, and manuscripts were shared by presenters (see Annex 3).

Presentations

Presentations, followed by a brief discussion, were made by Ingrid Felger and Christian Nsanzabana (Swiss Tropical and Public Health Institute, Basel, Switzerland); Eric Halsey, Mateusz Plucinski, and Venkatachalam Udhayakumar (US-CDC and President's Malaria Initiative, Atlanta, United States of America); Ian Hastings (Liverpool School of Tropical Medicine, Liverpool, United Kingdom of Great Britain and Northern Ireland); Jonathan Juliano (University of North Carolina, Chapel Hill, United States of America); Didier Ménard (Institut Pasteur, Paris, France); Daniel Neafsey (Broad Institute of MIT and Harvard, Cambridge, United States of America); and Pascal Ringwald (WHO, Geneva, Switzerland). Summaries of these presentations are included in Annex 4.

6. EVIDENCE AVAILABLE AND REVIEWED

Markers and genotyping

- In the 2008 document, WHO acknowledged the potential use of other markers as an alternative to *msp1*, *msp2* and *glurp*, but did not specifically recommend microsatellites *(1)*.

- For more than a decade, the US-CDC Malaria Laboratory has used neutral microsatellite genotyping, arguing that this approach can include more markers, has higher discriminatory power, and avoids confounding owing to selection pressure on the markers *(6)*.

- Two African countries (Mali and Uganda) use a combination of markers: *msp1*, *msp2*, and between one and four microsatellites including *CA1*, *TA87* and *TA99* (Mali), and *TA40, TA60, TA81* and *PfPK2* (Uganda).

Analysis algorithms and modelling

- Several molecular correction algorithms were assessed using *msp1, msp2* and *glurp* data provided by WHO from a large trial carried out in Rwanda and Cambodia. The model developed by the Liverpool School of Tropical Medicine indicated that the 2/3 algorithm was the most appropriate. Depending on the parameters and assumptions, a two-fold increase in the recrudescence rate was reported compared to using the WHO/MMV algorithm. There was no significant difference in the recrudescence rate using data from Cambodia, a low transmission setting *(5)*.

- Using the same model developed by the Liverpool School of Tropical Medicine, simulation data were used to compare the two main approaches in order to analyse paired microsatellite data *(7)*:
    - match-counting – all detectable alleles in all tested loci are compared between day 0 and day of recurrent infection;
    - Bayesian algorithm – a probabilistic approach quantifies the probability that a sample is either a recrudescence or a new infection.

- WHO/GMP re-analysed all PCR corrections for which *msp1, msp2* and *glurp* data were collected to compare the WHO/MMV and the 2/3 algorithms in different transmission settings.

- In collaboration with the Institut Pasteur Paris, WHO/GMP is evaluating the replacement of the *glurp* marker with one or more microsatellite markers or AmpSeq, as previously suggested *(7,8)*.

7. CONCLUSIONS AND RECOMMENDATIONS

The expert panel addressed the following key questions, with reference to the situation in sub-Saharan Africa, and put forward the following conclusions and recommendations for consideration:

1. What are the advantages and disadvantages of each of the following possible changes to the current methodology (considering simplicity and applicability at the country level)?

Changing markers

- Replacing one or more of the currently recommended markers (in particular *glurp*) with alternatives?
- Basing the reinfection/recrudescence classification on *msp1, msp2* and microsatellite data?
- Basing the reinfection/recrudescence classification on *msp1* and *msp2* data and AmpSeq?
- Basing the classification on microsatellite data only?
- Basing the classification on AmpSeq only?

Changing the analysis algorithm

- Changing to a 2/3 algorithm?
- Changing a simple match-counting algorithm to analyse microsatellite data using a Bayesian algorithm?
- Using another algorithm?

Conclusions

Each methodology for genotyping and analysis used to differentiate recrudescence from reinfection has advantages and limitations, and clearly give different results. This has important consequences because, in some cases, the difference in the number of recurrences identified as recrudescence drives the efficacy rate below 90%, which requires a change in treatment policy. In considering changing the current methodology, it is important to consider the following:

- A change in methodology could result in consistently low efficacy estimates for drugs in use or development. This could require revision of the threshold level for drug efficacy that prompts antimalarial treatment policy change in the country.
- It is important to balance the need for a methodology that can be carried out broadly across malaria-endemic countries and that will provide the minimum amount of information needed to inform decisions. Highly specialized laboratories can serve as reference centres to provide additional or more detailed information.
- AmpSeq is considered the method likely to be used for genotyping once this technology is more widely available. The transition to AmpSeq is anticipated once the method is validated and capacity is widely available.

The data examined clearly indicate that *glurp* is not an ideal marker and more discriminatory markers are needed. Microsatellite markers can provide an interim solution and a validated set of microsatellite markers should be evaluated for each country/region.

TESs are expensive and have important implications for drug treatment policy and drug development. It is critical to use methods that are reliable and reproducible across a range of transmission intensities. While underestimating recrudescence could lead to a delayed recognition of spreading drug resistance, it is important to recognize that overestimating recrudescence can also have negative consequences for patients and malaria control efforts. Prematurely and mistakenly changing a drug could mean less access to treatment and increased cost to malaria programmes.

Recommendations

The genetic comparison of initial and recurrent parasites to estimate recrudescence and reinfection rates remains a valuable component of TESs. However, the panel suggested some changes to the current approach.

- *msp1* and *msp2* should continue to be used, but the *glurp* marker should be replaced with one microsatellite from the following: *Poly-α*, *Pfpk2* and *TA1*. From these three microsatellites, the marker with the greatest diversity in the study area should be selected. A recrudescence would be classified if three loci match (3/3 approach).
    - In polyclonal infections, a match requires only one allele to be the same between the initial and recurrent infection samples at each locus; although this is not a change to the previous guidelines, it is an item often misinterpreted.
    - Combinations of *msp1/msp2/*microsatellites are currently used in Mali and Uganda. This approach would ideally use CE, but could also be done with agarose gel electrophoresis with validation.
- For simplicity and reasons of practical implementation, WHO/MMV match-counting (3/3) should be used as the primary analysis method and for policy change. The Bayesian and 2/3 analysis approaches should be additionally applied, whenever possible, for validation and for more thoroughly evaluating the advantages and limitations of these methods compared to simpler approaches.
- For a transition period, data should be analysed and reported using both the current (*msp1/msp2/glurp*) and new (*msp1/msp2/*microsatellites and AmpSeq) methods to enable historical comparison and to understand the implications of the new methods, for example, in terms of thresholds for treatment policy change.
    - This transition period would permit cross-validation of detection methodologies.
    - This transition period would support the move from length polymorphism to microhaplotype methods, and provide evidence to evaluate simple counting analysis versus more complex probabilistic analysis.
    - Countries are not required to use *glurp* if they have already switched microsatellite markers or *msp1/msp2/*microsatellites.
- There is a need for transparency on the method used, and methods should be clearly described when reporting the data.
- All genotyping data should be reported in full to enable comparative and exploratory analyses.

2. In which transmission settings would a change in methodology improve the classification of recurrent *P. falciparum* infection as recrudescence or reinfection?

Conclusions

The same general methodology should be applied to all transmission settings within sub-Saharan Africa. However, there needs to be careful investigation of how the different methods perform at different transmission levels. Outside Africa, the methods could remain unchanged as per the current guidance, and countries could transition directly to AmpSeq (see below).

Recommendations

For sub-Saharan Africa, all studies should use *msp1* and *msp2* with a panel of two to three informative microsatellite markers (such as *Poly-α*, *Pfpk2* and *TA1*). Analysis should use the match-counting method. The details and rationale for this is described above.

3. Are there other tools in development that could be used for this purpose in the future?

- If so, what are the advantages and disadvantages compared to the recommended methodology?
- What research is needed to validate new tools?

Conclusions

AmpSeq appears to be the most robust and reliable genotyping method that could become an affordable and field-deployable technology in the medium term (within five years). However, current capacity in Africa is insufficient to recommend this technology as a standard. This could be addressed by capacity building in African countries or, because AmpSeq is most cost-effective with high throughput, by sending samples to African regional facilities or to central laboratories outside of Africa. However, care must be taken to ensure continued ownership for implementers. This could be achieved by processing samples in regional African centres or at central facilities, but siting the analysis and interpretation of data in the country of origin. Whole genome sequencing (WGS) and molecular inversion probes (MIPs) are perhaps further in the future in terms of deployment in Africa, but could be used in research settings.

From a scientific perspective, the analysis method provides a degree of uncertainty that is associated with the results. Methods that include uncertainty, such as Bayesian analysis, need further validation. There are additional concerns around how these methods can be practically implemented in countries. The communication of uncertainty in the Bayesian analysis may be challenging in a public health context. More comparative data with simpler analysis methods would enable an evaluation of the relative strengths and limitations of this approach.

The trend is moving towards genetic analysis of infectious diseases, and these approaches are likely to become more widely accepted and better understood over the next few years.

Recommendations for the future

- AmpSeq should be targeted as the next method for evaluating recrudescence and reinfection rates in TESs.
- These data could be analysed using simple match-counting at first, although the level of uncertainty for this method is not defined.
- The addition of a Bayesian algorithm provides a measure of uncertainty around the results and is tolerant to errors.
    - The validity and evaluation of the advantages of a Bayesian method need to be demonstrated at different transmission levels.
    - Automation could potentially compensate for the increased complexity of this approach.
- Drug development studies may offer an opportunity for initial application of these new methods in parallel with the standard methodology. Such studies would also generate comparative data for validation and provide evidence to regulatory authorities to increase acceptance of this approach.

Research recommendations

- Identify the most polymorphic microsatellite markers per country or region.
- Develop automated pipelines to enable TES design and interpretation in order to increase the feasibility of new approaches.
- Explore alternative methods of model validation and different modelling approaches.
- Define a systematic process for validating genotyping methods and data analysis.
- Compare genotyping and analysis methodologies using the same datasets.
    - Evaluate the performance of the different approaches in different transmission settings.
    - Use longitudinal data to confirm signals that suggest the emergence and spread of drug resistance.
    - Use fully validated drug resistance markers to improve the confidence in methodology for detecting recrudescence caused by drug resistance (e.g., compare data from day 0 and day of recurrent infection samples).
    - Use drug plasma level or ex vivo data to confirm if 'failing' drugs have reduced efficacy because of poor adherence or absorption, not because of resistance.
- Evaluate Bayesian analysis to more fully investigate initial findings of a trend of increased recrudescence rates in areas of high transmission in order to determine whether these findings represent a methodological artefact, whether there is some reason for higher failure rates in these areas (e.g., a high MOI may be more challenging for antimalarial drugs to clear) or whether there is emergence of true antimalarial resistance that needs rigorous confirmation.
    - Looking at the MOI versus recrudescence rate for the different markers/analysis methods could be informative, as well as for drug types and transmission levels.
- The interaction between increased sensitivity of methods and the detection of gametocytaemia (rather than recrudescence) requires evaluation.

8. REFERENCES

1. Methods and techniques for clinical trials on antimalarial drug efficacy: genotyping to identify parasite populations. Geneva: World Health Organization; 2008 (https://apps.who.int/iris/bitstream/handle/10665/43824/9789241596305_eng.pdf).

2. Greenhouse B, Dokomajilar C, Hubbard A, Rosenthal PJ, Dorsey G. Impact of transmission intensity on the accuracy of genotyping to distinguish recrudescence from new infection in antimalarial clinical trials. Antimicrob Agents Chemother. 2007;51(9):3096-103. https://www.ncbi.nlm.nih.gov/pubmed/17591848

3. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. Am J Trop Med Hyg. 2006;75(5):836-42. https://www.ncbi.nlm.nih.gov/pubmed/17123974

4. Mwangi JM, Omar SA, Ranford-Cartwright LC. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing *Plasmodium falciparum* infections from reinfections in Kenya. Int J Parasitol. 2006;36(3):329-36. https://www.ncbi.nlm.nih.gov/pubmed/16442537

5. Jones S, Kay K, Hodel EM, Chy S, Mbituyumuremyi A, Uwimana A, et al. Improving methods for analyzing antimalarial drug efficacy trials: molecular correction based on length-polymorphic markers *msp-1, msp-2*, and *glurp*. Antimicrob Agents Chemother. 2019;63(9). https://www.ncbi.nlm.nih.gov/pubmed/31307982

6. Plucinski MM, Morton L, Bushman M, Dimbu PR, Udhayakumar V. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrob Agents Chemother. 2015;59(10):6096-100. https://www.ncbi.nlm.nih.gov/pubmed/26195521

7. Jones S, Plucinski M, Kay K, Hodel EM, Hastings IM. A computer modelling approach to evaluate the accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of antimalarial drug efficacy. Antimicrob Agents Chemother. 2020;64(4). https://www.ncbi.nlm.nih.gov/pubmed/31932376

8. Nyachieo A, C VAN Overmeir, Laurent T, Dujardin JC, D'Alessandro U. *Plasmodium falciparum* genotyping by microsatellites as a method to distinguish between recrudescent and new infections. Am J Trop Med Hyg. 2005;73(1):210-3. https://www.ncbi.nlm.nih.gov/pubmed/16014861

ANNEX 1. AGENDA

MONDAY, 17 MAY 2021		
14.00–14.10	Welcome	P. Alonso, Director GMP D. Wirth, Chair
14.10–14.15	Declarations of Interest	P. Ringwald
14.15–14.40	Objectives of the meeting Presentation of the WHO recommendations and conclusions of the TEG meeting 2017	P. Ringwald
14.40–15.30	New evidence on *msp1, msp2* and *glurp* as markers for reinfection and recrudescence	I. Felger, C. Nsanzabana
15.30–16.20	Microsatellites and Bayesian algorithm in Africa: added value of microsatellites, validation of Bayesian algorithm, comparison with counting methodology and impact on study results on TES in Africa	E. Halsey, M. Plucinski, V. Udhayakumar
16.30–17.20	Reinfection/recrudescence: pros and cons of available and potential new methods and feasibility within national programmes	D. Neafsey
17.20–18.10	Role of modelling to validate algorithms to distinguish reinfection from recrudescence	I. Hasting
TUESDAY, 18 MAY 2021		
14.00–14.50	Experience of using *msp1, msp2* and *glurp* in different transmissions settings: comparison between sequential and 2/3 algorithm and impact on TES results	P. Ringwald
14.50–15.40	Comparison of *msp1, msp2* and *glurp* vs *msp1, msp2* and microsatellites vs *msp1, msp2* and amplicon sequencing in African samples collected in TES	D. Ménard
15.40–16.30	Comparison of whole sequencing vs *msp1, msp2* and *glurp* vs microsatellites using African samples	J. Juliano
16.45–17.45	Overall discussion	Expert panel, Secretariat, and Rapporteur
17.45–18.45	Formulation of recommendations	Expert panel, Secretariat, and Rapporteur
18.45	Closing remarks	P. Alonso, D. Wirth

ANNEX 2. LIST OF PARTICIPANTS

EXPERT PANEL

Dyann Wirth
Harvard T.H. Chan School of Public Health
Boston, United States of America (Chair)

Caroline Buckee
Harvard T.H. Chan School of Public Health
Boston, USA

Umberto D'Alessandro
Medical Research Unit
Banjul, Gambia

Abdoulaye Djimde
University Science, Techniques and Technologies
Bamako, Mali

Lucy Okell
Imperial College, London
United Kingdom

Gauthier Mesia Kahunu
University of Kinshasa
Kinshasa, Democratic Republic of the Congo

Daouda Ndiaye
University Cheikh Anta Diop
Dakar, Senegal (unable to attend)

Harald Noedl
Malaria Research Initiative Bandarban
Vienna, Austria

Philip Rosenthal
University of California
San Francisco, USA

Sarah Volkman
Harvard T.H. Chan School of Public Health
Boston, USA

Marian Warsame
University of Gothenburg
Gothenburg, Sweden

PRESENTERS

Ingrid Felger
Swiss Tropical and Public Health Institute
Basel, Switzerland

Monica Golumbeanu
Swiss Tropical and Public Health Institute
Basel, Switzerland

Eric Halsey
Centers for Disease Control and Prevention and President's Malaria Initiative
Atlanta, USA

Ian Hastings
Liverpool School of Tropical Medicine
Liverpool, United Kingdom of Great Britain and Northern Ireland

Jonathan Juliano
University of North Carolina
Chapel Hill, USA

Didier Ménard
Institut Pasteur
Paris, France

Daniel Neafsey
Broad Institute of MIT and Harvard
Cambridge, USA

Christian Nsanzabana
Swiss Tropical and Public Health Institute
Basel, Switzerland

Mateusz Plucinski
Centers for Disease Control and Prevention and President's Malaria Initiative
Atlanta, USA

Venkatachalam Udhayakumar
Centers for Disease Control and Prevention and President's Malaria Initiative
Atlanta, USA

OBSERVERS

Stephan Duparc
Medicines for Malaria Venture
Geneva, Switzerland

Scott Filler
The Global Fund to Fight AIDS, Tuberculosis and Malaria
Geneva, Switzerland

Jennifer Gardy
Bill & Melinda Gates Foundation
Seattle, USA

Scott Miller
Bill & Melinda Gates Foundation
Seattle, USA

Jörg Möhrle
Medicines for Malaria Venture
Geneva, Switzerland

Jonathan Russell
Bill & Melinda Gates Foundation
Seattle, USA

Rick Steketee
United States Agency for International Development
Washington D.C., USA

Meera Venkatesan
United States Agency for International Development
Washington D.C., USA

WHO SECRETARIAT

Pedro Alonso
Director, Global Malaria Programme

Dorina Maria Bustos
Medical Officer, WHO Country Office for Thailand, Bangkok

Abdisalan Noor
Unit Team Lead, Strategic Information for Response Unit, Global Malaria Programme

Charlotte Rasmussen
Technical Officer, Diagnostics, Medicines and Resistance Unit, Global Malaria Programme

Pascal Ringwald
Coordinator, Director's Office, Global Malaria Programme

RAPPORTEUR

Naomi Richardson
Magenta Communications
Abingdon, United Kingdom

ANNEX 3. SUPPORTING DOCUMENTS

Early AM, Daniels RF, Farrell TM, Grimsby J, Volkman SK, Wirth DF, et al. Detection of low-density *Plasmodium falciparum* infections using amplicon deep sequencing. Malar J. 2019;18(1):219. https://www.ncbi.nlm.nih.gov/pubmed/31262308

Falk N, Maire N, Sama W, Owusu-Agyei S, Smith T, Beck HP, et al. Comparison of PCR-RFLP and Genescan-based genotyping for analyzing infection dynamics of *Plasmodium falciparum*. Am J Trop Med Hyg. 2006;74(6):944-50. https://www.ncbi.nlm.nih.gov/pubmed/16760501

Felger I, Snounou G, Hastings I, Moehrle JJ, Beck HP. PCR correction strategies for malaria drug trials: updates and clarifications. Lancet Infect Dis. 2020;20(1):e20-e5. https://www.ncbi.nlm.nih.gov/pubmed/31540841

Halsey E, Plucinski M, Udhayakumar V. Microsatellites and a Bayesian algorithm for efficacy calculation: the PMI/CDC approach. Manuscript in preparation.

Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. Am J Trop Med Hyg. 2006;75(5):836-42. https://www.ncbi.nlm.nih.gov/pubmed/17123974

Greenhouse B, Dokomajilar C, Hubbard A, Rosenthal PJ, Dorsey G. Impact of transmission intensity on the accuracy of genotyping to distinguish recrudescence from new infection in antimalarial clinical trials. Antimicrob Agents Chemother. 2007;51(9):3096-103. https://www.ncbi.nlm.nih.gov/pubmed/17591848

Gruenberg M, Lerch A, Beck HP, Felger I. Amplicon deep sequencing improves *Plasmodium falciparum* genotyping in clinical trials of antimalarial drugs. Sci Rep. 2019;9(1):17790. https://www.ncbi.nlm.nih.gov/pubmed/31780741

Jones S, Kay K, Hodel EM, Chy S, Mbituyumuremyi A, Uwimana A, et al. Improving methods for analyzing antimalarial drug efficacy trials: molecular correction based on length-polymorphic markers *msp-1*, *msp-2*, and *glurp*. Antimicrob Agents Chemother. 2019;63(9). https://www.ncbi.nlm.nih.gov/pubmed/31307982

Jones S, Plucinski M, Kay K, Hodel EM, Hastings IM. A computer modelling approach to evaluate the accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of antimalarial drug efficacy. Antimicrob Agents Chemother. 2020;64(4). https://www.ncbi.nlm.nih.gov/pubmed/31932376

Jones S, Kay K, Hodel EM, Gruenberg M, Lerch A, Felger I, et al. Should deep-sequenced amplicons become the new gold-standard for analysing malaria drug clinical trials? bioRxiv. 2021:03.23.436602. https://www.biorxiv.org/content/10.1101/2021.03.23.436602v1.full.pdf+html

Juliano J, Bailey J, Mesia Kahunu G, Plucinski M, Halsey E, Giesbrecht D, et al. Genotyping recurrent parasitemia in malaria clinical trials through size polymorphisms, microsatellites and molecular inversion probes: a case study in Mikalayi, Democratic Republic of the Congo. Manuscript in preparation.

Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 2017;18(1):864. https://www.ncbi.nlm.nih.gov/pubmed/29132317

Messerli C, Hofmann NE, Beck HP, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. Antimicrob Agents Chemother. 2017;61(1). https://www.ncbi.nlm.nih.gov/pubmed/27821442

Mwangi JM, Omar SA, Ranford-Cartwright LC. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing *Plasmodium falciparum* infections from reinfections in Kenya. Int J Parasitol. 2006;36(3):329-36. https://www.ncbi.nlm.nih.gov/pubmed/16442537

Nyachieo A, Van Overmeir C, Laurent T, Dujardin JC, D'Alessandro U. *Plasmodium falciparum* genotyping by microsatellites as a method to distinguish between recrudescent and new infections. Am J Trop Med Hyg. 2005;73(1):210-3. https://www.ncbi.nlm.nih.gov/pubmed/16014861

Plucinski MM, Morton L, Bushman M, Dimbu PR, Udhayakumar V. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrob Agents Chemother. 2015;59(10):6096-100. https://www.ncbi.nlm.nih.gov/pubmed/26195521

Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, Felger I. Comparison of *Plasmodium falciparum* allelic frequency distribution in different endemic settings by high-resolution genotyping. Malar J. 2009;8:250. https://www.ncbi.nlm.nih.gov/pubmed/19878560

Methods and techniques for clinical trials on antimalarial drug efficacy: genotyping to identify parasite populations. Geneva: World Health Organization; 2008 (https://apps.who.int/iris/bitstream/handle/10665/43824/9789241596305_eng.pdf).

ANNEX 4. PRESENTATIONS

Day 1

New evidence on *msp1, msp2* and *glurp* as markers for reinfection and recrudescence (I. Felger, C. Nsanzabana)

Limitations of current markers and genotyping methods

The currently recommended genotyping protocol is for multiplex nested PCR for *msp1, msp2* and *glurp* with CE high-resolution fragment sizing in an automated sequencer. However, some issues have arisen.

- Confusion around CE: In some cases, this has been interpreted as fragment sizing by Bioanalyzer, which is not sufficiently precise *(1)*. Fragment sizing should be performed in an automated sequencer.
- Definition of genotype bins: CE by automated sequencer discriminates a 1–2 base pair (bp) difference. Genotype bins of 3 bp are adequate for protein coding markers, and their boundaries are carefully selected based on discrete size steps when plotting a large number of fragments by size.
- Competition between allelic families: To avoid amplification bias, accuracy of genotyping outcomes should be improved through separate amplification reactions per allelic family *(2,3)*.
- Within-family competition also exists: PCR is biased towards short fragments. (Such bias in PCR was also reported for microsatellites during the meeting) *(2)*.
- Limitations of *glurp* as a marker: Long allele sizes increase competition; within one allelic family there is direct competition between all alleles; it is prone to stutter peaks, requiring increased cut-off limits and increasing the risk of excluding clones *(2)*.
- Markers *msp1* and *msp2* should be maintained, as both are highly diverse and amplification bias can be reduced (but not fully avoided) by using allelic-family-specific, single-tube PCR.

Adherence to the agreed WHO definitions of new infection and recrudescence is still insufficient, as evidenced by published reports of PCR correction. Automated sequencing should be used for maximum resolution of fragments. All PCRs should be performed separately to reduce template competition. Reporting of findings should include the cut-off for peak height, the strategy used for cut-off validation and the minority clone detection limit. In the wider context, external quality control activities should be encouraged for all laboratories and communication among genotyping laboratories should be enhanced as a platform for improvement.

Other markers and genotyping methods

The development and evaluation of five AmpSeq markers in samples from clinical trials was presented. The three best-performing markers were *cpmp*, *cpp* and *ama1-D3 (4,5)*. Despite high reproducibility in triplicates and robust detection of minority clones as low as 1%, this technique is operationally challenging (laborious pipetting, contamination prevention) and requires substantial technical expertise, particularly in bioinformatics.

The analytical sensitivity of five different genotyping techniques for minority clone detection was compared under standardized conditions at the Swiss Tropical and Public Health Institute using well-characterized laboratory strains (3D7, K1, HB3, FCB1):

- *msp1/2/glurp* by CE *(2)*;
- *msp1/2/glurp* by Bioanalyzer *(6)*;
- *TA60, TA40, TA81, PfPK2* by CE *(7)*;
- *ama-1-D3, csp, cpmp, cpp, msp7* by AmpSeq *(4,5)*;
- *msp1/msp2* by high-resolution melt analysis *(8)*.

For each technique, minority clone detectability, robustness, time and cost were evaluated. CE and AmpSeq showed the highest sensitivity for detecting minority clones and are considered robust techniques. Microsatellites improved sensitivity in minority strain detection versus *glurp*. Note that these results may not be wholly representative of biological samples, and with all techniques the biological constraints remain.

Microsatellites and Bayesian algorithm in Africa: added value of microsatellites, validation of Bayesian algorithm, comparison with counting methodology and impact on study results on TES in Africa (E. Halsey, M. Plucinski, V. Udhayakumar)

Microsatellites

The US-CDC uses seven neutral microsatellites (*TA1, Poly-α, Pfpk2, TA109, TA2490, 313_C2*, and *383_C3*) *(9–11)*. Ideally, reinfections would be identified by finding all different markers between day 0 and day X, and recrudescences by finding identical markers; however, there is usually a mixture. In fact, repeat genotyping of the exact same sample shows that variations can sometimes occur in microsatellite markers, even when none would be expected. These differences could be the result of measurement error, unobserved minority strains (often due to limited DNA concentration), and varying allele frequencies in a population.

Bayesian analysis

A good analytic approach should account for the variability outlined above, be reproducible, enable classifications to be generated based on published genotype data, and provide precision, with a measure of uncertainty around the classifications. To address these needs, the US-CDC has developed a Bayesian statistical approach that outputs a posterior probability of recrudescence based on an estimate of the degree to which the observed evidence supports reinfection versus recrudescence from microsatellite genotyping data. This approach incorporates parameters such as population frequencies, the probability of allelic suppression, and the error rate of fragment length measurement *(12)*.

The algorithm is implemented using a Monte Carlo Markov chain method, an open-source online tool investigators can use to calculate posterior probabilities without the use of additional statistical software. The algorithm has been used to analyse microsatellite data genotyping for US-CDC/PMI studies since 2015. Across 26 sites in nine African countries (Guinea, Benin, Democratic Republic of the Congo, Angola, Ethiopia, Uganda, Rwanda, Madagascar, Mozambique), the most recent PCR-corrected efficacy rates from each country were all over 90% except for:

- Busia, Uganda (AL 87%);
- Mikalayi, Democratic Republic of the Congo (AL 86%, DP 84%);
- Lunda Sul, Angola (AL 88%).

Conclusions

The advantage of using a Bayesian approach is the ability to jointly estimate the posterior distribution of the probability of recrudescence, as well as the unknown functions and parameters and hidden (unobserved) alleles. Therefore, it is possible to report the uncertainty around antimalarial drug efficacy estimates.

Consistency in approach and transparency in reporting methods are necessary, with the reporting of full genotype data, in order to enable alternative analysis. New genotyping methods require new statistical approaches and algorithms. Because of the differences in methods over time, trends in efficacy should be interpreted with caution.

Reinfection/recrudescence: pros and cons of available and potential new methods and feasibility within national programmes (D. Neafsey)

Comparison of genotyping methods

There are four approaches available: *msp1/msp2/glurp* genotyping, microsatellites, AmpSeq and WGS. The advantages and disadvantages of each were compared based on the following criteria and are summarized below.

Sensitivity

- Absolute sensitivity: Will the assay detect ANY parasites in the sample?
- Minor strain sensitivity: Will the assay detect ALL parasites in the sample? *(8)*.

Reproducibility *(7,8,13)*

- Is the assay reliable?
- Is there opportunity for subjective interpretation?
- How comparable are results between different studies?

Information content

- Are the assay targets sufficiently diverse in the local population? Marker number and heterozygosity reduce uncertainty, particularly in polyclonal infections *(14)*. AmpSeq targets can be highly diverse *(4,15)*.
- Marker number and heterozygosity are fungible, but higher marker number decreases feasibility.

Feasibility

- Equipment, expertise, cost and analysis requirements in the setting in which it is to be applied need to be considered.
- Targeted next generation sequencing is uncommon in Africa, but centres of excellence are already being established to provide high-quality sequencing services for several laboratories and research groups *(16)*. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has also highlighted the need for greater capacity for genomic analysis.

Summary

METHOD	SENSITIVITY (absolute)	SENSITIVITY (minor strains)	REPRODUCIBILITY	INFORMATION CONTENT	FEASIBILITY
msp1/msp2/glurp	High	5:1	Poor (gels) Medium (CE)	High	Very high
Microsatellites	High	5:1	Poor (gels) Medium (CE)	High	Very high
AmpSeq	High	100:1	High	High	High
WGS	Medium	~20:1	High	Very high	Medium

Analysis considerations

- Probabilistic infection classifications can capture uncertainty with any method, enabling comparison across studies/methods if genotyping error profiles are understood.
- All probabilistic infection classification approaches require knowledge of the diversity of the markers in the local parasite population. With AmpSeq, this can be done from locally generated data, and whole genome data are available for parasites in many sites.
- Error profiles are differentially quantifiable by genotyping method: AmpSeq error profiles are more measurable and so more easily incorporated into the analysis.
- Marker number and marker heterozygosity reduces uncertainty, particularly in polyclonal infections.
- Cloud-based common analysis pipelines can assist in transparency and portability.

Conclusions

Overall, AmpSeq is the best genotyping method, although issues around affordability and skills acquisition need to be addressed for its use in Africa. However, the advantages of this method should incentivize the community to address the feasibility challenges.

Role of modelling to validate algorithms to distinguish reinfection from recrudescence (I. Hasting)

Mechanistic pharmacokinetic/pharmacodynamic modelling

A model was developed to estimate the antimalarial cure rate on a per-infection (clone) basis *(17)*. This is particularly relevant in high transmission areas where the MOI is >1 and individuals may experience multiple reinfections that may or may not be cleared *(18)*. Genotypes were added to each clone randomly and 5000 patient simulations were conducted. Note that within simulations the 'true' failure rate is known so the performance of the different genotyping methods can be evaluated.

msp1, msp2* and *glurp

Four analysis methods using *msp-1, msp-2* and *glurp* were evaluated *(17)*:

- WHO/MMV: Initial and recurrent samples must share alleles at all three markers for recrudescence.

- No *glurp*: Same as for WHO/MMV but based on *msp1* and *msp2* only.
- 2/3 markers: Initial and recurrent samples must share alleles for at least 2/3 markers for recrudescence.
- Allelic family switch: Identical alleles in *msp1* and *msp2* indicate recrudescence, and absence of shared alleles in both markers indicate reinfection. If markers are discordant, a complete allelic family shift in the no-sharing marker is required to classify reinfection.

The WHO/MMV method tended to underestimate the true failure rate by a factor of two. The 2/3 method most closely matched the true failure rate because the misclassification of recrudescence as reinfection and vice versa was more closely balanced. None of the algorithms correctly classified all recurrent infections.

Microsatellites

Seven microsatellite markers and their allele distributions from three sentinel sites in Angola were examined and classified using a simple match-counting algorithm or Bayesian analysis *(19)*.

- The match-counting method was unstable and unreliable.
- The Bayesian algorithm was unable to accurately identify low-density recrudescence, but this did not appear to compromise its utility as a highly effective molecular correction method for analysing microsatellite genotypes.

The model findings were largely consistent with field data *(12)*. The US-CDC Bayesian algorithm does not model for the possibility of spurious allele readings in the data (false positive); this effect would increase with the MOI and force Bayesian statistics to infer more 'missing' alleles. Therefore, the distribution of the posterior probabilities should be checked.

Deep sequenced amplicons

Five markers were evaluated (*cpmp, ama1-D3, cpp, csp, msp-7*), with allelic diversity obtained from published sources *(20)*. Recrudescence was determined by counting the number of markers with a shared allele between the initial and recurrent samples.

- By using *cpmp, ama1-D3* and *cpp* and a matching threshold of ≥2 or =3 to classify recrudescence, the method returned highly robust estimates under a range of MOI distributions and force of infection values.
- This method has not been compared to field data.

Conclusions

- The advantage of using AmpSeq over other markers is the improved ability to detect genetic signals from low-density clones.
- A simple matching algorithm for three amplicons provided robust estimates of recrudescence and reinfection across a range of simulated transmission settings.
- False positives have not been addressed in the Bayesian algorithm.
- Bayesian analysis has not been tested in this model for AmpSeq.

Day 1: Discussion

Sequential WHO/MMV versus 2/3 decision algorithm

The WHO/MMV sequential analysis method allows early stopping. The 2/3 method might have cost implications in some settings. However, with this algorithm, *msp1* and *msp2* can be done together and then *glurp* is only done if the findings from *msp1* and *msp2* are discordant. The primary limitation of *glurp* is its size, although it does have extensive genetic diversity. Therefore, it has some discriminatory value, particularly in low transmission zones where there is no MOI. The advantage of 2/3 is that if the initial and recurrent *msp1* and *msp2* share the same genotype, then the possibility of it being a recrudescence is high and *glurp* can be discounted.

According to modelling, the 2/3 method superficially appears to provide better classification than the WHO/MMV method. This is only because the misclassification errors are more balanced between reinfection and recrudescence, whereas MMV/WHO is more likely to classify recrudescence as reinfection. The major limitation of the 2/3 algorithm is that even if one of the markers clearly indicates a reinfection, it is ignored. This method also appears to overestimate recrudescence for AL in high transmission areas, so would not be appropriate for use across all transmission settings and for all antimalarial drugs. When applied to *msp1/msp2/glurp*, the 2/3 algorithm gives equal weight to *glurp* when there is discordance between *msp1/msp2*, even though *glurp* is a suboptimal marker.

AmpSeq and the detection of minority clones

Although the different genotyping techniques use different methods to identify recrudescent parasites versus reinfection, they can be compared on their ability to detect minority clones.

The ability to more effectively detect minority clones might affect the threshold for determining whether a drug is failing. In addition, low-density recrudescences that would previously have been missed might be identified, although these may not be clinically relevant. AmpSeq is not significantly more sensitive than CE; if only samples from patients with microscopically determined infection are analysed, then these represent clinically significant parasitaemia levels. However, there may be some very low-density minority infections identified that may change the outcome of the genotyping.

The parasite most likely detected on day 0 is the dominant clone responsible for the clinical symptoms of malaria; although minority clones may be present, these parasites were not the cause of the clinical disease. So, if the aim is to determine if the drug cleared the parasite responsible for clinical disease, then only the majority clone at day 0 will be relevant. However, the detection of minority clones is important because a minority clone at day 0 could be the dominant clone at day 28. Additionally, any clone that survives to cause a recurrence is the one most likely to be drug-resistant. Therefore, identifying and clearing the minor clones is important for preventing the spread of resistance within the population.

With more markers, there is the opportunity to resolve parasites that are highly similar to each other but not identical. However, this might be an issue in lower transmission settings where parasite diversity is likely to be more limited. In sub-Saharan Africa, in many settings with moderate to high levels of transmission, most parasites are sufficiently dissimilar so that a modest panel of markers could suffice. Ongoing comparisons with WGS data could help to establish the optimal number of markers.

Implications of sensitivity

The sensitivity required depends on the question being asked, and sensitivity might be limited by the facilities available in a country versus in a central laboratory. However, if a method is too sensitive, there is a possibility that gametocytes might be detected. This would undermine the reliable estimation of drug efficacy, as most antimalarial treatments do not clear gametocytes.

Microsatellite genotyping analysed using a Bayesian algorithm

The variation in microsatellites across the population requires consideration. The Bayesian algorithm is tuned to population-specific allele frequencies of the microsatellite markers, and a match with two frequent alleles is not weighted the same as a match with two rare alleles. The algorithm can also incorporate unpaired day 0 samples to enrich the information on allelic frequencies.

The microsatellite markers used by the US-CDC were chosen after pilot testing of various markers. The combination of the seven markers has a high discriminatory power to identify different clones. The US-CDC has collected a large body of data on the diversity of the seven microsatellites that they use and would be happy to share those data with the group.

The US-CDC switched from *msp1/msp2* to microsatellites because, in highly endemic settings, although they were getting fragment lengths that were identical, when they sequenced the samples, they found that they were different parasites. With seven markers, if there is a high MOI, it may be difficult to obtain a complete set of information for all clones. However, seven markers were chosen to provide the diversity needed to discriminate parasite types across all study sites, so that one set of markers could be used for all study sites.

Characteristics of the Bayesian algorithm

An advantage of the Bayesian approach is that informative priors can be assigned. The current algorithm is run with uninformed priors, so every treatment failure is treated as equally likely to be a recrudescence or reinfection. However, it is possible to modify the model for the likely outcomes for a particular site. All of the data are run on a site-specific basis at present because of the variability in allele frequencies between sites. The Bayesian approach has been applied to 30 studies across 11 sites, peer-reviewed, made open-source and investigated in the pharmacokinetic/pharmacodynamic model.

The finding that the same samples give different genotyping results indicates either that things are being missed or that these are false positives. The influence of false positives on the Bayesian approach was raised. There are two things that increase the probability of a false positive or a chance match: 1) when there are very frequent alleles; and 2) high MOI because of the high numbers of comparisons. The latter in particular is a characteristic of high transmission settings.

The possibility of automation was discussed. Georgia Tech has already coded the algorithm in Python and it is publicly available and open-source on a web interface. The aim going forward is to integrate this into an analytic pipeline where the data are generated, automatically analysed and outputted in a standardized and understandable format. Explaining the rationale and methods for genotyping to malaria programmes is complex. However, the US-CDC has provided seminars for countries on microsatellite genotyping and Bayesian analysis where further information has been requested.

Pharmacokinetic/pharmacodynamic modelling

The Bayesian algorithm has not been applied to *msp1*, *msp2* and *glurp* within this model.

The model assumed failure rates of 10%. At this level, the misidentified recrudescences and misidentified reinfections were balanced with the 2/3 method. At higher failure rates, although this balance might be disturbed, the drug would still be identified as a failing drug.

The model assumes that everyone attends their assessments on time. This is because the timing of the assessment influences the number of clones detected. If there is a long gap between follow-up visits, it provides time for the minority clones to emerge and replicate, so the MOI will increase.

In low transmission areas, reinfection by related parasites is more likely, so three amplicons might not be enough to discriminate recrudescence from reinfection. The model is focused on moderate to high transmission, so this aspect has not been considered. However, co-transmission has been observed in high transmission areas, so it might be useful to examine how the degree of relatedness might affect the model outcomes.

Both *msp1* and *msp2* are subject to immune pressure, whereas microsatellites are neutral markers. The effect of immunity on the model has not been investigated relative to the decision algorithm. However, US-CDC has done an in-house adaptation of the Bayesian algorithm for *msp1/msp2/glurp* and the Bayesian estimates very closely match the 2/3 approach.

Immunity is not built into the model because if patients have clinical malaria, it is assumed that they are immunologically susceptible to that parasite. However, this does not consider the possibility that other parasites are present but are immunologically suppressed.

In the model, the Bayesian algorithm can produce profiles with mostly clear recrudescences or reinfections, which is generally consistent with field data. However, some field data show a significant number of intermediate values for the posterior probabilities. At present, the reasons for this pattern have not been explored in the model. However, intermediate results from the Bayesian algorithm are a flag that the genotyping data are not fully discriminatory. This is useful information and can enable identification of samples for further downstream analysis, for example, *msp1/msp2* for samples with intermediate microsatellite results. An advantage of the probabilistic approach is that it incorporates uncertainty into the final efficacy calculations; for example, in the efficacy calculations, a late treatment failure with a posterior probability of recrudescence of 0.35 would count as 35% of a recrudescence.

The key factor determining the outcomes of the model is the sensitivity of the different methods to detect minority clones.

Communicating the outputs of different analysis methods

Regarding interpretation, there was a concern that the Bayesian approach to analysing microsatellites generally reports decreased efficacy for antimalarial drugs compared to the *msp1/msp2/glurp* WHO/MMV approach. Should this decrease be regarded as a concerning sudden drop in drug efficacy, or should it prompt a

change in the threshold for drug efficacy (currently 90%)? Although the Bayesian approach does not always decrease efficacy, this issue highlights that any change in methods that gives rise to a change in efficacy presents a communication challenge. The methods and discussion need to make clear that the methods and analysis are not comparable with drug efficacy data obtained in previous studies. However, communicating these issues in a reasonable and credible manner is something that requires careful thought.

Operational considerations

Although AmpSeq appears to be the best method for genotyping, there were questions around whether it would be feasible to deploy AmpSeq to study sites, or whether it would be best at a limited number of central laboratories. There are several examples of Illumina sequencing being performed at central sites in Africa.

Most public health groups would be comfortable sending samples to a central centre for analysis. However, some countries may not want samples sent outside the country for processing, but may not have the capacity for AmpSeq.

There are also supply chain issues (e.g., cold-chain provision of sequencing reagents) that might limit the development of these capabilities in every country with a TES. One issue is that AmpSeq is not cost-effective for small numbers of samples, for example, for repeating an assay; batches of several hundred samples are required. Therefore, for these reasons, a 'centres of excellence' approach or regional hubs for AmpSeq is a more viable strategy in Africa.

A further advantage of AmpSeq is that the same platform can be used to detect mutations associated with drug resistance. This offers the potential to integrate drug efficacy and drug resistance data. SARS-CoV-2 has highlighted the need for capacity in genome sequencing, which can be leveraged to establish AmpSeq as the preferred approach to evaluating drug efficacy.

The Bill & Melinda Gates Foundation has been supporting the Africa CDC to look at the role it could play as a continent-wide organizer for pathogen genomics. Before the SARS-CoV-2 pandemic, they were planning on setting up a series of six laboratory sites across Africa. These are not public health laboratories, but are more closely identified as research institutes. The laboratories are within institutions where genomics research has been previously conducted. In phase I, the plan was for these laboratories to work on a range of pathogens, but their efforts are currently directed mainly towards SARS-CoV-2 genomics. However, the pandemic has accelerated capacity in these laboratories, with tens of thousands of SARS-CoV-2 genomes generated largely on the Illumina platform. It has also driven the establishment of shared procurement networks. Africa CDC is looking to return to conducting routine surveillance for a range of pathogens, including malaria. In particular, histidine rich protein 2 gene deletion and drug resistance surveillance are priorities. It will likely be a least a year before the laboratories can transition from their focus on SARS-CoV-2. However, these laboratories represent a potentially impactful network of sites. There are advantages to being under the umbrella of Africa CDC, such as the authority of the organization, which promotes Member States' willingness to share data and cooperate, and shared procurement mechanisms; procurement is a major challenge for the establishment of any laboratory facility in sub-Saharan Africa. This network of laboratories supported by Africa CDC is a potentially valuable asset for deploying AmpSeq for antimalarial drug efficacy studies across Africa.

Day 2

Experience of using *msp1, msp2* and *glurp* in different transmissions settings: comparison between sequential and 2/3 algorithm and impact on TES results (P. Ringwald)

Methods

For 145 TESs conducted between 2010 and 2019, the results of PCR correction (*msp1, msp2* and *glurp*) using the WHO/MMV algorithm were compared to the 2/3 algorithm for 1086 recurrences of *P. falciparum* malaria (2332 samples). Overall PCR success for each marker was high (*msp1* 96.0%, *msp2* 94.0%, *glurp* 93.6%).

Low to moderate transmission areas outside Africa (Greater Mekong, Middle East and South America)

A total of 60 TESs were conducted from 2011 to 2019 in Cambodia, Lao People's Democratic Republic (PDR), Myanmar, Viet Nam, Pakistan, Yemen and Guyana monitoring artesunate-amodiaquine (ASAQ), artesunate-mefloquine, artesunate-pyronaridine, artesunate+sulfadoxine-pyrimethamine (ASSP), artemether-lumefantrine (AL), dihydroartemisinin-piperaquine (DP), and artesunate monotherapy (297 patients, 587 samples). Failures classified as recrudescences were 86.5% (244/282) with WHO/MMV versus 94.3% (266/282) with 2/3 (p=0.002); the discrepancy was caused by *glurp* in 72.7% of cases.

- Kaplan–Meier efficacy estimates were ≤2% lower for 2/3 than for WHO/MMV, except:
    - DP in Dak Lak (Viet Nam) in 2019: 35.9% with MMV/WHO versus 28.4% with 2/3.

For 93 samples from Cambodia (2014–2016), *P. falciparum* 10-single nucleotide polymorphism (SNP) barcodes were performed. For two discordant samples (MMV/WHO reinfection; 2/3 recrudescence), one was a reinfection, and one was a recrudescence.

Low to moderate areas in the Horn of Africa

A total of 33 TESs were conducted from 2010 to 2019 in Eritrea, Somalia and Sudan monitoring ASAQ, ASSP, AL and DP (175 patients, 350 samples). Failures classified as recrudescences were 54.4% (92/169) with WHO/MMV versus 71.6% (121/169) with 2/3 (p<0.001); the discrepancy was caused by *glurp* in 58.6% of cases.

- Kaplan–Meier efficacy estimates were ≤3.4% lower for 2/3 than for WHO/MMV, except:
    - ASSP in Sinnar (Sudan) in 2011: 98.9% with MMV/WHO versus 94.4% with 2/3.

Low to moderate transmission areas in West Africa

A total of 10 TESs were conducted from 2010 to 2018 in Gambia and Mauritania monitoring ASAQ, AL and DP (27 patients, 53 samples). Failures classified as recrudescences were 57.7% (15/26) with WHO/MMV versus 69.2% (18/26) with 2/3 (p=NS); the discrepancy was caused by *glurp* in 100% of cases.

- Kaplan–Meier efficacy estimates were ≤3.0% lower for 2/3 than for WHO/MMV.

Moderate to high transmission in West Africa

A total of 14 TESs were conducted from 2011 to 2018 in Liberia, Sierra Leone and Togo monitoring ASAQ, AL and DP (91 patients, 174 samples). Failures classified as recrudescences were 16.7% (14/84) with WHO/MMV versus 50.0% (42/84) with 2/3 ($p<0.001$); the discrepancy was caused by *glurp* in 89.3% of cases.

- Kaplan–Meier efficacy estimates were ≤2.4% lower for 2/3 than for WHO/MMV, except:
    - ASAQ in Montserrado (Liberia): 93.7% with MMV/WHO versus 84.9% with 2/3;
    - AL in Bo (Sierra Leone): 100% with MMV/WHO versus 94.4% with 2/3;
    - AL in Eastern (Sierra Leone): 100% with MMV/WHO versus 94.6% with 2/3;
    - AL in Kara (Togo): 97.3% with MMV/WHO versus 93.2% with 2/3.

Moderate to high transmission in Central Africa

A total of 13 TESs were conducted from 2010 to 2018 in Congo, the Democratic Republic of the Congo and Equatorial Guinea monitoring ASAQ and AL (84 patients, 168 samples). Failures classified as recrudescences were 29.8% (25/84) with WHO/MMV versus 53.6% (45/84) with 2/3 ($p<0.001$); the discrepancy was caused by *glurp* in 75.0% of cases.

- Kaplan–Meier efficacy estimates were ≤2.3% lower for 2/3 than for WHO/MMV, except:
    - AL in Haut Katanga (Democratic Republic of the Congo): 97.6% with WHO/MMV versus 92.2% with 2/3;
    - AL in Ebibeyin (Equatorial Guinea): 93.7% with WHO/MMV versus 84.5% with 2/3;
    - ASAQ in Ebibeyin (Equatorial Guinea): 98.7% with WHO/MMV versus 93.6% with 2/3.

Low to moderate South-eastern Africa

A total of nine TESs were conducted from 2012 to 2015 in Burundi and Rwanda monitoring ASAQ, AL and DP (218 patients, 420 samples). Failures classified as recrudescences were 14.4% (27/187) with WHO/MMV versus 28.3% (53/187) with 2/3 ($p<0.001$); the discrepancy was caused by *glurp* in 53.8% of cases.

- Kaplan–Meier efficacy estimates were ≤2.7% lower for 2/3 than for WHO/MMV, except:
    - AL in Eastern 2012 (Rwanda): 97.1% with WHO/MMV versus 92.2% with 2/3;
    - AL in Kigali (Rwanda): 95.7% with WHO/MMV versus 90.9% with 2/3;
    - AL in Eastern 2013 (Rwanda): 99.3% with WHO/MMV versus 95.4% with 2/3;

Summary

Overall, the proportion of recurrent parasitaemia classified as recrudescence was higher with the 2/3 algorithm (68.8%) than with the WHO/MMV method (46.0%) ($p<0.001$). However, this did not always translate into a significant difference in Kaplan–Meier estimates of treatment outcome (i.e., >4% difference, 12/145 studies).

- Differences in the Kaplan–Meier estimates of treatment outcome were more evident in areas of moderate to high transmission than in areas of low to moderate transmission.

- AL was the most affected drug (66.7%) > ASAQ (16.7%) > ASSP and DP (8.3%) based on a failure rate >4%.

- The main marker leading to discordance between the two analyses was *glurp* (68.2%) > *msp2* (16.2%) and *msp1* (15.6%).

Conclusions

- As a marker to discriminate recrudescence from reinfection, *glurp* is inadequate and has a disproportionate influence on the outcome of the analysis algorithms. These data support abandoning *glurp* as a marker.

- The disproportionate effect on AL efficacy from using the 2/3 method versus the WHO/MMV method is concerning, given that lumefantrine has the shortest half-life of any partner drug. In the absence of a "gold standard", this analysis does not provide evidence that 2/3 provided any added value compared to the WHO/MMV method in high transmission regions in particular for AL. The other possible interpretation is that lumefantrine resistance has emerged, but it seems unlikely for it to emerge only in high transmission areas.

Comparison of *msp1, msp2* and *glurp* vs *msp1, msp2* and microsatellites vs *msp1, msp2* and amplicon sequencing in African samples collected in TES (D. Ménard)

Methods

Markers were evaluated from paired samples obtained from TESs conducted between 2016 and 2019 in countries with low and high endemicity.

- *msp1, msp2* and *glurp* with bands detected using gel electrophoresis (60%) or CE (40%): For gel electrophoresis, bands were considered different between initial and recrudescent samples if the size of the bands differed >20 bp for *msp1/msp2* and >50 bp for *glurp*, and for CE, if the size of the bands differed >10 bp for msp1/msp2 and > 20 bp for glurp.

- Microsatellites (*Poly-α, Pfpk2, TA1*) with bands detected using CE *(21)*: Bands were considered different between initial and recrudescent samples if the size of the bands differed >5 bp.

- *ama1, cpmp, cpp* polymorphic genes detected by AmpSeq *(5)*.

Different combinations of markers were analysed using either the WHO/MMV or the 2/3 algorithm both overall and for low transmission areas (Eritrea [n=23] and Gambia [n=2]) and high transmission areas (Burundi [n=30], Congo [n=3], Equatorial Guinea [n=14], Liberia [n=20]).

Markers

- The degree of concordance within markers was 38% for *msp1/msp2/glurp*, 58% for microsatellites and 63% for AmpSeq.

- The degree of concordance between markers was highest with AmpSeq, with 29% of results classified as 'intermediate' compared to 42% for microsatellites and 62% for *msp1/msp2/glurp*.

Algorithms

There was considerable variation in the percentage recrudescence rates determined using the different combinations of markers and algorithms (MMV/WHO or 2/3):

- 26% *msp1/msp2/*microsatellites (MMV/WHO); 29% *msp1/msp2/glurp* (MMV/WHO); 33% *msp1/msp2*/AmpSeq (MMV/WHO); 37% microsatellites (MMV/WHO); 50% *msp1/msp2* (MMV/WHO); 65% AmpSeq (MMV/WHO); 70% *msp1/msp2/glurp* (2/3).

- In low transmission areas, there was no significant difference in recrudescence rates between the different marker/algorithm combinations, with outcomes ranging from 52% for *msp1/msp2/*microsatellites (MMV/WHO) and microsatellites (MMV/WHO) to 80% for *msp1/msp2/glurp* (2/3) and *msp1/msp2* (MMV/WHO).

- By contrast, in high transmission areas, the range of outcomes was broader, ranging from 13% for *msp1/msp2/glurp* (MMV/WHO) to 66% for *msp1/msp2/glurp* (2/3).

Conclusions

- AmpSeq was the most robust technique with less discordance than with microsatellites or *msp1/msp2/glurp*.

- Adding *msp1/msp2* to microsatellites or AmpSeq increased the percentage of new infections in moderate to high transmission areas.

- AmpSeq is simple, rapid and accurate, but the equipment can be costly and it is more expensive if only a few samples are tested.

- The algorithm applied to determine recrudescence rates had no significant effect in areas of low endemicity, but in areas of high transmission with the same markers (*msp1/msp2/glurp*), the recrudescence rate was 66% with the 2/3 algorithm versus 13% using WHO/MMV.

- Although AmpSeq (WHO/MMV) and *msp1/msp2/glurp* (2/3) gave similar recrudescence rates, the results for the individual samples were not concordant between the two methods (concordance = 61%, 56/92 paired samples).

- In high transmission settings, ideally AmpSeq would be used, but microsatellites could be applied if AmpSeq is not operationally feasible.

Note: This analysis was performed on a random sample and the impact on therapeutic efficacy rates cannot be estimated. However, expanding the study to include all the TES data could provide these data and enable assessment of the effect of the different genotyping and analysis methods on drug efficacy rates.

Comparison of whole sequencing vs *msp1, msp2* and *glurp* vs microsatellites using African samples (J. Juliano)

Methods

A TES was conducted at six sites in the Democratic Republic of the Congo in 2017/18 evaluating AL, ASAQ and DP in children with uncomplicated *P. falciparum* malaria. PCR correction employed seven neutral microsatellites analysed using two methods:

- Counting-based method: recrudescence defined as a pre-identified number of identical alleles between day 0 and day of failure sampled across the seven loci (7 of 7 match);

- Bayesian statistical method: a statistical algorithm assigns a probability of recrudescence to each late treatment failure given the observed data.

Findings and further investigation

The counting method resulted in higher PCR-corrected efficacy rates than the Bayesian approach, and, in Mikalayi, the Bayesian algorithm returned therapeutic efficacies less than 90% for AL and DP. These findings were investigated further using the following approaches to estimate antimalarial efficacy for AL (n=34), ASAQ (n=20) and DP (n=39) and the complexity of infection (COI) in Mikalayi:

- *msp1/msp2/glurp* using either a strict bandwidth (*msp1* and *msp2* binned at 3±1.5 bp size and *glurp* at 20±10 bp) or a wide bandwidth (*msp1* and *msp2* binned at 3±5 bp size and *glurp* at 20±50 bp) and analysed using the WHO/MMV or the 2/3 algorithm;

- microsatellites (*TA1, Poly-α, Pfpk2, TA109, TA2490, C2M34* and *C3M69*) analysed using a Bayesian algorithm;

- MIPs *(22)*: A genome-wide SNP image barcode (IBC) was used to examine *P. falciparum* parasite relatedness and COI *(23,24)*, and microhaplotypes from the *P. falciparum* heterozygome (HeOME) were used to examine *P. falciparum* parasite relatedness and COI *(25)*.

Interpretation of COI was defined as:

- WHO: the number of alleles detected at the most diverse of the three genotyped loci;

- microsatellites: the number of alleles detected at the most diverse of the seven genotyped loci;

- HeOME/IBC: calculated using a Markov Chain Monte Carlo method (THE REAL McCOIL) *(26)*.

Summary

- The IBC and HeOME data are incomplete, but at present show similar efficacy rates to the standard *msp1/msp2/glurp* (WHO/MMV) method, and higher efficacy than with microsatellites (Bayesian) or *msp1/msp2/glurp* (2/3) for all three drugs. However, further work is needed to refine the analysis for IBC and HeOME.

- The WHO/MMV method had the highest mean COI (3.4 clones), followed by HeOME (2.8 clones), IBC (2.2 clones), and microsatellites (2.1 clones). The higher value for the WHO/MMV method was driven predominantly by *msp1*.

Conclusions

- In high transmission settings, efficacy estimates are impacted by the genotyping method and the interpretation method, and the 'true' value cannot be determined. Consequently, there is a need to account for uncertainty, and this may become more important in next generation sequencing.

- Probabilistic infection classification will be needed to account for the underlying population distribution of SNP/haplotypes; missing genotype loci or SNP haplotypes in the data; depth of analysis at each loci; and parasitaemia of the sample (risk of false positives).

- Higher multiplex and WGS methods may be more appropriate in low to moderate transmission settings where within-host complexity decreases but parasites within the population are more highly related. A smaller number of highly diverse targets is likely superior in high transmission settings.

Day 2: Discussion

Microsatellites

Analysis of microsatellites using the Bayesian algorithm for the Mikalayi (Democratic Republic of the Congo) data resulted in a decrease in the adjusted efficacy rate compared to the counting method; this was particularly evident for AL and DP, although the difference for ASAQ was much smaller.

The choice of microsatellites is important, as some are not diverse enough to be useful for this analysis. The distribution of the microsatellite sizes for Mikalayi was around 20–35 bins for most of the microsatellites, following a generally normal distribution; however, one microsatellite was less diverse with seven bin sizes (*TA2490*). Even in high transmission settings, there is not an insignificant amount of sharing of alleles, so the lower efficacies may be partially driven by that effect. For example, in the SNP data, only around 100–200 out of the 500 SNPs analysed show differences; because there are so many that are heterozygous at those locations, there is a lot of sharing, i.e., the high COI causes both alleles to be present in all the samples.

The choice of microsatellites would need to be justified by the data showing the diversity at the sites where the samples were collected. There may be scope for determining a particular combination of microsatellites or choosing which microsatellites to use for a particular area.

The bins used for microsatellite data were 5 bp or 7 bp, while some strains differed for microsatellites with only 3 bp. Therefore, wider bins tend to overestimate recrudescence by including artificially similar strains that are actually different in the same bin. Using a Bioanalyzer, sensitivity is lower, so that a wider range of bins is needed for the different haplotypes. Where the strict and wide binning approaches were applied (J. Juliano presentation), it was clear that the matches differed between the two methods with a wider bin width increasing the number of recrudescences.

Instead of using a wide bin range, the percentage of the highest peak can be used to remove most of the stutter peaks. This strategy was used for *msp1* and *msp2* to try to remove most of the stutter peaks, even though this also removed some of the repeats, with thresholds of 10% for *msp1/msp2* and 20% for *glurp*. Therefore, although this strategy decreases the possibility of false positives, it also decreases the sensitivity, meaning that some minority strains will be missed. It is not clear where the optimal balance lies, but it may depend on the transmission characteristics of the study site.

Use of *glurp*

Although *glurp* has limitations as a marker, when using the 2/3 method, *glurp* impacts the outcome only when *msp1* and *msp2* are discordant. This is in contrast to the WHO/MMV method in which *glurp* can indicate a reinfection even if *msp1* and *msp2* both indicate recrudescence. Therefore, it might be a good compromise to retain *glurp*, but to limit its impact on the overall outcome.

However, the data indicate that *glurp* is not an optimal marker for discriminating reinfection from recrudescence. The main driver of the discordance between the 2/3 and WHO/MMV methods was *glurp*, which suggests that there is a problem with this marker. One option is that *glurp* could be abandoned and replaced with, for example, microsatellites as an interim solution. Although there is some template competition for some microsatellites, it is not to the same extent as for *glurp*. The use of *msp1/msp2* plus microsatellites would not cost significantly more than the current genotyping methods. The strategy would then move from length polymorphism to microhaplotype comparisons in the future.

AmpSeq

This appears to be the best method overall for genotyping. It has good resolution when the MOI is >2, so this is particularly relevant to areas of high transmission. Although the AmpSeq method (match-counting) gave similar recrudescence rates to *msp1/msp2/glurp* analysed using the 2/3 algorithm, the results for the individual samples were discordant. This is because the errors in the 2/3 method balance each other out *(17)*, whereas AmpSeq is more accurate.

The cost may be a barrier to implementation at present. In Africa, there are some places where AmpSeq is already being applied. This enables field testing of the methodology during a transition period. The cost of doing large runs is no more than for *msp1/msp2/glurp*. However, the re-sequencing of samples on a small scale is expensive. In addition, there are implementation questions regarding access to reagents in Africa.

The evaluation of new genotyping approaches should be evaluated in tandem with new analysis approaches because they inform each other. Therefore, if AmpSeq is being considered, the analysis method it is paired with should be able to indicate the depth of coverage, how many replicates are required, and therefore what the final cost might be and how sustainable it could be.

When it is done well, CE has similar results to AmpSeq. Therefore, the issue is more which algorithm should be used to interpret the genotyping data (match-counting, 2/3, Bayesian, or another). However, an additional advantage of AmpSeq is that drug resistance markers can be examined on the same platform. It is also possible to quantify whether some molecular markers have been selected following antimalarial treatment. This might help to verify if there is really a recrudescent infection.

Whole genome sequencing

At present it is unknown how minor allele frequency of SNPs in the population impacts the distribution. This will be looked at by J. Juliano when re-sequencing. It is expected that most SNPs will have a very low minor allele frequency, and if those with higher values are selected, it might improve their ability to resolve infections. Furthermore, co-transmission may impact MOI/COI estimations and may need to be considered when interpreting the data in high transmission settings. There will likely be variation by site.

Implications of sensitivity

Increasing sensitivity may reach the point where gametocytes that may survive drug treatment are detected. For example, if there is a higher level of gametocyte carriage with a particular drug, it might lead to a greater likelihood of gametocytes being identified as recrudescence if more sensitive genotyping methods are used. Therefore, it is important to set the sensitivity to a level that best reports drug efficacy against asexual parasites. There is also a danger of increasing sensitivity in the sample selection. If light microscopy is not used but a more sensitive method, then parasites that are not clinically relevant may be detected.

Conversely, high sensitivity may be needed to detect resistance. One issue with the TES is that it is not sensitive enough to detect resistance; in Africa, there is no correlation between the prevalence of the molecular marker and the result of the TES. So, to detect resistant parasites early, the genotyping methods need to be very sensitive.

Triangulation with other data

Drug resistance markers: The addition of four microsatellite markers to *msp1* and *msp2* genotyping resulted in a reclassification of outcomes that strengthened the association between *dhfr59R*, an anti-folate resistance mutation, and recrudescence *(7)*. This association could be looked at in order to compare the different methods.

High versus low transmission: Parasitaemia levels, MOI, parasite relatedness, immunity, etc. are all factors that can vary by transmission setting. For both field data and modelling, looking at the markers/analysis methods with respect to these factors could indicate where there might be biases.

Managing uncertainty

The Bayesian method has uncertainty already incorporated. However, measures of uncertainty could be generated for other methods based on the degree of discordance in the classification of reinfection or recrudescence, particularly in the case of common alleles that are more likely to re-occur for whatever marker is being used. No data type is perfect, and this is why quantifying uncertainty is important. Probabilistic classification models can incorporate error models that tolerate mistakes, e.g., THE REAL McCOIL method to estimate COI, which has a website with a graphical user interface to encourage wide uptake.

Field data versus modelling

- The data presented by P. Ringwald indicate that the 2/3 algorithm increased recrudescence rates and that this effect was magnified in areas of high transmission. This is consistent with the modelling predictions, with the same magnitude – about a two-fold increase versus the WHO/MMV algorithm.
- The AmpSeq and 2/3 algorithm presented by D. Ménard are aligned with the simulations that show higher recrudescence than with the WHO/MMV method.
- The Bayesian analysis of microsatellite data presented by J. Juliano also show good agreement with the model predictions and again return higher recrudescence rates than with the WHO/MMV method.

All the methods differ, and it is not possible to determine which is the most accurate. As there is no gold standard, the 'true' underlying failure rate cannot be determined. Accordingly, it is a question of what evidence is needed to drive policy.

Note that the model applied a sensitivity analysis using a drug with a low failure rate (1%). The results indicated that none of the methods assessed would incorrectly identify an effective drug as a failing drug. In the model, there was little difference between the 2/3 and the WHO/MMV algorithms in low transmission settings. However, the 2/3 algorithm was closer to the model 'true' rate in high transmission areas, except in the case of AL (see below) for which the WHO/MMV method was more appropriate, as the 2/3 method overestimated recrudescence. This is in agreement with the TES data presented by P. Ringwald.

Treatment efficacy with artemether-lumefantrine

The 2/3 method and the Bayesian approach both increased recrudescence rates relative to the WHO/MMV method, particularly for AL, but mainly in high transmission settings. Applying the current treatment efficacy thresholds of 90% efficacy could result in AL being removed from common use in some countries. However, there is not enough evidence to support any policy change for AL, which is a very popular and safe drug.

Any recommendation to switch from AL cannot be taken lightly. In particular, there are cost and access issues to consider for the management of malaria in the field.

There are several alternative explanations for the findings with AL:

- They could be an artefact of the decision algorithms leading to an overestimation of the recrudescence rate for drugs with a short half-life in areas with high reinfection rates.
    - Lumefantrine has the shortest half-life of all artemisinin-based combination therapies (ACTs) and the effect was only seen in areas of high transmission, i.e., where reinfection rates are higher.
    - It may be possible to investigate this by examining treatment efficacy as a function of the MOI. However, as the MOI increases, failure rates may also increase as more clones have to be cleared *(18)*.
    - The modelling studies predict the higher failure rate of AL using the 2/3 method. When there are more than 10 infections per year, with AL, the WHO/MMV method appears to be more accurate, while the 2/3 method overestimates recrudescence by ~3–4% for that setting/drug (Jones, et al. AAC, 2019: Supplementary file 2, Figure S15) *(17)*.
    - This is consistent with the TES data that indicate apparent reductions in AL efficacy in areas of high transmission when the 2/3 analysis is used. By contrast, the WHO/MMV method indicates higher rates of drug efficacy at these sites.
- AL could be losing efficacy in African high transmission settings because of emergence of resistance to lumefantrine.
    - The available data in Africa do not indicate a change over time in AL efficacy. Regular TESs in the same sites are from Angola collected by the US-CDC: AL efficacy has hovered around the 90% mark since 2013 with no clear trend. However, these TESs may not be large enough to detect small changes in drug efficacy around 90%.
    - Ex vivo data do not suggest AL resistance in Africa.
    - AL efficacy over time could be examined in the WHO database to see whether there is any trend towards reduced efficacy in Africa, which was not really observed in the Greater Mekong Subregion.
    - In high transmission areas, higher numbers of patients need to be enrolled so that by study end there is a sufficient number of patients (at least 100) available for a valid analysis of the primary efficacy endpoint.
    - The possibility of resistance could be further investigated by triangulation with ex vivo data and lumefantrine blood levels.
- Adherence to drug therapy might be lower with AL given the twice-daily dosing regimen compared to once-daily ACTs, causing a higher failure rate. AL also has high inter-patient variation in pharmacokinetics and must be taken with a fatty meal to improve absorption; this could impact effectiveness.
- There is the possibility that gametocytaemia could be greater with AL and this might influence the recrudescence rate.

Thresholds for efficacy

It is clear that there is an interaction between the genotyping/analysis method and the estimated efficacy. Accordingly, efficacy may be below the 90% threshold for some methods and above the threshold for others. Different markers and more complex

methods for classifying efficacy rates may require a more nuanced approach to setting the threshold below which a drug is considered to be failing.

More sensitive techniques for detecting minority strains can change the landscape of the investigation beyond drug efficacy, as asymptomatic carriage might also be detected. Consequently, a new approach to considering what is meant by drug efficacy may be required.

If the same markers are used (*msp1/msp2/glurp*) but a different analysis method is applied (2/3 rather than WHO/MMV), perhaps the same threshold should be used. However, the application of thresholds has major implications for drugs currently in use and ones in development. It is not just the percentage value that needs to be considered, but also the context.

Note that WHO studies are now being published using both the 2/3 method and the WHO/MMV method of analysis.

Implications of methods and thresholds for drug development versus TES

Although the differences in treatment efficacy between the different methods are generally <4% and not statistically significant, for drug development, the difference between 96% and 92% efficacy determines whether a drug progresses through the pipeline. Therefore, the interaction between the genotyping methods/analysis and thresholds at which decisions are made regarding a drug's efficacy needs to be examined. Furthermore, in drug development, the intention-to-treat population must be considered, not just the per-protocol population.

Getting towards the 'truth' in terms of showing complete clearance of parasites is important in drug development. The questions around AL are already affecting drug development in terms of how these data should be interpreted considering the uncertainty around the values. In clinical trials for drug development, different standards are required compared to what is acceptable for TESs.

A key consideration is how data that are generated using new methods can be interpreted against historical data. This may require the data to be analysed in two ways – the old way and the new way – while the implications of any changes are being reviewed. In particular, for TESs, there needs to be a mechanism to evaluate if there has been a change in efficacy that indicates resistance emergence.

Regulatory drug trials are comparative, so the methods used within a trial to evaluate reinfection versus recrudescence enables new drugs to be evaluated against existing treatments. However, in countries, the data on new drugs will inevitably be compared with historical data on established therapies. It may, therefore, be difficult to persuade countries that new treatments are effective if the efficacy rates appear to be lower. Even just changing *glurp* to microsatellites changes the efficacy rates, as does using the 2/3 method instead of the WHO/MMV method of analysis. Consequently, genotyping data may need to be generated and analysed in both the 'old' and 'new' ways, at least for a transition period.

Malaria drug treatment can be a challenge for regulators because of the complexities of the disease. At present, the United States Food and Drug Administration (FDA) does not recognize PCR-adjusted efficacy because it has not been validated. There is a risk that the European Medicines Agency (EMA) will also decide that this method is not acceptable. AmpSeq may be the way forward, but it might be too expensive

for countries; it can be used in regulatory trials, however. The regulatory authorities may be reassured if it can be shown that the methods being used are state of the art. However, there is a need to compare regulatory studies conducted in different settings, and this may be difficult if different methods have been used. There cannot be one standard for drug registration and another standard for TESs or post-registration studies. However, it may be possible to agree with regulators on a transition method to enable comparisons while AmpSeq capacity is expanded in Africa. In the meantime, AmpSeq and the selected interim TES methodologies could both be reported in the initial post-licensure studies to benchmark the field efficacy of the new drug.

Issues around implementation

The Bayesian approach is complex and may be difficult to implement at the surveillance level in countries within malaria case management groups. Analytical methods need to be as simple as possible for implementation as tools in national malaria control programmes. Not everyone has access to an external research group and they must be able to do the work themselves. However, although uncertainty may take more time to explain, it may help decision-makers to understand the degree of uncertainty around findings, thereby supporting more informed policy decisions.

From an operational perspective, there needs to be one method that applies to both low and high transmission areas. The interim situation may require the best of the available methods to be used, but with a longer term aim to research and validate new technologies.

Improving quality control for laboratory methods in both field and reference laboratories is critical, and control panel parasites are lacking for most in the field. Quality control should include both microscopy and PCR. Moving forward, setting up regional reference laboratories for molecular methods in Africa and other endemic regions is also important.

References

1. Plucinski MM, Hastings IM, Moriarty LF, Venkatesan M, Felger I, Halsey ES. Variation in calculating and reporting antimalarial efficacy against Plasmodium falciparum in Sub-Saharan Africa: a systematic review of published reports. Am J Trop Med Hyg. 2021. https://www.ncbi.nlm.nih.gov/pubmed/33724925

2. Messerli C, Hofmann NE, Beck HP, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. Antimicrob Agents Chemother. 2016;61(1):e01500-16. https://www.ncbi.nlm.nih.gov/pubmed/27821442

3. Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, Felger I. Comparison of Plasmodium falciparum allelic frequency distribution in different endemic settings by high-resolution genotyping. Malar J. 2009;8:250. https://www.ncbi.nlm.nih.gov/pubmed/19878560

4. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 2017;18(1):864. https://www.ncbi.nlm.nih.gov/pubmed/29132317

5. Gruenberg M, Lerch A, Beck HP, Felger I. Amplicon deep sequencing improves Plasmodium falciparum genotyping in clinical trials of antimalarial drugs. Sci Rep. 2019;9(1):17790. https://www.ncbi.nlm.nih.gov/pubmed/31780741

6. Adegbite BR, Edoa JR, Honkpehedji YJ, Zinsou FJ, Dejon-Agobe JC, Mbong-Ngwese M, et al. Monitoring of efficacy, tolerability and safety of artemether-lumefantrine and artesunate-amodiaquine for the treatment of uncomplicated Plasmodium falciparum malaria in Lambarene, Gabon: an open-label clinical trial. Malar J. 2019;18(1):424. https://www.ncbi.nlm.nih.gov/pubmed/31842893

7. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal Plasmodium falciparum infections. Am J Trop Med Hyg. 2006;75(5):836-42. https://www.ncbi.nlm.nih.gov/pubmed/17123974

8. Beshir KB, Diallo N, Sutherland CJ. Identifying recrudescent Plasmodium falciparum in treated malaria patients by real-time PCR and high resolution melt analysis of genetic diversity. Sci Rep. 2018;8(1):10097. https://www.ncbi.nlm.nih.gov/pubmed/29973679

9. McCollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA. Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrob Agents Chemother. 2007;51(6):2085-91. https://www.ncbi.nlm.nih.gov/pubmed/17283199

10. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol Biol Evol. 2000;17(10):1467-82. https://www.ncbi.nlm.nih.gov/pubmed/11018154

11. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, et al. A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. Science. 1999;286(5443):1351-3. https://www.ncbi.nlm.nih.gov/pubmed/10558988

12. Plucinski MM, Morton L, Bushman M, Dimbu PR, Udhayakumar V. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrob Agents Chemother. 2015;59(10):6096-100. https://www.ncbi.nlm.nih.gov/pubmed/26195521

13. Mwangi JM, Omar SA, Ranford-Cartwright LC. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing Plasmodium falciparum infections from reinfections in Kenya. Int J Parasitol. 2006;36(3):329-36. https://www.ncbi.nlm.nih.gov/pubmed/16442537

14. Henden L, Lee S, Mueller I, Barry A, Bahlo M. Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. PLoS Genet. 2018;14(5):e1007279. https://www.ncbi.nlm.nih.gov/pubmed/29791438

15. Early AM, Daniels RF, Farrell TM, Grimsby J, Volkman SK, Wirth DF, et al. Detection of low-density Plasmodium falciparum infections using amplicon deep sequencing. Malar J. 2019;18(1):219. https://www.ncbi.nlm.nih.gov/pubmed/31262308

16. Ghansah A, Kamau E, Amambua-Ngwa A, Ishengoma DS, Maiga-Ascofare O, Amenga-Etego L, et al. Targeted next generation sequencing for malaria research in Africa: current status and outlook. Malar J. 2019;18(1):324. https://www.ncbi.nlm.nih.gov/pubmed/31547818

17. Jones S, Kay K, Hodel EM, Chy S, Mbituyumuremyi A, Uwimana A, et al. Improving methods for analyzing antimalarial drug efficacy trials: molecular correction based on length-polymorphic markers msp-1, msp-2, and glurp.

Antimicrob Agents Chemother. 2019;63(9). https://www.ncbi.nlm.nih.gov/pubmed/31307982

18. Jaki T, Parry A, Winter K, Hastings I. Analysing malaria drug trials on a per-individual or per-clone basis: a comparison of methods. Stat Med. 2013;32(17):3020-38. https://www.ncbi.nlm.nih.gov/pubmed/23258694

19. Jones S, Plucinski M, Kay K, Hodel EM, Hastings IM. A computer modelling approach to evaluate the accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of antimalarial drug efficacy. Antimicrob Agents Chemother. 2020;64(4). https://www.ncbi.nlm.nih.gov/pubmed/31932376

20. Jones S, Kay K, Hodel EM, Gruenberg M, Lerch A, Felger I, et al. Should deep-sequenced amplicons become the new gold-standard for analysing malaria drug clinical trials? bioRxiv. 2021:03.23.436602.

21. Mohd Abd Razak MR, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, et al. Genetic diversity of Plasmodium falciparum populations in malaria declining areas of Sabah, East Malaysia. PLoS One. 2016;11(3):e0152415. https://www.ncbi.nlm.nih.gov/pubmed/27023787

22. Aydemir O, Janko M, Hathaway NJ, Verity R, Mwandagalirwa MK, Tshefu AK, et al. Drug-resistance and population structure of Plasmodium falciparum across the Democratic Republic of Congo using high-throughput molecular inversion probes. J Infect Dis. 2018;218(6):946-55. https://www.ncbi.nlm.nih.gov/pubmed/29718283

23. Verity R, Aydemir O, Brazeau NF, Watson OJ, Hathaway NJ, Mwandagalirwa MK, et al. The impact of antimalarial resistance on the genetic structure of Plasmodium falciparum in the DRC. Nat Commun. 2020;11(1):2107. https://www.ncbi.nlm.nih.gov/pubmed/32355199

24. Moser KA, Madebe RA, Aydemir O, Chiduo MG, Mandara CI, Rumisha SF, et al. Describing the current status of Plasmodium falciparum population structure and drug resistance within mainland Tanzania using molecular inversion probes. Mol Ecol. 2021;30(1):100-13. https://www.ncbi.nlm.nih.gov/pubmed/33107096

25. Tessema SK, Hathaway NJ, Teyssier NB, Murphy M, Chen A, Aydemir O, et al. Sensitive, highly multiplexed sequencing of microhaplotypes from the Plasmodium falciparum heterozygome. J Infect Dis. 2020. https://www.ncbi.nlm.nih.gov/pubmed/32840625

26. Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, et al. THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. PLoS Comput Biol. 2017;13(1):e1005348. https://www.ncbi.nlm.nih.gov/pubmed/28125584

For further information please contact:

Global Malaria Programme
World Health Organization
20, avenue Appia
CH-1211 Geneva 27
Switzerland
Email: GMPinfo@who.int

9789240038363