Towards evidence-based malaria guidelines in low- and middle-income countries.

David Jonathan Sinclair MB.ChB, MPH, DTM&H

Submitted for consideration for the degree of Doctor of Philosophy by published work.

Health Sciences
Warwick Medical School
University of Warwick
January 2017
Table of Contents

Acknowledgements ........................................................................................................ iii
Statement of ethical considerations ........................................................................ iv
Submission declaration .......................................................................................... iv
Word count ................................................................................................................ iv
Statement of candidate’s contribution to the published work ............................... v
Abstract ...................................................................................................................... ix
Abbreviations (in order of appearance) .................................................................. x

Chapter 1. Developing global and national clinical guidelines ..................... 1
Introduction ............................................................................................................... 1
Current standards in guideline development ......................................................... 1
The role of the Cochrane Collaboration ................................................................. 3
The role of the World Health Organization ........................................................... 4
The aim and scope of this PhD submission ......................................................... 5

Chapter 2. Which first-line treatment for uncomplicated malaria? .......... 7
Background ............................................................................................................. 8
Policy question ...................................................................................................... 9
Methods .................................................................................................................. 9
Key findings ........................................................................................................... 11
Strengths and limitations of this approach ......................................................... 12
Influence on WHO recommendations ............................................................. 16

Chapter 3. Which first-line treatment for severe malaria? ....................... 19
Background ........................................................................................................... 20
Policy question ..................................................................................................... 21
Methods ................................................................................................................ 21
Key findings .......................................................................................................... 22
Strengths and limitations of this approach ....................................................... 24
Influence on WHO recommendations ............................................................. 26

Chapter 4. Which diagnostic approach for malaria? ............................... 27
Background ........................................................................................................... 28
Policy question ..................................................................................................... 29
Methods ................................................................................................................ 30
Key findings .......................................................................................................... 31
Strengths and limitations of this approach ....................................................... 34
Influence on WHO recommendations ............................................................. 36

Chapter 5. Facilitating national guideline development ............................ 37
Background ........................................................................................................... 38
Policy questions .................................................................................................... 39
Methods ................................................................................................................ 39
Key findings .......................................................................................................... 40
Strengths and limitations of this approach ....................................................... 41
Influence on WHO recommendations ............................................................. 44

Chapter 6. Conclusions .................................................................................... 47
The scientific contribution of this work ............................................................. 47
The developing role for academics at the interface between research and policy.... 52
Our experience as info-mediaries .................................................................... 53
Preparing global guidelines for contextualization at national level........................................57

References .............................................................................................................................. 59

Appendix 1. Full bibliography of candidate’s published work .............................................. 64

Appendix 2. Statements of candidate’s contributions .......................................................... 67

Appendix 3. The published work ......................................................................................... 68

List of figures

Figure 1: The GRADE approach to rating the quality of evidence. Adapted from Guyatt 2008. [20] ................................................................. ................................................................. 2

Figure 2: The GRADE approach to defining the strength of a recommendation. Adapted from Guyatt 2008. [21] ................................................................. ................................................................. 2

Figure 3: The six domains of the AGREEII Instrument for appraising the quality of healthcare guidelines. Adapted from Brouwers 2010. [22] ................................................................. 3

Figure 4: A standardized approach to handling missing data across trials. Developed from Bloland 2003. [37] ................................................................. 10

Figure 5: Forest plot of randomized trials, comparing dihydroartemisin-piperaquine (DHA-P) against all comparators, for the primary outcome: PCR-adjusted treatment failure at the longest available time-point. Taken from Sinclair 2009. [1] ................................................................................................. 12

Figure 6: Summary of findings table for adverse events with dihydroartemisin-piperaquine versus artemether-lumefantrine. Adapted from Zani 2014. [6] ........................................ 15

Figure 7: Evidence box in support of the WHO recommendation to use artesunate in preference to quinine in adults with severe malaria. Taken from the WHO Malaria Treatment Guidelines: Second edition. [30] ................................................................. 20

Figure 8: Forest plot of randomized trials comparing artesunate with quinine for the primary outcome: death. Taken from Sinclair 2012. [2] ................................................................................................. 24

Figure 9: Logic framework for modeling the effect on health outcomes of introducing RDT-supported diagnosis compared to clinical diagnosis alone. Taken from Odaga 2014. [3] .... 29

Figure 10: Table of characteristics of interventions in trials comparing RDT-supported diagnosis with clinical diagnosis alone. Modified from Odaga 2014. [3] ................................................................. 33

Figure 11: Forest plot of randomized trials comparing RDT-supported diagnosis with clinical diagnosis alone for the outcome: anti-malarial prescriptions. Taken from Odaga 2014. [3] 34

Figure 12: Forest plot of zinc supplementation versus placebo for the outcome: duration of diarrhoea. Adapted from Lazzerini 2008. [71] ................................................................................................. 43
Acknowledgements

Special thanks to Paul Garner and Anne-Marie Stephani for 10 years of enthusiasm, inspiration, mentorship, and friendship.


Finally, thanks to my PhD supervisors Noel McCarthy and Olalekan Uthman for encouragement and guidance.
Statement of ethical considerations

Of the submitted work, only one included the recruitment of living people:


The study protocol for the qualitative aspect of this paper was discussed with both the Liverpool School of Tropical Medicine Ethics Committee and the WHO Ethics Review Committee and received a formal written waiver. Permission to conduct the study was granted by the Assistant Director General of the WHO and all participants provided written informed consent prior to being interviewed.

Submission declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree, apart from Odaga 2014 which also formed part of the PhD thesis of the first author (Odaga J, Sinclair D, Lokong JA, Donegan S, Hopkins H, Garner P. Rapid diagnostic tests versus clinical diagnosis for managing people with fever in malaria endemic settings. Cochrane Database of Systematic Reviews. 2014; Issue 4).

All of the submitted published works were collaborative. My individual contribution to each of them is clearly stated in the section titled ‘Statement of candidate’s contribution to the published work’, and supported by signed statements by the senior author of each paper where necessary (see appendix 3).

Word count

9903 excluding tables and references.
Statement of candidate’s contribution to the published work

<table>
<thead>
<tr>
<th>Main submitted work</th>
<th>Published authorship statement</th>
<th>Additional statement agreed with co-authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinclair D, Zani B, Donegan S, Olliaro P, Garner P. Artemisinin-based combination therapy for treating uncomplicated malaria. Cochrane Database of Systematic Reviews. 2009; Issue 3. [1]</td>
<td>All authors were involved in the conception and design of the protocol. Data extraction and assessment of risk of bias was performed by DS and BZ. DS, PO, and PG worked on the analysis of secondary outcomes. Data input and analysis was conducted by DS with input from PO and PG and statistical advice from SD. The text was drafted by DS with input from all other authors.</td>
<td>David Sinclair designed the protocol with input from all other authors. David Sinclair and Babalwa Zani conducted the data extraction and assessment of risk of bias. David Sinclair conducted the analysis and wrote the first draft of the paper with input from all other authors. David Sinclair led the response to referees and finalized the paper.</td>
</tr>
<tr>
<td>Sinclair D, Donegan S, Isba R, Laloo DG. Artesunate versus quinine for treating severe malaria. Cochrane Database of Systematic Reviews. 2012; Issue 6. [2]</td>
<td>Katharine Jones (KJ) and SD assessed the eligibility and methodological quality of trials, extracted and analysed data, and completed the first published version of the review. DS replaced KJ for the 2011 update of this review. DL contributed to the design and writing of the review. RI designed and conducted the economic commentary.</td>
<td>David Sinclair led the new editions of this review in 2011 and 2012. He checked the data entry and updated the risk of bias assessments. He conducted data extraction and analysis for the additional trials with Sarah Donegan, completed the summary of findings tables and rewrote the text with input from all other authors.</td>
</tr>
<tr>
<td>Odaga J, Sinclair D, Lokong JA, Donegan S, Hopkins H, Garner P. Rapid diagnostic tests versus clinical diagnosis for managing people with fever in malaria endemic settings. Cochrane Database of Systematic Reviews. 2014; Issue 4. [3]</td>
<td>JO wrote the protocol and the first draft of the review during his PhD programme at Liverpool School of Tropical Medicine, with input from PG and JAL. JO was responsible for data entry and analysis. JO and JAL independently screened studies for inclusion, assessed risk of bias from the included trials and extracted data from them. PG provided supervision throughout the review process. SD provided advice</td>
<td>David Sinclair became an author on this review after the first draft was written. The review was in a very draft form, with limited analysis, some of which needed checking. David Sinclair checked and revised the data entry and risk of bias assessments, created the logic framework, created the additional tables describing the characteristics of included studies, the intervention, and the relationship between endemicity and health worker adherence to the test. He revised the analysis in light of these tables, revised the</td>
</tr>
</tbody>
</table>
on the statistical methods. DS reviewed all sections of the review; in particular the trial outcomes, Summary of findings tables, main results, discussion, and conclusion. HH provided unpublished trial data and contributed to the interpretation, and discussion. All authors contributed to the final interpretation and results.

| Sinclair D, Isba R, Kredo T, Zani B, Garner P. World Health Organization guideline development: An evaluation. PLoS ONE. 2013;8(5):e63715. [4] | Conceived and designed the experiments: DS, RI, TK, BZ, HS, PG. Performed the experiments: DS, RI, TK, BZ, PG. Analyzed the data: DS, RI, TK, BZ, PG. Wrote the paper: DS, RI, TK, BZ, HS, PG. | David Sinclair designed the project and wrote the methods with input from all other authors. He coordinated the AGREEII appraisals and analyzed the data. Paul Garner and David Sinclair carried out the qualitative interviews jointly. David Sinclair analyzed the transcripts and conducted the qualitative analysis with input from Paul Garner and Helen Smith on the framework, themes and coding. David Sinclair took the lead responsibility in journal submission and responding to referees. |
| Sinclair D, Gyansa-Lutterodt M, Asare B, Koduah A, Andrews E, et al. Integrating global and national knowledge to select medicines for children: The Ghana National Drugs Programme. PLoS Medicine. 2013;10(5):e1001449. [5] | Wrote the first draft of the manuscript: DS. Contributed to the writing of the manuscript: DS MG BA EA AK PG. ICMJE criteria for authorship read and met: DS MG BA EA AK PG. Agree with manuscript results and conclusions: DS MG BA EA AK PG. | David Sinclair designed the project with guidance from Paul Garner. David worked in Ghana with the National Drugs Programme; designed and conducted the training and support to participants throughout the project; and designed the summary sheets for the committee to consider. David Sinclair wrote the first draft of the manuscript with input from all other authors, finalized the manuscript, and led the response to referee comments. |

*See appendix 1 for statements of contribution confirmed and signed by co-authors.*
<table>
<thead>
<tr>
<th>Additional contributing work</th>
<th>Published authorship statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zani B, Gathu M, Donegan S, Olliaro PL, <strong>Sinclair D.</strong> Dihydroartemisinin-piperaquine for treating uncomplicated <em>P. falciparum</em> malaria. Cochrane Database of Systematic Reviews. 2014; Issue1. [6]</td>
<td>DS, BZ, SD, and PO developed the protocol as used in [1]. For this update, BZ and MG reviewed the reference list, extracted data, and entered it into Review Manager. BZ, MG, and DS conducted the analyses, constructed summary of findings tables, and evaluated the quality of evidence using the GRADE approach. All authors reviewed and edited the final draft.</td>
</tr>
<tr>
<td>Bukirwa H, Unnikrishnan B, Kramer CV, <strong>Sinclair D</strong>, Nair S, Tharyan P. Artesunate plus pyronaridine for treating uncomplicated <em>Plasmodium falciparum</em> malaria. Cochrane Database of Systematic Reviews. 2014; Issue 3. [7]</td>
<td>BU and SN co-drafted the initial version of the protocol. HB revised the protocol, and together with PT independently selected trials, assessed quality, extracted and entered data that was checked by BU and SN. CVK extracted adverse events data. HB used GRADE profiler to create and import ‘Summary of findings’ tables. HB wrote the initial draft of the review and worked with all the authors to finalise the review. All authors approved the final review version.</td>
</tr>
<tr>
<td>Isba R, Zani B, Gathu M, <strong>Sinclair D.</strong> Artemisinin-naphthoquine for treating uncomplicated <em>Plasmodium falciparum</em> malaria. Cochrane Database of Systematic Reviews. 2015; Issue 2. [8]</td>
<td>DS and BZ contributed to the development of the standard protocol as used in [1]. BZ drafted the background. RI and MG reviewed the reference list, extracted data, and entered it into Review Manager. RI, MG, and DS conducted the analyses, constructed 'Summary of findings' tables, and evaluated the quality of evidence using the GRADE approach. RI wrote the first draft and all authors reviewed and contributed to the final draft.</td>
</tr>
<tr>
<td><strong>Sinclair D</strong>, Gogtay N, Brand F, Olliaro PL. Artemisinin-based combination therapy for treating uncomplicated <em>Plasmodium vivax</em> malaria. Cochrane Database of Systematic Reviews. 2011; Issue 7. [9]</td>
<td>All authors were involved in the conception and design of this protocol. NG, FB and DS extracted the data. DS analyzed the data and wrote the first draft with NG. All authors assisted with revising subsequent drafts and preparing the final draft for publication.</td>
</tr>
<tr>
<td>Asare B, <strong>Sinclair D.</strong> Evidence summary for artesunate. Supplementary material to [5]. [10]</td>
<td></td>
</tr>
<tr>
<td>Buabin D, Mensah NO, <strong>Sinclair D.</strong> Evidence summary for artemether. Supplementary material to [5]. [11]</td>
<td></td>
</tr>
<tr>
<td>Ochodo E, Garner P, <strong>Sinclair D.</strong> Achieving universal testing for malaria. British Medical Journal. 2016;352:i107. [12]</td>
<td>EO had the original idea, wrote the first draft, provided examples, and coordinated the paper; PG provided examples and contributed to the manuscript; DS guided structure and content of the paper, provided examples, and contributed to the manuscript.</td>
</tr>
<tr>
<td>Okwundu CI, Nagpal S, Musekiwa A, <strong>Sinclair D.</strong> Home- or community-based programmes for treating</td>
<td>Charles Okwundu conceptualised and wrote the first draft of the review. Sukrti Nagpal and Dave Sinclair contributed to the background information, study selection and data</td>
</tr>
</tbody>
</table>
Alfred Musekiwa commented on the draft review and contributed to the data extraction, synthesis, and analysis.


---

Abstract

This PhD submission presents a case study of an academic group working as infomediaries at the interface between research and global policy, and at the interface between global policy and national decisions: advising on methodological issues, conducting systematic evidence reviews in response to information needs, and developing approaches for reinterpreting global guidance for national decision-making.

The included systematic reviews were among the first to adopt innovative elements such as: summary of findings tables, standardized language reflecting the level of certainty in effect estimates, logic frameworks, and brief economic summaries; and have contributed to the further development of these methods. This work has helped to establish formal and transparent methods within global malaria guidance, and contributed to improved standards in global guidance more broadly.
Abbreviations (in order of appearance)

WHO World Health Organization
GRADE Grading of Recommendations Assessment, Development and Evaluation
AGREEII Appraisal of Guidelines for Research and Evaluation Instrument
HIV Human Immunodeficiency Virus
ACT Artemisinin-based Combination Therapies
PCR Polymerase Chain Reaction
RR Risk Ratio
95%CI Ninety-five percent Confidence Intervals
DHA-P DiHydroArtemisinin-Piperaquine
AL6 Artemether-Lumefantrine (6 doses)
M-H Mantel-Haenszel – A statistical approach to meta-analysis
ECG Electrocardiogram
RCT Randomized Controlled Trial
EMA European Medicines Agency
RDT Rapid Diagnostic Test (for malaria)
pLDH plasmodium Lactate DeHydrogenase
HRP-2 Histidine-Rich Protein 2
STG Standard Treatment Guidelines
SE Standard Error
MTG Malaria Treatment Guidelines
GRC Guideline Review Committee
ADAPTE (The meaning of the acronym is unclear)
DECIDE The Developing and Evaluating Communication strategies to support Informed Decisions and practice based on Evidence project
SUPPORT The SUPporting POlicy relevant Reviews and Trials project
Chapter 1. Developing global and national clinical guidelines

Introduction

Over the past 20 years, many high-income countries have recognized the value of national evidence-based clinical guidelines, and established independent bodies to oversee their development [15][16]. Clinical guidelines aim to improve the health of populations by promoting proven treatments and raising standards. They provide a meaningful basis for audit of clinical practice and healthcare provision, and promote the rational, cost-effective use of resources, such as medicines, therapies and laboratory tests.

While the autonomous production of clinical guidelines by individual countries is likely to promote ownership of the recommendations, and influence dissemination and implementation, it is also a duplication of efforts, and potentially wasteful use of resources. Indeed, the cost for developing a clinical guideline following current standards is substantial, and likely to be prohibitive for most low- and middle-income countries.

As such, it seems important to establish ways of reducing duplication, without reducing ownership, and the World Health Organization (WHO) and the Cochrane Collaboration are two organizations ideally placed to facilitate this at a global level [17].

Current standards in guideline development

Currently, the most widely used system for developing evidence-based healthcare recommendations is the ‘Grading of Recommendations Assessment, Development and Evaluation’ approach (GRADE) [18]. This approach was developed by an
international consensus of methodologists in response to the diverse collection of unsatisfactory systems in use in the early 2000s [19].

Crucially, the GRADE working group recognized that evidence of intervention effects was not sufficient to guide recommendations, and inevitably a number of further judgments are required. The GRADE approach therefore separates the appraisal of the ‘quality of the evidence’ (see figure 1) from the ‘strength of the recommendation’ (see figure 2) [20][21].

Figure 1: The GRADE approach to rating the quality of evidence. Adapted from Guyatt 2008. [20]

<table>
<thead>
<tr>
<th>Quality Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Further research is very unlikely to change our confidence in the estimate of effect.</td>
</tr>
<tr>
<td>Moderate</td>
<td>Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.</td>
</tr>
<tr>
<td>Low</td>
<td>Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.</td>
</tr>
<tr>
<td>Very Low</td>
<td>We are very uncertain about the estimate.</td>
</tr>
</tbody>
</table>

Figure 2: The GRADE approach to defining the strength of a recommendation. Adapted from Guyatt 2008. [21]

<table>
<thead>
<tr>
<th>Strength</th>
<th>For patients</th>
<th>For clinicians</th>
<th>For policy makers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Most people in your situation would want the recommended course of action and only a small proportion would not.</td>
<td>Most patients should receive the recommended course of action.</td>
<td>The recommendation can be adopted as policy in most situations.</td>
</tr>
<tr>
<td>Weak</td>
<td>Many people in your situation would want the recommended course of action, but also many would not.</td>
<td>Each patient should be provided with sufficient information to arrive at their own decision, consistent with their values and preferences.</td>
<td>Policy-making at the local level will require substantial debate among stakeholders.</td>
</tr>
</tbody>
</table>
To be truly useful to their target audience, clinical guidelines also need to be free from bias, well presented, and user-friendly. The Appraisal of Guidelines for Research and Evaluation Instrument (AGREEII) was first published in 2003, and lays out the elements of a good guideline. It provides a framework for documenting the guideline process, and can be used to appraise the quality of clinical guidelines against an international standard (see figure 3)[22].

Figure 3: The six domains of the AGREEII Instrument for appraising the quality of healthcare guidelines. Adapted from Brouwers 2010. [22]

<table>
<thead>
<tr>
<th>Scope and purpose:</th>
<th>The description of the overall aim of the guideline, the scope of the questions, and the target audience.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder involvement:</td>
<td>The appropriate involvement of all stakeholders, including the intended users and beneficiaries of the guideline.</td>
</tr>
<tr>
<td>Rigour of development:</td>
<td>The methods used to search for, synthesize, and appraise evidence, formulate recommendations, and keep them updated.</td>
</tr>
<tr>
<td>Clarity of presentation:</td>
<td>The general language, structure, and format of the guideline.</td>
</tr>
<tr>
<td>Applicability:</td>
<td>Discussion of the barriers and facilitators to implementation, and advice or tools to improve uptake and implementation.</td>
</tr>
<tr>
<td>Editorial independence:</td>
<td>The methods used to declare and manage potential conflicts of interest related to the funding body or the guideline group members.</td>
</tr>
</tbody>
</table>

The role of the Cochrane Collaboration

Systematic reviews are a key step in ensuring that clinical recommendations are based on empirical evidence. Following pre-specified protocols, researchers search for, collate, and appraise all the available primary research appropriate to answering current clinical questions [23].

The Cochrane Collaboration has been a major player in the development of systematic review methods and remains a leading producer of systematic reviews worldwide [24]. Since its inception the collaboration has emphasized the need to be independent, and free from commercial, financial and academic conflicts of interest [25]. As the name suggests, it encourages collaboration rather than competition,
bringing together multinational authorship teams. The most common type of Cochrane review assesses the benefits and harms of a particular treatment or intervention, but a full guideline process may also require systematic reviews on disease prognosis, diagnostic test accuracy, or patient experience and preferences [26].

Typically a Cochrane systematic review includes and appraises the global evidence-base regardless of language. However, for many interventions, the benefits and harms that can be expected will be modified by contextual factors such as disease prevalence and virulence, population age, sex, lifestyle, and genetics. Systematic reviews conducted at the global level will therefore only be useful at the national and local levels if they present this information clearly to the reader and adequately explore the nature of the effect modification [27].

**The role of the World Health Organization**

It follows from the GRADE separation of ‘quality of evidence’ from ‘strength of recommendation’ that even when the global evidence-base is broadly applicable, policy makers in different settings may make different decisions [21][28]. Local priorities, local health system resources, and cultural variation in values and preferences all influence if, where, and how much of a particular intervention should be implemented. So what then is the role of the WHO as a producer of global guidance?

One traditional role has been for the WHO to provide guidance for use in countries with little local capacity to develop their own. In this role, WHO guidelines may be targeted directly at local implementers (such as clinical staff), and bypass local
decision-making processes. This paternalistic approach, though necessary at times, risks being biased by global priorities, and ignorant of local insight.

As capacity in many low- and middle-income countries increases, it is likely that this role will diminish with national ministries demanding greater control over their own priorities and resources. As such, the WHO may be required to re-align itself as a facilitator of autonomous national decision-making; setting a framework for global disease control, and providing synthesized evidence on effects, cost, and feasibility for consideration of individual decisions at the national level. In doing so, the primary audience for WHO documents would need to shift from local implementers to local policy makers.

The aim and scope of this PhD submission

I have been part of the WHO Malaria Treatment Guidelines (MTG) committee since 2007, initially as a technical advisor (presenting my systematic reviews of anti-malarial drug efficacy), and later as a guideline methodologist – with input into the guideline process and document format. While the main focus of all three editions has been the drug treatment of malaria illness, some recommendations are also made around malaria diagnosis and prevention [29][30][31].

The primary aim of the included papers was to ensure that malaria recommendations made at the global level were informed by reliable summaries of the available evidence-base. Through the course of the work I have become interested in how the global evidence-base and global recommendations are then utilized by national policy makers.
Chapters 2 and 3 present systematic reviews of anti-malarial drug efficacy for the treatment of uncomplicated (non-severe) malaria and severe malaria respectively, and chapter 4 presents a systematic review of the effects of introducing rapid diagnostic tests in areas without access to light microscopy. These reviews were used to inform the recommendations of the second and third editions of the WHO Malaria Treatment Guidelines.

Chapter 5 presents a formal appraisal of the quality of WHO guidelines, alongside a demonstration project utilizing and re-interpreting global guidance at the national level.

In chapter 6, I discuss my overall conclusions about the implications for guideline development at the World Health Organization, and in low- and middle-income countries.
Chapter 2. Which first-line treatment for uncomplicated malaria?

Main paper


Additional contributing work


Background

The first edition of the WHO Malaria Treatment Guidelines endorsed ‘combination therapy’ as a global strategy for malaria treatment [29]. The development and spread of anti-malarial resistance had rendered most older mono-therapies obsolete, and the panel considered combination therapy essential to protect new anti-malarials against further resistance [32]. This policy mirrored similar approaches in tuberculosis and HIV, and emerged at the time when artemisinin-based combination therapies (ACTs) were being developed.

In 2006 the WHO formally recommended four ACTs: artesunate plus mefloquine, artemether-lumefantrine, artesunate plus amodiaquine, and artesunate plus sulfadoxine-pyrimethamine; and one non-artemisinin-based combination still available and used in several countries: amodiaquine plus sulfadoxine-pyrimethamine [29].

By the time of the second edition, however, national malaria programmes were asking the WHO for more guidance on how to select their national first-line treatment, and consequently we undertook a systematic review to summarize the relative benefits and harms of the five recommended combinations and one new combination which had shown promise in recent trials: dihydroartemisinin-piperaquine.

While several Cochrane reviews had already synthesized the evidence around individual ACT comparisons [33][34][35][36], this review aimed to provide all
relevant head-to-head comparisons within a single review, and allow policy makers to directly compare the relative benefits and harms of each option.

**Policy question**

What are the relative benefits and harms of the recommended ACTs?

**Methods**

We followed standard Cochrane methods to formulate the key questions, construct a protocol, and then formally search for, and appraise, all randomized head-to-head trials of ACTs in uncomplicated *P. falciparum* malaria.

To facilitate appropriate meta-analysis, we needed to standardize the way losses-to-follow-up and missing data were handled across all trials. Our primary analysis therefore followed the WHO's protocol for assessing and monitoring anti-malarial drug efficacy [37][38]. This protocol describes a complete case analysis, which excludes all participants with incomplete or missing outcome data from the efficacy estimate. To test the robustness of this approach, we then conducted a series of sensitivity analyses, which aimed to restore the integrity of the randomization process (see figure 4).

Our primary outcome was treatment failure, defined as the presence of recurrent parasitaemia during the period of follow-up, with Polymerase Chain Reaction (PCR) genotyping used to adjust this result to exclude probable new infections [39]. Secondary outcomes were parasite and fever clearance during the first week, gametocyte clearance, anaemia, and adverse effects.
We used the GRADE approach to evaluate the quality of evidence, and presented the effect estimates for the main outcomes alongside the GRADE judgments in summary of findings tables.

Figure 4: A standardized approach to handling missing data across trials. Developed from Bloland 2003. [37]

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Participants</th>
<th>PCR³-unadjusted</th>
<th>PCR-adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Numerator</td>
<td>Denominator</td>
</tr>
<tr>
<td>Primary analysis</td>
<td>Exclusions after enrolment</td>
<td>Excluded</td>
<td>Excluded</td>
</tr>
<tr>
<td></td>
<td>Missing or indeterminate PCR</td>
<td>Included as failures</td>
<td>Included</td>
</tr>
<tr>
<td></td>
<td>New infections</td>
<td>Included as failures</td>
<td>Included</td>
</tr>
<tr>
<td>Sensitivity analysis 1 c</td>
<td>As 'Primary analysis' except: missing or indeterminate PCR</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sensitivity analysis 2 d</td>
<td>As 'Sensitivity analysis 1' except: new infections</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sensitivity analysis 3 e</td>
<td>As 'Sensitivity analysis 2' except: exclusions after enrolment</td>
<td>Included as failures</td>
<td>Included</td>
</tr>
<tr>
<td>Sensitivity analysis 4 f</td>
<td>As 'Sensitivity analysis 2' except: exclusions after enrolment</td>
<td>Included as successes</td>
<td>Included</td>
</tr>
</tbody>
</table>

Note: participants who were found not to satisfy the inclusion criteria after randomization are removed from all calculations.

- **PCR**: polymerase chain reaction.
- **Excluded**: means removed from the calculation.
- **To re-classify all indeterminate or missing PCR results as treatment failures in the PCR-adjusted analysis.**
- **To re-classify all PCR-confirmed new infections as treatment successes in the PCR-adjusted analysis.** (This analysis may overestimate efficacy as PCR is not wholly reliable and some recrudescences may be falsely classified as new infections. Also some participants may have gone on to develop a recrudescence after the new infection.)
- **To re-classify all exclusions after enrolment (losses to follow-up, withdrawn consent, other antimalarial use, or failure to complete treatment) as treatment failures.** For PCR-unadjusted total failure this represents a true worst-case scenario.
- **To re-classify all exclusions after enrolment (losses to follow-up, withdrawn consent, other antimalarial use, or failure to complete treatment) as treatment successes.**
**Key findings**

Fifty studies met the inclusion criteria, with most conducted between 2003 and 2009. Thirty-one trials were conducted in Africa, 17 in Asia, one in South America, and one in Oceania. Pregnant and lactating women were excluded from all trials. The study populations in Asian trials were older, with exclusion of children aged less than one year, while African studies concentrated on children including those as young as six months.

All five ACTs achieved PCR-adjusted failure rates of less than 10%, in line with the WHO recommended standard (established by consensus in the first edition of the guidelines), at most study sites.

The new ACT, dihydroartemisinin-piperaquine, performed well compared to the most commonly used drugs in specific settings: versus artesunate plus mefloquine in Asia (PCR-adjusted treatment failure: RR 0.39, 95% CI 0.19 to 0.79; three trials, 1062 participants); and versus artemether-lumefantrine in Africa (RR 0.39, 95% CI 0.24 to 0.64; three trials, 1136 participants) (see figure 5).

Artesunate plus mefloquine performed well in trials from Asia and South America, with failure rates consistently low, but was rarely studied in the African context.

Artemether-lumefantrine and artesunate plus amodiaquine performed well in almost all studies they were involved in but single trials from Uganda found failure rates in excess of 10%.

Amodiaquine plus sulfadoxine-pyrimethamine performed poorly in East Africa versus artemether-lumefantrine (RR 0.12, 95% CI 0.06 to 0.24; two trials, 618
participants), and versus artesunate plus amodiaquine (RR 0.44, 95% CI 0.22 to 0.89; three trials, 1515 participants).

Figure 5: Forest plot of randomized trials, comparing dihydroartemisinin-piperaquine (DHA-P) against all comparators, for the primary outcome: PCR-adjusted treatment failure at the longest available time-point. Taken from Sinclair 2009. [1]

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>DHA-P Events Total</th>
<th>Control Events Total</th>
<th>Weight</th>
<th>Risk Ratio</th>
<th>M-HL Random, 95% CI</th>
<th>Year</th>
<th>Risk Ratio</th>
<th>M-HL Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.1.1 Day 63: DHA-P vs Artesunate plus mefloquine</td>
<td>8 191 9 191 29.7%</td>
<td>6 185 6 185 29.7%</td>
<td>0.83 (0.69, 1.00)</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.1.2 Day 42: DHA-P vs Artemether-lumefantrine</td>
<td>16 336 16 336 18.3%</td>
<td>10 336 10 336 18.3%</td>
<td>0.45 (0.37, 0.55)</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.1.3 Day 28: DHA-P vs Artesunate plus amodiaquine</td>
<td>10 232 15 232 75.6%</td>
<td>5 232 5 232 75.6%</td>
<td>0.59 (0.47, 0.74)</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.1.4 Day 42: DHA-P vs Artemisinin plus sulfadoxine-pyrimethamine</td>
<td>10 108 20 108 70.0%</td>
<td>10 108 20 108 70.0%</td>
<td>0.55 (0.39, 0.79)</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.1.5 Day 28: DHA-P vs Amodiaquine plus sulfadoxine-pyrimethamine</td>
<td>10 108 20 108 70.0%</td>
<td>10 108 20 108 70.0%</td>
<td>0.55 (0.39, 0.79)</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Strengths and limitations of this approach

Systematic reviews are not always welcomed by disciplines new to their use, and within infectious diseases some prominent figures had expressed skepticism about the approach due to expected variability in drug effects over time and place associated with drug resistance. It was therefore critical that this review fully considered this heterogeneity in its analysis and conclusions.
To facilitate this, we made several modifications to the standard presentation of forest plots (see figure 5):

- A three-letter country code denoting the trial setting was added to each study ID,
- The year each study finished recruitment was added in the final data column,
- Within each sub-group, trials were arranged chronologically with oldest trials first.

These modifications allowed readers to make their own assessments of applicability and declining efficacy over time, and facilitated our investigation of heterogeneity through sub-group analyses by time and place.

When drawing conclusions about the applicability of results, it was then important to consider how the known resistance pattern of each partner drug might influence where trials were conducted. It seemed likely to us that the equipoise required to ethically justify a randomized trial could restrict the evaluation of specific ACTs in settings with documented resistance, and we found some evidence of this with artesunate plus sulfadoxine-pyrimethamine. In our analyses this combination performed well against other ACTs but the number of trial settings was severely limited (with widespread resistance to sulfadoxine-pyrimethamine in most of Africa). As a consequence we noted that systematic reviews are unlikely to detect declining efficacy of anti-malarials over time, and alternative methods will be required to justify withdrawal of recommended drugs, such as single-arm efficacy studies, or in-vitro monitoring of known markers of resistance [40].
More positively, knowledge of the known resistance patterns can permit generalization of results to additional settings beyond those of the included trials. For example, in this version of the review, there were no African studies evaluating artesunate-mefloquine (due to historical concerns about mefloquine causing excessive vomiting in children). However, as mefloquine resistance has not been documented in Africa it is highly likely that this combination will be effective even without local trials.

Following publication of the review, and the subsequent recommendation in favour of dihydroartemisinin-piperaquine, the European Medicines Agency (EMA) raised concerns about potential cardio-toxicity with this combination [41]. Our review had found few signals of major safety concerns, but this was primarily due to inadequate safety monitoring in the underlying trials, and consequently insufficient to be reassuring to those making decisions. None of the included trials had conducted ECG monitoring of participants, but two randomized controlled trials (RCTs) published after the Cochrane review had noted an increased risk of QT prolongation in those treated with dihydroartemisinin-piperaquine.

Given the high efficacy of the recommended ACTs, the adverse event profile became a critical consideration for national policy makers when selecting first-line treatments. Consequently, for subsequent reviews of individual anti-malarial combinations we developed a new approach to assessing and documenting the adequacy of adverse event monitoring and completeness of reporting, and produced detailed adverse event summary of findings tables, detailing the absolute risk of adverse events and the confidence in those estimates (see figure 6).
Figure 6: Summary of findings table for adverse events with dihydroartemisinin-piperaquine versus artemether-lumefantrine. Adapted from Zani 2014. \[6\]

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Number of participants having adverse events (95% CI)</th>
<th>Number of participants (trials)</th>
<th>Quality of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Serious adverse events</strong> (including deaths)</td>
<td>AL6: 6 per 1000 (6 to 17)</td>
<td>DHA-P: 10 per 1000 (2 to 5)</td>
<td>7022 (8 trials)</td>
</tr>
<tr>
<td>Gastroenterological</td>
<td>Early vomiting: 2 per 100 (2 to 5)</td>
<td>3 per 100 (8 to 11)</td>
<td>2695 (3 trials)</td>
</tr>
<tr>
<td></td>
<td>Vomiting: 9 per 100 (8 to 11)</td>
<td>9 per 100 (1 to 7)</td>
<td>6761 (9 trials)</td>
</tr>
<tr>
<td></td>
<td>Nausea: 2 per 100 (8 to 11)</td>
<td>2 per 100 (1 to 7)</td>
<td>547 (2 trials)</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea: 12 per 100 (10 to 14)</td>
<td>12 per 100 (1 to 7)</td>
<td>4889 (7 trials)</td>
</tr>
<tr>
<td></td>
<td>Abdominal pain: 19 per 100 (12 to 20)</td>
<td>16 per 100 (12 to 17)</td>
<td>911 (5 trials)</td>
</tr>
<tr>
<td></td>
<td>Anorexia: 15 per 100 (12 to 17)</td>
<td>14 per 100 (12 to 17)</td>
<td>3834 (5 trials)</td>
</tr>
<tr>
<td><strong>Neuro-psychiatric</strong></td>
<td>Headache: 27 per 100 (25 to 44)</td>
<td>33 per 100 (25 to 44)</td>
<td>309 (2 trials)</td>
</tr>
<tr>
<td></td>
<td>Sleeplessness: 1 per 100 (1 to 9)</td>
<td>3 per 100 (1 to 9)</td>
<td>547 (2 trials)</td>
</tr>
<tr>
<td></td>
<td>Dizziness: 3 per 100 (2 to 11)</td>
<td>4 per 100 (2 to 11)</td>
<td>547 (2 trials)</td>
</tr>
<tr>
<td></td>
<td>Sleepiness: 0 per 100 (0 to 0)</td>
<td>0 per 100 (0 to 0)</td>
<td>384 (1 trial)</td>
</tr>
<tr>
<td></td>
<td>Weakness: 17 per 100 (15 to 21)</td>
<td>18 per 100 (15 to 21)</td>
<td>1812 (5 trials)</td>
</tr>
<tr>
<td><strong>Cardio-respiratory</strong></td>
<td>Cough: 42 per 100 (40 to 45)</td>
<td>42 per 100 (40 to 45)</td>
<td>4342 (5 trials)</td>
</tr>
<tr>
<td></td>
<td>Coryza: 68 per 100 (60 to 72)</td>
<td>66 per 100 (60 to 72)</td>
<td>832 (2 trials)</td>
</tr>
<tr>
<td></td>
<td>Prolonged QT interval (Bazett's correction): 7 per 100 (6 to 11)</td>
<td>9 per 100 (6 to 11)</td>
<td>1548 (1 trial)</td>
</tr>
<tr>
<td></td>
<td>Prolonged QT interval (Fridericia's correction): 0 per 100 (0 to 2)</td>
<td>0 per 100 (0 to 2)</td>
<td>1548 (1 trial)</td>
</tr>
<tr>
<td><strong>Musculoskeletal/dermatological</strong></td>
<td>Pruritus: 2 per 100 (2 to 6)</td>
<td>4 per 100 (2 to 6)</td>
<td>2033 (5 trials)</td>
</tr>
<tr>
<td></td>
<td>Facial oedema: 0 per 100 (0 to 0)</td>
<td>0 per 100 (0 to 0)</td>
<td>384 (1 trial)</td>
</tr>
</tbody>
</table>

The assumed risk of adverse events in the AL6 group is an average risk across trials. The corresponding risk with DHA-P (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).
CI: Confidence interval; RR: Risk ratio.

1. No serious risk of bias: All but one of the trials are open label. However, we did not down grade for this outcome.
2. No serious inconsistency: The finding is consistent across all trials. Statistical heterogeneity is low.
3. No serious indirectness: Trials were mainly conducted in children in Africa, with few trials in Asia or in adults.
4. Downgraded by 1 for serious imprecision: No statistically significant difference was detected between treatments. However the current sample size does not exclude the possibility of rare but clinically important differences.
5. Downgraded by 1 for risk of bias: The majority of trials are open label.
6. No serious imprecision: The finding is of no effect and the CIs around the absolute effect excludes clinically important differences.
7. Downgraded by 1 for serious imprecision: There are limited data.
8. Downgraded by 1 for serious imprecision: The result does not reach statistical significance.
9. No serious imprecision: The total number of participants is high and findings are precise.
10. Downgraded by 1 for serious risk of bias: This trial is unblinded. Only a few of the recorded prolonged QT intervals were registered as adverse events which removed the statistical significance. The reasons for this are unclear.
11. No serious indirectness: This single trial was conducted in children in Uganda, Kenya, Mozambique, Zambia, and Burkina Faso.

Influence on WHO recommendations

The WHO formally considered the first draft of the review in 2008. This was the first time the WHO malaria department had utilized the GRADE approach to formulating recommendations, and a GRADE sub-group was formed to learn the methodology and formally go through the evaluations of ‘quality of evidence’. Following discussions, the committee added dihydroartemisinin-piperaquine to the list of recommended ACTs, and withdrew the recommendation for amodiaquine plus sulfadoxine-pyrimethamine.

This process established a transparent and systematic approach to the evaluation and consideration of new anti-malarial combinations, and for the third edition in 2015 we utilized the same protocol and several new review teams to develop three further Cochrane reviews in response to current questions:
• ‘Dihydroartemisinin-piperaquine for treating uncomplicated *P. falciparum* malaria’. Following the EMA report, the WHO committee formally considered the evidence for cardio-toxicity with dihydroartemisinin-piperaquine, and concluded with a strong recommendation in favour of dihydroartemisinin-piperaquine (*Strong recommendation, high quality evidence*).

• ‘Artesunate plus pyronaridine for treating uncomplicated *P. falciparum* malaria’. Prior to the WHO meeting, some expected this new combination would be added to the list of recommended drugs. However, the review highlighted the lack of data in children, and the concerns about liver toxicity. The committee therefore did not recommend artesunate-pyronaridine for general use, but made allowance for its use in settings with resistance to other combinations (*Conditional recommendation, moderate quality evidence*).

• ‘Artemisinin-naphthoquine for treating uncomplicated *P. falciparum* malaria’. Although there was some pressure from the pharmacological company for the WHO to consider this combination, the review highlighted that very little data was available and provided justification for the panel not to recommend it (*Strong recommendation, very low quality evidence*).

A variation of this review was also used to assess the use of artemisinin-based combination therapies in the treatment of *Plasmodium vivax* malaria, and this was also presented and considered by the panel.
Chapter 3. Which first-line treatment for severe malaria?

Main paper


Additional contributing work

Asare B, Sinclair D. Evidence summary for artesunate [10]


Background

Many deaths from severe malaria occur during the first 24 to 48 hours following hospital admission and, consequently, to be effective anti-malarial drugs need to achieve rapid therapeutic blood concentrations following administration [42].

The artemisinin derivatives have been shown to clear malaria parasites from the blood faster than other anti-malarials [43][44], and importantly they are also more effective against the young ring forms of the parasite, which sequester in the microcirculation of vital organs causing severe disease [42]. Artesunate is the most studied artemisinin derivative, and was first recommended in the second edition WHO Malaria Treatment Guidelines.

Figure 7: Evidence box in support of the WHO recommendation to use artesunate in preference to quinine in adults with severe malaria. Taken from the WHO Malaria Treatment Guidelines: Second edition. [30]

At that time an earlier version of the Cochrane review was used as the basis for GRADE evidence profiles presented to the guideline panel [45]. After consideration
of the evidence the WHO recommended artesunate for use in adults, but concluded there was insufficient evidence to make recommendations about its use in children (see Figure 7).

In 2010 we conducted an update of the review, to include newly published trials in children, and in 2012 we added an appraisal and commentary on cost-effectiveness studies.

**Policy question**

Does parenteral artesunate reduce deaths from malaria compared to parenteral quinine in children?

**Methods**

We used standard Cochrane methods to formally search for, appraise, and synthesize randomized controlled trials comparing intravenous or intramuscular artesunate with intravenous or intramuscular quinine, in adults and children with severe malaria. The primary outcome was death, and secondary outcomes were: neurological sequelae, coma recovery time, time to hospital discharge, fever clearance time, parasite clearance time, and adverse effects.

This was the first systematic review by the Cochrane Infectious Diseases Group to use the GRADE approach to evaluate the quality of evidence. We used the following considerations to assist our judgments:

- Risk of bias: we conducted sensitivity analyses to assess the robustness of the results against the risk of bias criteria.
- Consistency: we assessed the consistency in the size and direction of the individual effect estimates, and the overlap of the confidence intervals of
individual studies. We further assessed this statistically by applying the Chi\textsuperscript{2} test and the \(\text{i}^2\) statistic. To explore possible causes of heterogeneity we conducted the sub-group analyses described below.

- Directness: we evaluated the directness and applicability of the results by conducting sub-group analyses by: participant age (children versus adults), type of severe malaria (cerebral versus non-cerebral malaria), geographical region, drug regimen (loading dose versus no loading dose of quinine, and use of any additional anti-malarials), route of administration (intravenous versus intramuscular route), and time since admission to hospital.

- Precision: we considered the clinical importance of the upper and lower limits of the 95\% confidence interval, and the power of the individual studies and overall meta-analysis to detect effects.

- Publication bias: we used funnel plots to consider whether we had any evidence that publication bias was adversely influencing the result.

We presented the effect estimates for the main outcomes alongside the GRADE judgments in summary of findings tables.

**Key findings**

We found eight trials enrolling 1664 adults and 5765 children. Six trials were conducted in Asia and two in Africa. Of these, two were large multicentre studies: Dondorp 2005 [46], 11 centres in four countries (Bangladesh, Myanmar, India, and Indonesia); and Dondorp 2010 [47], 11 centres in nine countries (Mozambique, The Gambia, Ghana, Kenya, Tanzania, Nigeria, Uganda, Rwanda, and the Democratic Republic of the Congo).
Dondorp 2005 and Dondorp 2010 used rapid diagnostic tests (RDTs) to confirm *P. falciparum* parasitaemia, and all the other trials used standard microscopy. Although standardized clinical definitions for severe malaria exist, the entry criteria were not consistent across trials.

All trials compared artesunate with quinine, but only three administered both artesunate and quinine using the current recommended dosing schedules (artesunate: 2.4 mg/kg on admission, at 12 hours, at 24 hours, and then once daily until starting oral therapy; quinine: 20 mg/kg loading dose, then 10 mg/kg every 8 hours until starting oral therapy).

Overall, treatment with artesunate significantly reduced the risk of death both in adults (RR 0.61, 95% CI 0.50 to 0.75; 1664 participants, five trials; *high quality evidence*) and children (RR 0.76, 95% CI 0.65 to 0.90; 5765 participants, four trials; *high quality evidence*) (see Figure 8).

In children, treatment with artesunate was associated with an increase in the incidence of neurological sequelae at the time of hospital discharge (RR 1.41, 95% CI 1.05 to 1.88; 6422 participants, three trials).

However, one trial followed participants up until day 28 to see if these sequelae resolved. Of the 170 children with sequelae at the time of discharge, 129 (75.9%) were available for assessment on day 28, and 68 of these (52.7%) had fully recovered. At this time point the difference between groups was not statistically significant (RR 1.23, 95% CI 0.74 to 2.03; 4857 participants, one trial).
Figure 8: Forest plot of randomized trials comparing artesunate with quinine for the primary outcome: death. Taken from Sinclair 2012. [2]

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Artesunate</th>
<th>Quinine</th>
<th>Risk Ratio M.H. Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.1.1 Adults (Age &gt; 15/16 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anh 1998</td>
<td>2 16</td>
<td>7 22</td>
<td>1.3%</td>
</tr>
<tr>
<td>Anh 1995</td>
<td>6 59</td>
<td>18 91</td>
<td>3.6%</td>
</tr>
<tr>
<td>Dondorp 2005</td>
<td>102 633</td>
<td>153 678</td>
<td>29.9%</td>
</tr>
<tr>
<td>Hien 1992</td>
<td>6 31</td>
<td>8 30</td>
<td>1.0%</td>
</tr>
<tr>
<td>Newton 2003</td>
<td>7 50</td>
<td>12 54</td>
<td>2.4%</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td>841</td>
<td>823</td>
<td>38.8%</td>
</tr>
<tr>
<td><strong>Total events</strong></td>
<td>1244</td>
<td>1298</td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong>: Chi² = 2.23, df = 4 ( (P = 0.89) ), I² = 0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test for overall effect</strong> Z = 4.70 ( (P &lt; 0.00001) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **1.1.2 Children (Age < 15 years)** | | | |
| Cao 1997 | 4 37 | 5 35 | 1.0% | 0.76 [0.22, 2.66] |
| Dondorp 2005 | 5 67 | 11 105 | 21.1% | 0.49 [0.10, 1.67] |
| Dondorp 2010 | 230 2712 | 297 3713 | 67.7% | 0.77 [0.46, 0.91] |
| Elnashir 2010 | 1 33 | 2 33 | 0.4% | 0.50 [0.05, 5.26] |
| **Subtotal (95% CI)** | 2879 | 2880 | 61.2% | 0.76 [0.65, 0.90] |
| **Total events** | 240 | 315 | | |
| **Heterogeneity**: Chi² = 0.67, df = 3 \( (P = 0.83) \), I² = 0% |
| **Test for overall effect** Z = 3.31 \( (P < 0.00001) \) |
| **Total (95% CI)** | 3720 | 3709 | 100.0% | 0.71 [0.62, 0.80] |
| **Total events** | 364 | 513 | | |
| **Heterogeneity**: Chi² = 5.60, df = 8 \( (P = 0.59) \), I² = 0% |
| **Test for overall effect** Z = 5.45 \( (P < 0.00001) \) |
| **Test for subgroups differences**: Chi² = 2.69, df = 1 \( (P = 0.10) \), I² = 62.8% |

**Strengths and limitations of this approach**

This review provided clear evidence that artesunate was superior to quinine at reducing deaths, and in the conclusions we went as far as stating that further trials comparing these two anti-malarials in the general population were unnecessary [2].

The result appeared applicable to all populations and settings, and consequently the decision about whether to change a national first-line policy from quinine to artemether shifted to other considerations.

In a minor update in 2012, we added a brief appraisal of the cost-effectiveness of artesunate versus quinine. Following a formal search we identified two cost-effectiveness studies conducted alongside the two large multicentre efficacy trials [48][49]. Based on a number needed to treat of 41 (taken from the large multicentre trial in children), switching to artesunate was estimated to cost $123 per additional life saved; a highly cost-effective intervention by any standard [50].
However, later that year during a collaborative project with the Medicines for Children Project, the WHO Essential Medicines Department, and the Ghana National Drugs Programme, we encountered unexpected problems when reconsidering the evidence for the Ghanaian national context [5].

Firstly, it was noted that artemether (an alternative artemisinin derivative to artemisunate but with less favourable pharmacokinetics) had entered common use in Ghana due to its widespread availability and ease of use. Neither the Cochrane review nor the WHO guideline had commented on the comparative effects of artemisunate and artemether. To appropriately consider artemisunate as a first-line treatment in Ghana, the national committee therefore tasked a small team to utilize evidence-based approaches to appraise and summarize older non-Cochrane systematic reviews of artemether versus artemisunate and artemether versus quinine [11].

Secondly, and perhaps more importantly, there were major problems with the feasibility of rapid policy change. There remained a worldwide shortage of artemisunate, and the committee was rightly concerned about the potential for programmatic harms with a change from a drug with a reliable supply chain (quinine) to one with an unreliable supply and a need to quality-assure the product (due to the widespread problem of fake anti-malarials in Ghana).

The artemisunate review, the WHO recommendation, and the subsequent consideration by the Ghana National Drugs Programme demonstrated the importance of national debate before implementing WHO recommendations (even with strong recommendations based on high quality evidence). In this instance,
guideline development at a national level influenced subsequent recommendations at the global level.

**Influence on WHO recommendations**

The updated review was presented to the WHO panel in March 2011, and after consideration artesunate was recommended as first line treatment for all age groups worldwide *(Strong recommendation, high quality evidence)*.

Subsequent to this, and in response to our work in Ghana, I supported and advised a new review team to conduct a Cochrane review of artemether for treating severe malaria [10]. This review was presented to the WHO panel in 2014, and artemether was recommended in preference to quinine when artesunate was unavailable *(Conditional recommendation, low-quality evidence)*.
Chapter 4. Which diagnostic approach for malaria?

Main paper


Additional contributing work


Background

The gold standard for malaria diagnosis is confirmation of the presence of malaria parasites in a symptomatic person’s blood by light microscopy [51]. However, most malaria episodes and deaths occur in rural parts of Africa where light microscopy is difficult to implement and maintain. Consequently, prior to the development and availability of Rapid Diagnostic Tests (RDTs) the WHO recommended a pragmatic approach labelled ‘presumptive treatment for malaria’, whereby all children with fever were given anti-malarials regardless of the presence or absence of alternative causes for their fever [52][53].

This presumptive approach accepted the substantial overuse of anti-malarials because the available mono-therapies were cheap and well tolerated. However, the increased cost associated with artemisinin-based combination therapies and the declining malaria prevalence in many settings raised the relative importance of diagnosing and treating other causes of fever [54][55].

From a health systems perspective RDTs have the potential to substantially reduce the over-prescription of anti-malarial drugs by reducing the misclassification of fevers, especially in low-prevalence areas [56][57]. Consequently, in 2010, the WHO recommended a switch from presumptive treatment to parasitological diagnosis for all [30], and promoted this new strategy as a way to improve the care of people with malaria [58].

However, a simple logic model demonstrates that people with malaria are only likely to be harmed by the introduction of RDTs, as a small number will be
misclassified as not having malaria (false negatives) and the appropriate anti-malarial will be withheld or delayed (see figure 9). Instead, the potential health benefits of introducing RDTs are restricted to people whose fever is not due to malaria, for whom a negative RDT result should prompt the health worker to look for and treat the true cause of their fever.

We therefore undertook a review of RDT-assisted diagnosis versus presumptive treatment to establish the relative benefits and harms of introducing RDTs.

**Policy question**

Does the introduction of Rapid Diagnostic Tests (RDTs) into algorithms for diagnosing and treating people with fever improve health outcomes, and reduce anti-malarial prescribing?
Methods

We followed standard Cochrane methods to search for, appraise, and synthesize individual and cluster randomized trials comparing RDT-supported algorithms with algorithms using clinical diagnosis alone.

The primary outcomes were: the proportion of patients still unwell at follow-up, and the proportion of patients with fever prescribed anti-malarials. Secondary outcomes were: the proportion of patients with fever prescribed antibiotics, the proportion of microscopy-positive patients not prescribed anti-malarials, and the proportion of microscopy-negative patients prescribed anti-malarials.

In reviews of complex interventions such as this, clear descriptions of the intervention are essential to understanding the applicability, and feasibility, of achieving the same results outside of a trial setting. In this review we presented this information in additional tables (see figure 10), and these were used to guide our investigation of heterogeneity, and assessments of directness and applicability. The tables summarize details of the epidemiological setting, the training provided to health staff, and the protocols used to guide treatment for those testing positive and negative.

We used the GRADE approach to evaluate the quality of evidence, and presented the effect estimates for the main outcomes alongside the GRADE judgments in summary of findings tables.
**Key findings**

We included seven trials, enrolling 17505 people with fever or reported history of fever; two individually randomized trials and five cluster-randomized trials. All trials were conducted in rural African settings.

In most trials the health workers diagnosing and treating malaria were nurses or clinical officers with less than one week of training in RDT-supported diagnosis. Health worker prescribing adherence to RDT results was highly variable: the number of participants with a negative RDT result who received anti-malarials ranged from 0% to 81%.

Overall, RDT-supported diagnosis had little or no effect on the number of participants remaining unwell at four to seven days after treatment (6990 participants, five trials; *low quality evidence*); but using RDTs reduced prescribing of anti-malarials by up to three-quarters (17287 participants, seven trials; *moderate quality evidence*). As would be expected, the reduction in anti-malarial prescriptions was highest where health workers’ adherence to the RDT result was high, and where the true prevalence of malaria was lower (see figure 10).

Using RDTs to support diagnosis did not have a consistent effect on the prescription of antibiotics, with some trials showing higher antibiotic prescribing and some showing lower prescribing in the RDT group (13573 participants, five trials; *very low quality evidence*).

One trial reported malaria microscopy on all enrolled patients in an area of moderate endemicity, so we could compare the number of patients in the RDT and clinical diagnosis groups that actually had microscopy-confirmed malaria infection
but did not receive anti-malarials. No difference was detected between the two diagnostic strategies (1280 participants, one trial; *low quality evidence*).
Figure 10: Table of characteristics of interventions in trials comparing RDT-supported diagnosis with clinical diagnosis alone. Modified from Odaga 2014. [3]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Nurse/assistant</td>
<td>Nurse</td>
<td>Community health workers</td>
<td>Drs, nurses, and clinical officers</td>
</tr>
<tr>
<td>Who conducted the training?</td>
<td>Nurses, after a training course</td>
<td>Not described</td>
<td>Experienced IMCI trainers</td>
<td>Nurses after a two-week course</td>
</tr>
<tr>
<td>How long was the training? (days)</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>Half-day</td>
</tr>
<tr>
<td>Was a written guideline provided?</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Yes</td>
</tr>
<tr>
<td>What supervision was conducted?</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Record review and feedback (monthly)</td>
<td>Observation and feedback (once, 2-months after training)</td>
</tr>
<tr>
<td>Were staff incentives provided?</td>
<td>No</td>
<td>No</td>
<td>Bicycles</td>
<td>No</td>
</tr>
<tr>
<td>Who conducted the RDT tests?</td>
<td>Research staff</td>
<td>Research staff</td>
<td>Prescribers</td>
<td>Prescribers</td>
</tr>
<tr>
<td>Which RDT-type was used?</td>
<td>OptiMAL-IT (pLDH)</td>
<td>Paracheck (HRP-2)</td>
<td>ICT malaria PF (HRP-2)</td>
<td>Paracheck (HRP-2)</td>
</tr>
<tr>
<td>Were the RDTs provided free?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Were the anti-malarials provided free?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Were the antibiotics provided free?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Algorithm</td>
<td>Test all cases of fever with RDTs?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Prescribe only if RDT positive?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Do not prescribe if RDT is negative?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Management of RDT-negative cases?</td>
<td>Not described</td>
<td>Look for other causes and treat as per STG</td>
<td>Amoxicillin if signs of pneumonia; else refer</td>
<td>Not described</td>
</tr>
<tr>
<td>Guidance on prescribing antibiotics?</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>If pneumonia is suspected</td>
</tr>
</tbody>
</table>
Strengths and limitations of this approach

The GRADE approach to appraising diagnostic tests remains in development and has not yet been formally adopted by the Cochrane Collaboration [59]. However, the approach emphasizes the importance of questioning whether the introduction of a new diagnostic test actually improves patient-important outcomes, and the most direct way to evaluate this is through a randomized comparison of two diagnostic strategies. That was the approach taken in this systematic review.

The alternative is to model clinical outcomes using the sensitivity and specificity of the test, provided by a systematic review of diagnostic test accuracy [60], and evidence-based predictions of what will happen to those testing false positive, true positive, true negative and false negative; i.e. who will benefit and who will be harmed. We presented this approach in the background to the review alongside the conceptual framework (figure 9).
Our systematic review demonstrated clearly that even with fairly minimal training, the introduction of RDTs substantially reduces unnecessary prescriptions of anti-malarials (see figure 11); an important consideration from a health system perspective. However, the review provided only limited reassurance to health workers and patients that this did not result in a substantial increase in adverse health outcomes. The included trials missed the opportunity to demonstrate health benefits by concentrating follow-up care on those testing positive for malaria, rather than those testing negative (the only group who are likely to benefit or be harmed). While all trials gave a clear description of the subsequent care of people testing positive, very little detail was provided on what would happen to those who tested negative; there was little or no description of when a test would be repeated, of what protocols would be utilized to look for alternative causes, or of when and how antibiotics would be used.

In our follow-up paper in the British Medical Journal, we highlighted and discussed some of these key questions, and it seems likely that the evidence to reassure health staff and patients will not come from randomized trials [12]. In an ideal world, local audits would provide feedback to health workers on the proportion of false negative results in their setting (by conducting studies of parallel RDT and microscopy testing), and on the subsequent clinical course of those discharged home without anti-malarials. However, in reality these audits will probably be conducted as observational cohort studies due to the lack of routine data collection in these settings, and we found one published example of this from Tanzania [61]. Researchers tracked 1000 children for two weeks after attending health services with a fever, and, reassuringly, patients with negative RDT results seemed to re-
present to health services if they failed to get better, and there were no recorded instances of malaria-related deaths in people discharged home without anti-malarials. This result is probably only generalizable to settings with similar access to healthcare services (geographically and financially), and further similar studies would help guide future recommendations.

Influence on WHO recommendations

The results of the review were presented to the WHO MTG committee in October 2014. However, to date the committee has been reluctant to make specific recommendations around the use of diagnostic tests, and instead simply makes a good practice statement that ‘the diagnosis should be confirmed parasitologically prior to treatment’ [30].

In future editions of the guideline, evidence-based approaches could be used to make statements about:

- When to do a rapid diagnostic test – the clinical criteria that should initiate testing in specific epidemiological settings.
- When to repeat a test – the clinical criteria for re-testing.
- When to treat presumptively - the epidemic threshold for reverting to presumptive treatment due to a shift in the balance of benefits and harms (and cost-utility) of testing.
- Interventions to improve adherence to test results.
Chapter 5. Facilitating national guideline development

Main papers


Additional contributing work

Background

An internal evaluation of WHO guidelines in 2003 was never published but is reported to have found that ‘most WHO guidelines did not meet most of the AGREE criteria’ [62]. The assumption was that although the organization had internal guidance for the development of recommendations, it was rarely followed [63][64]. To investigate this, Oxman and colleagues conducted in-depth interviews with senior staff about the processes they used to formulate recommendations and published the results in the Lancet in 2007. While there was variation between departments, most described a process of expert consensus without any formal collation, appraisal, or consideration of evidence [62].

In response to this public criticism, the WHO moved quickly to strengthen guideline processes within the institution [65]. A guideline review committee (GRC) was established and tasked with overseeing the implementation of evidence-based processes, and the GRADE approach to formulating recommendations was formally adopted [66].

In 2011, in response to a request from the WHO GRC, we undertook a re-evaluation of WHO guidelines and guideline processes to look for and document any changes following the introduction of stricter quality control and the implementation of the GRADE approach. Also in 2011, we conducted a separate pilot project in Ghana, which provided insights into the applicability and usability of global recommendations.
Policy questions

Has the quality of WHO guidelines improved since the formation of the GRC, and what are the experiences and perceptions of WHO departments about the implementation of GRADE?

Methods

For the evaluation of WHO guidelines, we utilized a matched before-and-after study design to compare 10 guidelines published in 2010 (from a spread of WHO departments and across a range of broad topic areas) with 10 guidelines published before the GRC (earlier versions of the same guideline where possible). We used the AGREEII appraisal tool to assess the methods and presentation of each guideline [22]. We then interviewed senior staff from 18 WHO departments about their approaches to formulating recommendations, and their experiences with the GRC and the application of the GRADE approach.

In Ghana, working with the WHO Essential Medicines Department, we identified five paediatric medicines that were strongly recommended by the WHO for low-income countries such as Ghana, but were not yet adopted onto the Ghana essential medicines list. We then trained local Ministry of Health staff to retrieve, appraise, and interpret systematic reviews, and prepare evidence summaries utilizing the structure developed by the SUPporting POlicy relevant Reviews and Trials (SUPPORT) project [67]. These summaries included appraisals of the benefits and harms, the potential public health impact, costs, and feasibility of introducing each drug in Ghana, and were used to facilitate informed decision-making by the national expert committee in November 2011.
**Key findings**

We found marked improvements in GRC-approved guidelines, with average scores higher across all six AGREE domains compared with pre-GRC guidelines. The biggest improvements were noted for ‘rigor of development’ (up 37.6%, from 30.7% to 68.3%) and ‘editorial independence’ (up 52.7%, from 20.9% to 73.6%): both important markers of quality and reliability.

‘Stakeholder involvement’ and ‘applicability’ were now the lowest-scoring domains in GRC-approved guidelines, and these are probably the most difficult to do well at a global level. Indeed, many of the interviewees acknowledged that making recommendations that were intended to inform policies in many different settings was not straightforward. There was uncertainty about how to incorporate considerations about disease burden, health infrastructure and financing, and cultural values and preferences into WHO documents, or how to provide the necessary contextual guidance within the framework of existing methods.

This finding was further borne out in Ghana, where applying global recommendations was not straightforward for any of the medicines, regardless of the presence or absence of high-quality evidence of important clinical benefits.

Our experience across the five topics identified four factors that generated debate and uncertainty in the committee:

1) The available trials were primarily conducted in non-African settings, with plausible reason to doubt the effect could be generalized (for example, zinc sulphate for acute diarrhoea in children);
2) The trials used control groups which did not reflect current practice in Ghana (for example, trials of artesunate compared it with quinine but in Ghana artemether was widely used);

3) Evidence on cost and cost-effectiveness was unavailable (for zinc sulphate, chlorhexidine, caffeine citrate, and dispersible amoxicillin); and

4) Reliable suppliers were either limited or not immediately available (all five medicines).

**Strengths and limitations of this approach**

For the past decade, much of the focus on WHO guideline development has been around improving the transparency and reliability of the methods [62][63][65]. Our appraisal of WHO guidelines demonstrated unequivocally that improvements had been made. Even without any tests for statistical significance, the improvements were large and obvious, and subsequent WHO guideline appraisals have confirmed this [68][69].

Moving forward, these two papers highlighted the difficulty and importance of addressing applicability in global guidance (and systematic reviews). The Ghana project in particular demonstrated that a statement that a recommendation was based on high quality evidence at a global level was not enough to facilitate confident decision-making at the national level. The Ghana committee rightly wanted to know how the trial populations and settings (and control groups) related to their own, and this has implications for the future presentation of both systematic reviews and guidelines.
As an example of this, the Ghana national committee considered the WHO recommendation for zinc supplementation during acute diarrhoea in low- and middle-income countries [70]. The Cochrane review available at the time concluded that: ‘In areas where diarrhoea is an important cause of child mortality, research evidence shows zinc is clearly of benefit in children aged six months or more’ [71]. However, the analysis sub-grouped by continent presented in figure 12 demonstrates that almost all of the evidence for an effect on diarrhoea duration is from Asian studies, and the only African study to date failed to demonstrate an effect. This was problematic for the committee in Ghana, especially as there were good reasons to suspect that the effects of zinc vary by location dependent on the local diet and risk of zinc deficiency; the coastal populations in Ghana eat a lot of fish and are therefore unlikely to be zinc-deficient [72].

Traditionally, sub-group analyses are presented in Cochrane reviews where they have been used to explore and explain statistical heterogeneity. However, they have become a key part of the GRADE assessment of ‘directness’ or applicability, and facilitate interpretation of the review regardless of the presence or absence of statistical heterogeneity.
For global guideline developers this implies a need to pre-specify which factors might influence the effectiveness of an intervention and reduce applicability, and to formally consider each of these alongside each recommendation; explicitly stating where an effect has been proven, where it remains uncertain, and the plausible reasons why it may be more or less effective in these scenarios. Indeed, these steps are explicitly recommended by the GRADE working group [73][74]. For infectious diseases, background resistance would obviously be important (and time and place might be used as surrogate markers), but many other factors such as age, disease prevalence, or nutritional status might also be important.
**Influence on WHO recommendations**

The results of the WHO guideline evaluation were presented to a cross-departmental meeting of the WHO in Geneva in November 2012. Informal feedback from the GRC credits the paper with helping to strengthen the commitment of the organization to the improved standards, and reinforcing the authority of the GRC.

The findings of the Ghana paper were presented at a WHO meeting of the Essential Medicines for Children Project (November 2011), and at the Cochrane Colloquium in Hyderabad (September 2014). This paper highlights the importance of national debate of the evidence-base and implications of policy change, prior to adopting and implementing global recommendations, and fits within a broader discussion of the importance of documenting the steps between evidence and decisions.

The GRADE working group (through an extension project known as DECIDE: Developing and Evaluating Communication strategies to support Informed Decisions and practice based on Evidence), has now developed worksheets for use by guideline panels, to formally document the evidence and debate around each of the different factors influencing a decision: size of the problem, benefits and harms of the intervention, costs, feasibility, acceptability etc. [75].

In the third edition of the WHO Malaria Treatment Guidelines we included a simplified version of this worksheet (which captured some of the important factors influencing the decisions) but the panel did not openly discuss each of the considerations in a step-wise fashion as advocated by the GRADE group [14].

Although the current textbook format of the Malaria Treatment Guidelines seems popular with end users (it is a bestseller among WHO documents), future editions
might better serve the information needs of malaria programme managers if they utilized this worksheet approach to clearly document the rationale, evidence, and debate around each major recommendation.
Chapter 6. Conclusions

This PhD presents a case study of an academic group working as info-mediaries at the interface between research and global policy, and at the interface between global policy and national decisions: advising on methodological issues, conducting systematic evidence reviews in response to information needs, and developing approaches to reinterpret reviews for national decision-making.

It contains one of the very few formal evaluations of policy making within the WHO, and the first published demonstration of evidence-based, rigorous, processes being used to re-assess global recommendations for a specific low-income setting. The included systematic reviews facilitated new, transparent, global recommendations for dihydroartemisinin-piperaquine as a first line therapy for uncomplicated malaria, and intravenous artesunate as first-line treatment for severe malaria.

The scientific contribution of this work

This PhD contributed extensively, particularly in the field of malaria, to improved presentation and interpretation of systematic review results, and subsequently to improved transparency in global and national level recommendations. The included systematic reviews helped to establish the credibility of meta-analysis in the malaria field, and were among the first to include innovative elements such as: GRADE summary of findings tables, standardized language reflecting the level of certainty in effect estimates, logic frameworks, and brief economic summaries. My experiences contributed to the broader discussions on these topics and to further development of their methods [77][78][79].
The first edition of the WHO Malaria Treatment Guidelines recommended five drug combinations but stated only that these ‘had achieved the target of 95% treatment success’. There was little guidance on how to choose between these five combinations, and little comment on any potential adverse effects. Furthermore there were clear inconsistencies in the reporting of trial results across malaria research groups, making direct comparison between trials impossible.

My early work, summarizing trial data for artemisinin-based combinations, built on existing protocols to develop a standard set of treatment outcomes (including adverse effects), and a standardized approach to trial analysis that fully considered the potential effects of missing data on the primary efficacy outcomes. These were essential for the credibility of meta-analysis across anti-malarial drug trials, and are now standard for Cochrane reviews of artemisinin-based combination therapies.

To facilitate informed national drug selection, we also needed to provide clear summaries of the comparative safety of the main ACTs. We therefore developed a system for assessing the completeness of adverse event monitoring and reporting, and for the third edition of the malaria treatment guidelines I produced detailed adverse event GRADE profiles for each ACT. These adverse event profiles go beyond any current guidance of either Cochrane or the GRADE working group, and are an important avenue for further work. Typically health staff and patients require two types of information on adverse effects; the frequency and nature of common short term ‘side-effects’: for which there is now a reasonable amount of data from RCTs of ACTs; and the risk of rare but serious adverse effects: for which RCTs are unlikely to be sufficient. In a forthcoming review of mefloquine to prevent
malaria in travellers we have further developed these methods to incorporate large amounts of data from non-randomized studies to try and capture data on rare effects [80].

My work in Ghana highlighted the importance of being explicit about the applicability of systematic review results if they are to be truly useful to decision-makers in different settings. This is particularly important in infectious diseases where resistance (and consequently efficacy) vary with time and place, and geographical presentations of data were essential to generating confidence among the Ghana national guideline panel. Cochrane methods at the time only recommended sub-group analysis as part of the assessment of heterogeneity, and applicability was usually only considered informally in the discussion. To improve this, and to facilitate decision making at global and national levels, I incorporated the time and place of the primary trials into all forest plots, and conducted sub-group analysis against a pre-defined set of applicability criteria regardless of the presence or absence of heterogeneity. My approach was developed independently of the GRADE working group, but has subsequently been recommended as part of the assessment of ‘directness’ [81].

Prior to the development of the GRADE approach to assessing the quality of evidence, it was commonplace for the conclusions of Cochrane reviews to be based solely on the statistical significance of the primary outcomes, and early users of the GRADE approach often saw it simply as an additional step at the end of a review. However, the real value of the GRADE approach is its ability to pull together all the elements of a Cochrane review, into a reliable conclusion that fully considers the
risk of bias of the trials, the characteristics and applicability of the trials, and the statistical certainty in the results. As such, applying the GRADE approach is an integral part of structuring the analysis, writing the results, and drawing conclusions. This PhD includes the first two reviews by the Cochrane Infectious Diseases Group to contain GRADE summary of findings tables, and these reviews were instrumental in establishing these processes within the Cochrane Infectious Diseases group, for pushing their adoption wider in the collaboration, and for establishing the GRADE approach within the WHO Malaria Treatment Guidelines.

Alongside the adoption of GRADE, these reviews also pioneered the use of standardized wording of results, especially within the abstract and plain language summaries [82]. This wording was developed by researchers from the Cochrane EPOC group, and aims to reflect the size of the observed effects AND the overall confidence that this effect is real (the GRADE quality of the evidence). This innovation, combined with the formal applicability assessment, and GRADE assessment of the quality of evidence, combined to facilitate transparent and reliable statements about the results of systematic reviews. Subsequent WHO recommendations based on these reviews have also incorporated language reflecting both the estimated magnitude of effects and the certainty in these effects.

Logic frameworks are a more recent addition to systematic reviews, which require authors to think critically about the assumptions underlying a research question. They can be used to clarify the proposed mechanism of action of an intervention, and distinguish the important impact outcomes from the intermediate process
outcomes. They can highlight deficiencies in the outcomes being measured and reported in the existing trials, and contribute to improved understanding. The logic framework presented within this PhD helped to clarify where the potential benefits and harms of introducing RDTs lay (in those testing negative), and in what epidemiological settings the benefits and risks would be greatest. These insights had not been clearly articulated in any policy documents at the time. As reviews begin to tackle more complex questions, logic frameworks are likely to become increasingly common and play an important role in developing the question and framing the interpretation. During the course of this PhD, I have worked with multiple authorship teams to develop logic frameworks for their own topic, often leading to radical restructuring of the analyses and reframing of the review conclusions. These reviews have helped to establish logic frameworks within the Cochrane Infectious Diseases Group, and to promote their use wider in the collaboration.

Cost and cost-effectiveness are usually not included within Cochrane systematic reviews, and this is probably appropriate as cost only becomes relevant once the efficacy and safety of an intervention have been established. Cost considerations are also rarely (or only briefly) included in WHO guidelines (probably due to rapid changes and variability in pricing). The evidence for the clinical superiority of artemisinin in severe malaria was however so overwhelming that cost (and supply) became the primary concern for national policy-makers following the WHO recommendation. For this reason I supervised the inclusion of a brief summary of economic data in the most recent update of this review (which equates to a very
rapid second systematic review of cost-effectiveness data). This remains one of the few examples of this methodology, but a valuable addition to this review.

The developing role for academics at the interface between research and policy

For many years, ‘getting research into policy and practice’ was an elusive and frustrating problem for the research community, but over the last 20 years evidence synthesis has grown and developed to fill the research-to-policy gap. Subsequently, the development of the GRADE approach has provided a useable structure for converting systematic review findings into transparent, informed decisions.

To facilitate and strengthen the use of these methods, academic input is now required at all stages of the guideline process, particularly where panels are heavily weighted with clinicians, or lack sufficient experience in research synthesis. It is recognized that such groups may struggle to formulate clearly answerable questions, and identify where and how evidence synthesis may help [66][76].

Most guideline development manuals now recommend a ‘methodologist’ as part of a guideline panel, but very little has been written about the required competencies. The WHO guideline for guidelines state simply that the methodologist should be ‘an expert in guideline development processes’ [66], and writing in the Lancet prominent members of the GRADE working group wrote that methodologists should have ‘advanced training (usually Masters or PhD) in clinical epidemiology, and extensive experience in the interpretation, and usually the generation, of new knowledge from clinical research’ [83].
The GRADE and DECIDE methodologies, though logically structured, are not straightforward and their useful application requires experience, and usually innovation to adapt them to the specific circumstance. To achieve this, guideline methodologists usually require not only the obvious experience in evidence synthesis and the GRADE approach, but also a diverse group of skills such as group facilitation, consensus building, and plain language writing.

The systematic reviews presented within this PhD were primarily developed to contribute to the development of the WHO Malaria Treatment Guidelines, but are equally relevant to national policy decisions. Consequently, a core requirement for national policy groups, if unnecessary duplication is to be avoided, will be the ability to reinterpret or adapt systematic reviews developed outside of their own guideline process. As in my work in Ghana, this may involve an appraisal of the systematic review methods, re-analysis and structuring of the systematic review data to match the local question, and where necessary a rapid update of the review.

*Our experience as info-mediaries*

When the guideline process works well, evidence synthesis can facilitate discussion among the guideline panel, increase confidence in decisions, and improve the reliability and clarity of recommendations. Indeed this PhD presents some good examples of high quality research, synthesized in high quality reviews, leading to unchallenged and clear decisions. The underlying principles of both GRADE and DECIDE are simple and logical, and easily taught to inexperienced panels.

Inevitably however, the process is rarely this straightforward, even with more
experienced panels. Without adequate input from those making the decisions, systematic reviews can easily ask the wrong question, or the right question in the wrong way. Similarly, without sufficient content and methodological expertise, systematic reviews can simply fail to produce useable information rendering them useless and frustrating to the panel.

More commonly though, systematic reviews are criticized for taking too long to produce, or failing to reach a definitive conclusion due to the low quality evidence available (and therefore being a waste of time compared to expert opinion). High quality systematic reviews take time to produce (often up to two years unless there are authors with large amounts of dedicated time), and real and imagined time pressures (or industry pressures) can cause experts to push for practice change long before benefits or harms have been reliably demonstrated in clinical trials.

Within the world of evidence synthesis, there has been much talk about the need for ‘rapid reviews’ to ensure that evidence synthesis meets policy windows. In our experience however, guideline development is rarely an emergency, and important questions can usually be predicted in advance, through frequent dialogue with policy makers, and regular literature searches. Over three editions of the malaria treatment guidelines, the long-term engagement between the Cochrane Infectious Diseases Group and the WHO Malaria Treatment Guidelines committee (and independent funding to complete policy-relevant malaria reviews), facilitated timely evidence synthesis in anticipation of, and response to, emerging questions.

While more linkages such as this, between academic groups and guideline groups,
should be encouraged, where possible the evidence synthesis itself should probably remain independent of the committee. Throughout our involvement with the WHO Treatment Guidelines Committee we sought input for the questions and the outcomes of interest, and on occasion invited panel members to peer referee draft reviews. However, we usually resisted involvement of panel members on review teams as this can compromise their freedom to challenge prevailing opinions in light of the evidence, and conversely the freedom of the committee to challenge the conclusions and interpretation of the systematic review.

Systematic reviews are often perceived as rigid or mechanical, with defined methods leading to a set outcome. In reality though, while the core methods are now well established, and the GRADE approach has added a structure for formulating conclusions from results, there remains a great deal of freedom in both the presentation and interpretation of the summarized evidence. It is therefore helpful if methodologists have some oversight of the multiple systematic reviews (from multiple academic teams), required for a single guideline. For both the Malaria Treatment Guidelines, and the Ghana evidence summaries, my oversight enabled standardization of outcomes and approaches across reviews, which in turn facilitated understanding and interpretation by the guideline panels. Furthermore, it is not unusual for guideline groups to ask more focused questions after seeing the synthesized evidence, and it is useful to have access to the datasets to respond quickly to such questions from the panels.

There are also frequently tensions between mechanistic logic and demonstrable evidence, particularly when the quality of the evidence summarized by a systematic
review is perceived as low quality. Evidence synthesis may be discounted for providing a weak answer that pulls against established logic, or a weak answer when a strong message is wanted. Indeed it is not unusual for guideline groups to express a desire for strong clear messages, whatever the evidence, for fear of being ignored, or uncertainty that those enacting the recommendations will have the ability to interpret the evidence appropriately.

Our experience in Ghana however, was the opposite of this, with the panel adding a layer of interpretation to both high and low quality evidence which may be easily missed at a global level. With some training and facilitation, the panel (dominated by clinicians) quickly understood the magnitude and certainty in research findings in favor of artemunate for severe malaria but made insightful judgments about potential difficulties with implementation and the undesirable consequences of this. Similarly, despite the obvious limitations in both the applicability and feasibility of the evidence for chlorhexidine, they could see the problems with current practice and the benefits of further local research. Rather than weakening global recommendations, these insights fed back into, and improved, global guidance.

It is important to note however, that these insights into national conditions in Ghana were usually informal, based simply on the experiences of the panel, rather than reliable data. Although we routinely searched for national data, even when it existed, the panel had little confidence in its validity, at times choosing to trust the results of studies conducted elsewhere rather than their own routine data collection. The composition of the panel is therefore instrumental in gaining diverse opinions, and the DECIDE frameworks (or SUPPORT tools we utilized) at least
provide a transparent and structured approach for considering the important factors.

**Preparing global guidelines for contextualization at national level**

Moving forward, systematic reviews and global guidance will only act to facilitate autonomous, informed decision-making if their format and content is directly targeted at serving the information needs of national decision-makers.

When considering the evidence of effects of an intervention, there is a need for both systematic reviewers and guideline groups to pre-specify the key factors that might plausibly influence the effects in a given population or setting; to adequately explore this effect; and to clearly present the findings and implications [27][28][77]. The GRADE criteria of consistency and directness provide a framework for considering these factors.

Similarly, for complex interventions there is the need to clearly describe all facets of the interventions with the potential to influence the outcome, and consider the likelihood that real-life (pragmatic) interventions will really achieve these aims. As the GRADE approach has made clear, there are many factors that influence decisions beyond evidence of effects. The GRADE/DECIDE evidence-to-decisions frameworks go some way to increasing the transparency of these additional steps, by presenting both the evidence and debate around issues such as cost and feasibility.

The development and adoption of GRADE summary of findings tables, and DECIDE evidence-to-decision frameworks, has radically changed the format and content of
WHO guidelines. Moving forward it would be invaluable to garner feedback from the target-users of guidelines on the usefulness of this information, the clarity of their presentation, and the influence on subsequent national decisions.


10. Asare B, **Sinclair D**. Evidence summary for artesunate. Supplementary material to [5].

11. Buabin D, Mensah NO, **Sinclair D**. Evidence summary for artemether. Supplementary material to [5].


References

28. Eisenberg JM. Globalize the evidence, localize the decision: evidence-based medicine and international diversity. Health Affairs. 2002;21(3):166-8
References


82. Cochrane Effective Practice and Organisation of Care (EPOC). Reporting the effects of an intervention in EPOC reviews. EPOC Resources for review authors, 2017. Available at: http://epoc.cochrane.org/resources/epoc-resources-review-authors

83. Hirsh J, Guyatt G. Clinical experts or methodologists to write clinical guidelines? Lancet 2009; 374: 273–75 (Authors reply)
