Author's response to reviews

Title: Pre-referral rectal artesunate in severe malaria: a fundamentally flawed trial

Authors:

Karim F Hirji (kfhirji@aol.com)
Zulfiqarali G Premji (zulpremji688@gmail.com)

Version: 2 Date: 26 October 2010

Author's response to reviews: see over
Dear Editors-in-Chief of *Trials*:

We are submitting the revised version of the paper "Pre-referral rectal artesunate in severe malaria: a fundamentally flawed trial" by Karim F Hirji and Zulfiqarali G Premji for continued consideration for publication in your journal.

First we extend our sincere thanks to the two reviewers for their detailed comments on the first version of our paper. These comments have been very helpful in the production of this hopefully improved version.

The key issue which concerned the reviewers and the editor was the issue of ethics. We have realized that our case was not well documented on that matter. All discussions of ethics and commercial interests have thus been removed, and now we focus on the scientific aspects of the trial, avoiding strong language as suggested. The relevant topics previously discussed under the heading of “Ethics” are now discussed in relation to external validity, and with toned down language.

The reviewers note the practical difficulties of conducting clinical trials in rural Africa and Asia. These are, no doubt, considerable. But over past two decades, a large number of trials have been carried out there. Much has been learned and some high quality trials have been done. In particular, we refer to the trial of Yeboah-Antwi et al. (2010) *PLoS Medicine*, 7: e1000340 that successfully trained village health workers to differentiate between malaria and pneumonia, and use a rapid diagnostic test for malaria. The intervention reduced treatment failure rates for both conditions. Their report is exemplary in terms of the details given about eligibility criteria, sample size calculation, pre-specified primary and secondary outcomes, background of village health workers used, their training and other issues. The data analysis and presentation are well done. On the journal website, the trial protocol, CONSORT checklist, and two training manuals are available to them, there is no reason why Gomes et al. could not have attained even a higher standard of quality in design, conduct, analysis and reporting. A table in the revised version (Table 2) makes a comparison of these two trials.

The comments of the reviewers made us add material to the clarify issues. New material about Gomes et al. from two other papers in the literature written by these authors is also noted.

This version incorporates most of the specific suggestions given by the reviewers. These are detailed in a point-by-point response to their comments attached below. On some issues, the two reviewers give different verdicts. Thus, the absence of stratified analysis is deemed by Dr. Wittes as “serious violation that affects inference” while Dr. Cook says that “a solid argument can be made for treating the data as from a...
single study...” In such cases, we found ourselves to be mostly in agreement with Dr. Wittes.

The changes made in response to the comments of the reviewers appear in red color and the new material appears in blue color. Changes in style or arrangement of the contents are not marked. Kindly note that with the LaTeX template provided by your journal, we have not been able to produce a line numbered version.

We thank you, Dr. Cook and Dr. Wittes for your thoughtful comments and kind assistance,

Sincerely,

Karim F Hirji, Professor of Medical Statistics
Zulfiqarali G Premji, Professor of Parasitology
Muhimbili University of Health and Allied Sciences,
Dar es Salaam, Tanzania
kfhirji@aol.com, zulpremji688@gmail.com
Authors’ Responses to Review Reports

As requested, a detailed paragraph by paragraph response to the comments of the reviewers is given below. Our responses appear in red color.

**********

Response to Reviewer I

Reviewer’s report
Title: Pre-referral rectal artesunate in severe malaria: a fundamentally flawed trial
Version: 1 Date: 16 July 2010
Reviewer: Jonathan Alistair Cook
Reviewer’s report:

The manuscript is an extensive critique of a recently published large RCT of treatment for Malaria of immediate treatment in a remote by drug suppository before sending to health facility where diagnosis can be confirmed or treatment options may be available. The critique (as the Hirgi et al paper will be referred to in this review) covers a large number of aspects of the trials and highly where alternative options were available and also a number of flaws in aspect of the approach (particularly analysis).

*** Authors’ Response: None. ***

The authors of the critique conclude that this is a fundamentally flawed study whose conclusions remain subject to appreciable doubt. Whereas the Lancet comment as noted by the (critique) authors characterise as one of handful of important paper every decade it will influence the way malaria is treated. The critique covers a number of aspects of trial design, conduct and analysis which seeks to present the justification for such a conclusion. To some extent the criticism reflect an explanatory or idealistic approach to evaluation whereas the Gomes trial was implicitly at least conducted in a very pragmatic fashion.

*** Authors’ Response: We are not sure what is meant by an idealist or explanatory approach. In our view, we have used basic criteria drawn from the literature for making our critique. ***

However, there are a number of aspects (particularly related to the analysis) where the Gomes study was clearly conducted in a suboptimal way. Given the number of important issues this critique raise it has a lot of value and I would support its publication. Trialists may well differ (as I do) with some of the comments.

*** Authors’ Response: We agree that trialists differ, and with reason, on the importance of different features of a clinical trial. We make a note of this in the conclusion. ***
For me the paper and the critique together highlight three aspects:

1. The difficult of reporting in the space allowed in paper publications. A number of the criticism amount to we dont know and maybe this could have influenced. The critique strong highlights the value and need for further detail on the conduct of the trial though the original paper is still very important. Furth information is given on WHO website though further would be extremely helpful.

*** Authors’ Response: The Lancet allows extra material to be put on its web-pages but that was not utilized. The WHO website provides little in the way of details. In this era of electronic publishing and greater awareness about transparency, lack of space is not, in our view, a valid reason for the absence of critical information. We make a note of this in the conclusion.

Information about sample size calculation and eligibility criteria would have taken much less space compared to the extensive space devoted to subsidiary secondary analyses. ***

2. The difficult of conducting any trial in this area of research. The authors of the critique comment that a positive feature was that only 8 out of the almost 18000 subjects were completely loss-to-follow-up. For me this is more than positive more likely an exceptional undertaking (though I note the 7-30 days timeframe reflects upon this) in extremely difficult circumstances. Furthermore the authors note it is the largest trial conducted for severe malaria and only trial of this intervention in these circumstances. This reflects the practical challenges.

*** Authors’ Response: On issue of difficulty of conducting trials, please see the cover letter to the editor.

We thank you for connecting the 7-30 days time-frame with the issue of level of loss to follow up. The seemingly high degree of follow up can be actually an artifact of how they set up the time window and analyzed the data.

Say the planned follow up window was fixed at day 25-30 after randomization. But in the study the status of half the cases was known only to day 14, and for many, only to day 7. Under a binary endpoint analysis using the planned follow up window, the loss to follow up would be more than 50%. The usual way to address this problem is through a survival type of analysis. An erroneous way to deal with it is to artificially stretch the follow up window to 7-30 days and use the binary endpoint analysis. Gomes et al. did the latter. In the light of your comment, we have modified the discussion of this issue. ***

3. That different people will conduct a trial in different ways and neither are right or wrong that there is substantial grey area. The critique tends to state that cer-
tain aspects are right or wrong whereas some are very difficult issues with no clear consensus on what is appropriate.

*** Authors’ Response: While a nuanced stand on some issues is needed, we do not think everything is purely subjective. There are issues on which a general consensus exists in the methodology literature. We note this point in the Discussion. ***

Specific comments are given below

**Major Compulsory Revisions**

**Commercial Influence**

The section of Commercial Influence is highly speculative and doesn’t provide any evidence to support its accusations. There is as far as I can see no evidence of any company involvement in the study. Nor is there any evidence that the decision to exclude negative smears was taken to match the pharmaceutical companies interests. Given the strongly litigious nature of medical practice in the USA its hard to see why a pharmaceutical should or would want to have licence to give a drug to treat individuals who only may be suspected of having a disease some of which may turn out not to have it. In the USA it would be reasonable to expect the smear results to be available prior to deciding upon treatment and hence treatment limited to those diagnosed with the condition. To criticism the study on this basis is unwarranted without providing some concrete justification.

*** Authors’ Response: Now we have changed the title and contents, and reframed the issue in terms of external validity. ***

**Minor Essential Revisions**

**ITT**

A large proportion (31%) of the randomised participants were treated as post-randomisation exclusion due to being ineligible (either already being treated or having a negative blood smear and potentially not having the condition). The randomisation and then exclusion of the proportion of whom had a previous antimalerial infection is surprising and is an implicit recognition that they should not have been randomised and that this should have been an exclusion criteria (it was known as it was collected on the randomisation form). The other reason for exclusion is a more understandable and a trickier proposition as here we have participants who appear to be eligible but later it turns out are not eligible (they are diagnosed as not having Malaria). An acknowledgement of the difficulty (impossibility at present?) of identifying the desired population for randomisation is needed. How would the critique authors overcome this challenge?
Authors’ Response: A treatment for severe malaria in a rural setting has to be presumptive treatment and will include some cases who do not have malaria. The question is to reduce their number. We now refer to the paper of Yeboah-Antwi et al. (2010) *PLoS Medicine*, 7: e1000340 which provides practical guidance on this sort of questions.

As noted the formal definition of an intention to treat (ITT) analysis assumes a value for every randomised case though the terms has been used more generally for those which analyse in the groups they were allocated irrespective of compliance. However, if former is applied strictly then ITT is an unattainable standard in many circumstances as missing data will be present and therefore the value is unknown. Analyses with imputation methods (which is what the ITT analyse in this paper are) are not a complete solution to this problem. I feel the critique does not acknowledge the difficult of the issue of missing data for analysis. The inclusion of the post-randomisation exclusions increase p-values as the relative effect size is smaller but this is the heart of the issue as this included people we would not want to be treating and so we would expect a smaller effect size. However, they were randomised.

Authors’ Response: The issue of missing data is not easy to resolve. In ITT analysis, subjects with missing response data are usually analyzed under the best or worst case scenario, as appropriate. But the violations of ITT in Gomes et al. do not stem from the problem of missing data. The follow up status of the excluded cases (one third of the total) was as well known as the follow up status of those who were retained in the analysis. They were excluded for other reasons.

A related concern regarding ITT is the exclusion of such a large proportion of randomisation participant could undermine the comparability of the groups and possibly the applicability of the results. This reflects the inability to randomised only those who were eligible as the smear could not be carried out. Some data is given in Figure 1 which states that those excluded were similar in number by group and outcome across groups. Nevertheless it would have been preferably for an analysis including these participants to also have been conducted and the results presented as done in the critique (and also information on the characteristics of the excluded given).

Authors’ Response: We agree. The distribution of follow up times for the excluded and included cases, and center-specific variations are of interest. In the setting of this trial, smears cannot be done. Treatment is presumptive in nature, and so the “eligible” must be defined as those presumed to have severe malaria and all such cases included in the analysis.

Interpretation either way (if the results differ which they do in terms of significance at 5% but not greatly in terms of effect size) is difficult. The exclusion reflects the tension picked up by the authors when they consider the description of the study population.
For presumably practical reasons, the study randomised suspected severe malaria cases but want to make recommendations on patients with severe malaria. Unless some strategy could be proposed in future for immediate diagnosis then it would have been better to accept the trial is primarily providing evidence about treating suspected cases (which could lead to negative impact on those who do not actually have the disease) and retain though who are later found not to have the disease and treated accordingly. While the critique authors suggest using a diagnostic test, these (like smears) are not without their own problems (as they are inaccurate to some degree in terms of ruling in and out the disease). I do have concerns with the application of the study where suspected cases are given treatment. I expect this was commented on by a reviewer of the Gomes paper and hence why data on outcome was included in the consort on outcome for excluded participants. This would worthy of comment. Can the results be applied as presented in the Gomes paper?

*** Authors’ Response: Gomes et al. trial concerns pre-referral treatment based on presumptive diagnosis of severe malaria in rural Africa and Asia. On the issue of rapid diagnostic tests, Yeboah-Antwi et al. (2010) PLoS Medicine, 7: e1000340 successfully trained VHWs to differentiate between malaria and pneumonia and use a rapid diagnostic test for malaria. We now refer to this trial.

As analyzed, the results of Gomes et al. do not apply to the rural situations. That is what we state. ***

The sensitivity analysis presented in Table 2 though analysis seem to be based upon three categories and so difference in p-values between the Gomes reported analysis are not surprising. It is not clear from the critique and it should be explicitly stated whether the three level status is used (presumably it is).

*** Authors’ Response: We do not directly compare the p-values from our three category analysis with those from the binary analyses of Gomes et al. We say that when there are several outcome categories, it is better to begin with an overall analysis and then, if indicated, make separate binary comparisons. The latter are adjusted for multiplicity. ***

It is disappointing that the Gomes paper only carried out significance tests bar for some subgroups. The impact of using a different effect size, the risk difference, is given in table 3 and by centre and overall (from a random effects model). In this case the random effects result is similar to RD of overall results ignoring centre.

*** Authors’ Response: We agree. Note the authors use RR not RD. ***

As a minor point, the critique authors should be more explicit about what they mean by random effects model.
Centre differences

The study took place in three countries and covered a variety of settings. As noted by the critique authors and reported in the original paper there were a number of important differences between the centres. I think there is still a solid argument can be made for treating the data as from a single study though I would support using (or at least having a pre-specified) secondary analysis which stratified by centre to explore the impact of these differences. It would appear the trial authors were to some degree aware of this problem and hence reported the outcomes by location (Africa versus Asia). The inclusion of older aged individual in the Asian site is unusual and could presumably have been standardised across sites. Perhaps another oversight in design or for some reason not feasible? Other differences seem more a reflection of practical challenges (e.g. transportation and free hospitalisation and supportive care). Interpretation of the reason for any difference between sites is therefore difficult and the critique should acknowledge this.

Health centers in rural Ghana, Bangladesh, and Tanzania do not have funds to provide transport for or treat severe malaria patients free of charge, especially when a larger number come to their doorstep in the context of a clinical trial. The funds and arrangements for this form a part of the trial budget and plan. For an internationally funded trial whose resources easily exceed the budget of an entire health district in these nations, these are variations written into the protocol. Yes, in any multi-center study, there will be between center variations in terms of patient features, parasite type, quality of care, etc. The point is to reduce controllable differences with a good design and appropriate implementation.

We do not see why excluding older individuals would have been infeasible in Bangladesh. The fact was that in Bangladesh, different inclusion criteria were used; and this is not stated clearly.

Multiple analyses of the same outcome

On a minor point the sentence (page 3 last paragraph) about disentangling true positive findings from false positive findings is poorly stated. There are not true positive and false positive findings to be separated out because of a lack discipline in analysis; there are just findings judged at a specific criteria (e.g. significance level). As more outcomes are analysed the possibility of spurious results amongst them increases if judged by the same error rate (type 1). With regards to expected recruitment as individual sites this is rarely reported (though implicit in the sample size calculation) as they are generally known often to be rough estimates and can be very inaccurate.
Authors’ Response: We have changed the wording. 50% of cases were from Africa and 50% from Asia. These were specific targets, not rough estimates and seem to have been met. The question we pose is, how were these proportions decided? ***

Ethics

The authors raise the issue of standard care and whether it is acceptable for taking novices off the street and giving short unspecified training of varied duration. The author see this a blank and white issue: By putting sever malaria patients in the care of such recruiters this study compromised its ethical standing. For me this is unclear, if there are more qualified individual available then I agree. However, the principle of using the best resource available holds and if short-term training of individuals with limited training is the only possible/realistic first-line option then so be it. It is ethic if the benefits out way the harms (and a study is needed to evaluate this) then would we not want to implement such an intervention? For a parallel it doesnt seem too different from training first-aiders in countries with much more extensive healthcare systems because there is benefit in immediate low level treatment.

Authors’ Response: We agree with use of people with limited training to provide health care in developing nations. What we say is that Gomes et al. should have used village health workers and not people with no or little health background at all. This is what other malaria trials do. When this treatment is translated into practice, it is the village health care workers who will be used. No country in Africa or Asia will officially leave the care of critically ill severe malaria patients to people with no medical background and only a week of training. One has to set a higher standard for a life threatening condition like severe malaria than for skin rash. ***

Additionally the authors raise the issue of the transport to a health center. Again it is easy to criticise but it is realistic to expect this to be available if a national or regional programme were implemented? The example of placebo versus placebo is not particularly helpful as it is a-priori not ethical because there is no hope of one of the treatments being active whereas in the Gomes trial clearly there is.

Authors’ Response: Transport to health center, where it was provided, came from the trial budget. A national program does not just deal with one treatment for one disease but covers many diseases and treatments. Travel vouchers for serious conditions are a realistic and cost-effective possibility. In this trial, center differences on transport reflected decisions of national trial organizers. Some budgeted for it and others did not. This should have been centrally coordinated as time to reach a health facility was deemed a critical predictor of outcome.

On the question of placebo versus placebo, what you say is the point we make: A trial is justified on the basis of expectation of a clinically meaningful treatment difference,
realistic assessment of practical utility, good design and satisfactory implementation. The placebo versus placebo example responds to the justification on the WHO website that seems to say “do not worry about the problems in this trial since everyone who took part benefited in some way.”

Posing a placebo versus placebo comparison to make a point is not unheard of in the literature. For example, Wittes (2009) employed a 2-placebo scenario to illustrate a point about subgroup analysis (Circulation, 119:912-915). We have retained this comparison.

Similarly they criticise that some of the intramuscular injections in the Tanzania were at a risky site. However, this takes place in the clinic and so presumably (and sadly) is a common occurrence in practice.

*** Authors’ Response: Risky injections were given at only one Tanzania center. The subject then was in the trial, to be followed up subsequently. A common guideline for standard of care and training would have reduced this harm. That it was a major problem at one Tanzania site reflects failure of planning and supervision at that site. We now discuss the issue of patient welfare and its relation to external validity.***

They also raise the issue of consent in a more rural circumstance with a population with limited education. While an interesting discussion of the issue it is difficult to see what the relevance as a critique to the Gomes trial is bar a reference to monitors who not speak the local language.

*** Authors’ Response: We have removed this issue from the paper. See the letter to the editor.***

Overall I feel this section is too simplistic in its view and some acknowledgement of the difficulties are needed here which practical challenges are great and feasibility is a huge concern. At some reference to their being difference in opinion is needed.

*** Authors’ Response: We have removed the issue of ethics from the paper. On the question of practical challenges, please see the letter to the editor.***

Discretionary Revisions

Training

The section on training raised interesting questions about the specific of the training given and the value in more extensive reporting on this by the trialist. However, it seem to me unrealistic to expect much detail in a main trial paper when some many other aspect are required to be reported. I dont agree with the authors that it is important to know is if the training, supervision and monitoring of the recruiters
were more rigorous than the norm for severe malaria trials in the community setting. It is useful to know what about training, supervision and monitoring to determine applicability in general. A less intensive programme of these aspects could arguably be a positive feature of an intervention package if the treatment was effective.

*** Authors’ Response: The issue of training has been addressed in part above. Who was trained and how are central to the practical application of the findings of such a trial. On the issue of details of training, please see the paper Yeboah-Antwi et al. (2010) *PLoS Medicine*, 7: e1000340. We refer to this paper now.

Ordinarily, we need to know the specifics on training, monitoring and supervision. The WHO website material on this trials compares, without giving any details, the training in a trial with training usually given to village health workers. Our comment responds to what is said there, and has been retained. ***

**Analysis method**

It is surprising that the results were not analysed (at least in addition to the presented analysis) using standard survival analyses methods and the Kaplan-Meier curve presented given the variation in follow-up. This presumably is a result of the practical challenges to assess outcome at the same time in follow-up. This would presumably have been more of a problem for disability than for death which I would have thought could have been standardised more readily (when analysed on it own though not for the composite). A particular point, I dont agree with the authors that table 4 presents a particularly useful set of analyses (4 levels of status).

*** Authors’ Response: We agree. Because of the practical challenge in assessing all outcomes after the same time interval, the survival approach is indicated here. This point is related to the issue of loss to follow up, which we have responded to above. Please see the paper by Rabbani et al. dealing with a rural trial in Bangladesh, which we refer to now. This study had a very good, explicit follow up schedule and performance.

The analysis in Table 4 illustrates how to deal with a time dependent subgroup in an unbiased way (see Hirji and Fagerland (2009) *Trials*, 10:1). If complete data are available, better ways of handling this exist. But with the available data, we can do no better. ***

**Subgroups**

I agree with the critique authors that the use of time to reach clinic to stratify outcome is a serious concern as this is a factor that is unknowable until after the outcome and potentially confounded with outcome. It wasnt controlled in the trial design but to stratify the analysis by it implies that it was or at least could be.
Related is the division of death to within or after 6 hours. Using death within 6hrs or died with the entire follow-up period is a consistent approach. However to analyse separate those who were died after 6hrs (i.e. excluding those who died before 6hrs) is problematic. The lack of information on pre-specification of subgroup analyses in the trial paper is very disappointing. There is a general looseness of reporting on this aspect in the Gomes paper which makes it unclear what was pre-specified and what was post-hoc.

Sample size and outcome selection

The authors correctly put out the surprising absence of a sample size calculation from the paper. It is difficult to believe this was not done (given the large sample size they have) but if so it is not reported. If this was reported it may well have informed a related aspect of outcome selection and analysis. As pointed out it is not clear which whether there was a single primary outcome or not. It is very surprising this is not documented. The paper is written like the composite was the primary outcome and this might be worthy of comment.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I declare that I have no competing interests.

*** Authors’ Response: Thank you for your detailed comments. ***
Response to Reviewer II

Reviewer’s report
Title: Pre-referral rectal artesunate in severe malaria: a fundamentally flawed trial
Version: 1 Date: 6 August 2010
Reviewer: Janet Wittes

Reviewer’s report:
1. Gomes et al., in a 2010 paper in the Lancet, describe an 18,000 person trial in Africa and Bangladesh studying the effect of rectal artesunate on severe malaria. According to Gomes, children were randomized to an active or placebo suppository and then went to a clinic for standard injectable treatment. The paper showed that those who received the active suppository had lower mortality and lower rates of permanent disability. The accompanying editorial in the Lancet praised the study and commented on the importance of the results. As a consequence of the trial, the WHO has recommended rectal artesunate as an immediate treatment for severe malaria.

*** Authors’ Response: None. ***

2. Hirji and Premji (HP) lambaste the Gomes et al. paper by pointing to a host of serious methodological failures. HP conclude by saying they are not taking a position on the efficacy of rectal artesunate, but that, Our position is that this question is too important to be decided from an inadequately designed, poorly conducted, erroneously analyzed and selectively interpreted study.

*** Authors’ Response: None. ***

3. When I first read this paper I was struck by its angry tone. Surely, I thought, the Gomes paper could not have been as bad as HP asserted. After all, I reasoned, the paper had passed the ordinarily rigorous peer review at the Lancet and had led to a policy change at the WHO. The authors include some illustrious trialists (e.g., Richard Peto and Nicholas White). But when I read the Gomes paper, I found myself agreeing with many of the points HP made.

*** Authors’ Response: None. ***

4. I would, however, recommend that the authors tone down their language. When HP write about the methodologic problems with the paper, strong language is appropriate. But when HP question the ethics of the trial or the commercial influence on the Gomes paper, I am uncomfortable. These sections appear as ad hominem attacks; I strongly recommend sticking with science and methodology and avoiding accusations of questionable ethics.
5. Below I have some comments and some suggestions for changes. HP might consider classifying their criticisms into categories: those features of the Gomes study that clearly violate rigorous principles of clinical trials and have serious implications for inference (e.g., violations of ITT; subgrouping by a post-randomization variable); those that clearly violate principles but probably don’t matter much for inference (e.g., one slide vs. two); and those that appear to violate principles but because the paper did not report on them we cannot know whether the trial was actually conducted more rigorously than the Gomes would suggest (e.g., training of health workers).

*** Authors’ Response: We agree in principle. But for this study, as now explained in the Conclusion, it is not easy to do this classification. Issues that strongly affect external validity are also, in our view, violations of principle. Note that here some things are not what they seem. As we now explain, the use of only one slide introduces bias because it may lead to some artesunate arm getting an incorrect diagnosis at the health center, and receiving inappropriate treatment, and thus affecting their outcome. That effect would not been seen in the placebo cases with one slide. ***

6. Page 2. First line in second column. impacted is not a verb. Change to, had an impact on.

*** Authors’ Response: Done. ***

7. Page 2, second paragraph in the second column. I would delete the first four sentences of this paragraph it just sets the reader on edge. I would leave the rest of that paragraph unchanged but move it to be part of the paragraph above.

*** Authors’ Response: Changed almost as suggested. ***

8. Page 2, the application of the Jadad scale. This was very interesting and confirmed my sense that bad trials can do well on the Jadad scale and conversely (although not stated by HP) that high quality trials can do poorly.

*** Authors’ Response: We agree. ***

9. Page 3, line 17. Change he or she to the child

*** Authors’ Response: Done. ***

10. Page 3, line 22. Both the original Gomes paper and this manuscript describe the location of the trial as Africa and Asia. But both continents are huge and varied. I
think it is more accurate to say, Tanzania, Ghana, and Bangladesh. Then later refer to the first two countries as Africa but use Bangladesh throughout rather than Asia.

*** Authors’ Response: Done. ***

11. Page 3 and beyond: HP make the interesting point that because of major differences between the African sites and the Bangladeshi site, the study is half-way between a multi-center trial and two parallel trials. This raises a general issue about trials: when are the centers in a trial so different from each other that they represent different scientific questions. I would recommend that HP distinguish those features of the different sites that potentially have a major impact on inference (e.g., the age difference because the course of the disease differs in young children and adults) and those that don’t (e.g., one slide vs. two slide. I see having only one slide in Africa leading to a decrease in precision but as adding no bias.)

*** Authors’ Response: We agree. One slide also adds to bias. Please see the text and the response to point 5 above. ***

12. On page 3, second column, the HP criticize the Gomes paper for not following the CONSORT statement. My reaction was, so what? Just as the Jadad scale produces false positives and false negatives, so does CONSORT. Its a helpful guideline, but there are items in it that are very important and items that are not. For example, HP point to the failure of the Gomes paper to give a sample size calculation. The study is what it is—sample size calculations are relevant in the design of a study but they are much less relevant after the study is over. So, again, I would not toss all the violations into one bucket; I would stress the violations that matter in ultimate inference from the trial. (I feel the same way about Good Clinical Practice in my mind, it adds an unnecessary burden on clinical trials.)

*** Authors’ Response: We agree that not all violations should be thrown into one bucket. However, we differentiate the CONSORT Statement from something like the Jadad scale in that the former is based on an ongoing, formal process that brings together experts in the field, and is accepted as a standard for reporting by many health and medical journals. In particular, The Lancet, where Gomes et al. was published, requires that trial reports adhere to the latest version of the CONSORT Statement.

In our view, sample size calculations are important even after the trial is over in that they are indicative of the primary outcome, the effect measure, type of data analysis method to be used and the loss to follow up and level of missing data in the study. This has been explicitly noted now. ***

13. Page 4, first paragraph. HP want detailed entry criteria. This, I believe, is not realistic in the setting of the Gomes trial. Patients are treated quickly and presumptively. (But the presumptive nature of the treatment plays into the most serious criticism of the paper – the analysis of only 2/3 of the randomized patients.)***

*** Authors’ Response: Trials in rural settings cannot have the refined eligibility criteria used in hospital based trials. But they need some criteria, which must be stated clearly. Please see the cover letter to the editor, and the paper Yeboah-Antwi et al. (2010) *PLoS Medicine, 7*: e1000340. The pre-referral severe malaria study of Thera et al. we cite has clear statements of diagnostic and eligibility criteria.

In rural villages, pre-referral treatment for severe malaria, be it in a trial or in practice, and even with rapid diagnostic tests, can only be presumptive in nature. Including the subsequently confirmed cases only in the analysis compromises the external validity of the trial. We clarify this point in the revised version. ***

14. Page 4, last paragraph on the first column and first on the second. I found this criticism important, but I wonder whether Gomes had to exclude this material because of space limitations in the Lancet. I would recommend that HP tone down the criticism it is possible that Gomes had good methods for training but just didn’t report their approach.

*** Authors’ Response: The questions of who was trained and how are central to the practical applications of the findings of such a trial. *The Lancet* allows extra material to be put on its web-pages. The WHO website does not give any specific details of this trial. See the details given in the paper by Yeboah-Antwi et al. (2010) *PLoS Medicine, 7*: e1000340 that we now refer to.

In this era of electronic publishing and greater awareness about transparency, lack of space is not a valid reason for the absence of critical information. The two rural malaria trials we previously cited (Ajayi et al. and Thera et al.) give clear statements about the training given to village health workers and their supervisors.

As suggested, we have toned down the wording in this section. ***

15. Page 4. The section on the 7 to 30 day window. I agree with HP that this is a problem. I would have liked Gomes to have told us how the window was determined for each participant. I worry that this acts like another post-randomization subgroup. I realize that the study was performed in areas difficult to reach, but as HP points
out, the variability of this window may lead to problems in inference.

*** Authors’ Response: We agree. A fixed and narrow time window for follow up is difficult in this setting. Hence, the analysis should not have used a binary response. It should have used survival analysis methods. The mix-up of short term with medium term mortality is also an issue. Please see the paper by Rabbani et al. dealing with a rural trial in Bangladesh, which we refer to now. This study had a very good, explicit follow up schedule and performance. ***

16. Page 4 and beyond: time to clinic. I regard the whole issue of subgrouping by time to clinic as extremely important and I agree with HPs criticism of Gomes handling of the missing data. But I am uncomfortable with HPs characterization of the trial as poorly conducted given the difficulties of conducting a trial in the settings where it was done, I am much more tolerant of lapses in quality control. Very large trials correct for lapses in quality of conduct because their size reduces the variance of the estimates as long as no bias is introduces. And it is the latter that makes the violations of ITT and post-randomization subgroups so serious.

*** Authors’ Response: The settings in which this trial was done, there can be major problems with missing data. However, our main concern is that time to clinic data were routinely collected only at two of the four study centers, and many blood slides were not taken at one center. These point to problems with training and conduct at these centers. Given that these centers had about 50% of the trial subjects, the level of missing data for time to clinic could be as high as 50%. This is an excessive level for a variable that then becomes the center piece of the analysis done. ***

17. Page 6, last paragraph on the first column. When I read this paragraph, I didn’t believe that HP was correct, but when I checked the Gomes paper, I was shocked to realize that in fact Gomes had not analyzed the data in a stratified manner. I regard this criticism in the category of serious violation that affects inference.

*** Authors’ Response: We agree. ***

18. Page 7, second column, line 21. Please don’t give four significant figures for p-values. How about, .18, .11, .027?

*** Authors’ Response: Done. ***

19. Page 8 and the general discussion of exclusions from the primary analysis. I recommend that HD present a more nuanced discussion of the issues. I agree with HD that the treatment will be presumptive and therefore the entire randomized population should be included in the analysis, at least to assess safety. Nonetheless, the two exclusions (not having malaria and previously being treated with injection) are based on baseline variables. I think it is reasonable to ask what the effect of treat-
ment would be in the target population because (a) that is the population studied and (b) if there were a rapid malaria diagnostic and the histories were taken carefully, it would be interesting to know whether the treatment would be effective.

*** Authors’ Response: Not having malaria is a baseline variable whose value, in the setting of the trial, cannot be known at recruitment. Previously being treated with malaria is a baseline variable whose value, in the setting of the trial and given the recruitment form, can and should have been known at recruitment. Failure to exclude the latter at enrollment is a clear failure of implementation. But excluding these randomized subjects from data analysis violates the ITT principle. Gomes et al. could at least have commented on the impact of including and excluding these subjects from analysis. We note this now.

Yeboah-Antwi et al. (2010) *PLoS Medicine*, 7: e1000340 successfully trained village health workers to differentiate between malaria and pneumonia and use a rapid diagnostic test for malaria. We now refer to this trial.

By improving study design and implementation, the number of clearly and potentially ineligible cases can be reduced. Those that remain, however, should be included in the data analysis since that reflects the reality on the ground. ***

20. Page 8 and the 6 hours. Here the issues are much more dicey and I agree with HD. Not only was the 6 hours determined after the data were collected, but it is a post-randomization factor. (e.g., A child who died immediately could never get to the hospital.) It is very hard for me to understand why the Gomes paper did not put more caveats around this analysis (or not do it at all).

*** Authors’ Response: We agree. Analyzing a post-randomization subgroup is the major problem. Incorporating a post randomization event fully into the analysis requires a model with time dependent variables. ***

21. Page 9, Bottom of second column. Here I disagree with HD. Gomes gives risk ratios and HD prefers risk difference. Many of us involved in trials argue that risk ratios are more relevant for assessing whether a treatment works while risk differences are relevant for assessing the magnitude of the population effect. If HD wants to include this argument, they should make it clear that the choice between the two methods is a matter of taste. I have a similar problem with NNT. I know that CONSORT recommends reporting it, but NNT is a function of incidence and risk. Gomes failure to cite NNT is another matter of taste (and one where I side with Gomes).

*** Authors’ Response: People differ, with reason, in their preference for the risk ratio or risk difference. When the event rate is small, we feel that a risk ratio exaggerates the

Gomes et al. use risk ratio for most analyses and risk difference for a few analyses. They also refer to the NNT is an inconsistent way.***

22. Page 10, second full paragraph. I agree with the points here and would leave this paragraph unchanged.

*** Authors’ Response: None ***

23. Page 10, the ethical anomalies. Here I am very uncomfortable. Accusations of unethical behavior are dangerous. I believe that the HD could write this section without loaded words. I would change the topic from Ethical anomalies to Other Problems in the Design and Conduct of the Trial.

*** Authors’ Response: We have removed all discussion of ethics, and changed the material and terminology in this section, and avoided loaded terms. ***

24. I would change this paragraph to read, This trial fulfilled the basic formal ethical requirements of a randomized clinical trial. Ethical approvalparticipants. In the course of the trial, however, several important anomalies occurred. They were (i) enrollment of ineligible case, (ii) low standards of care, (iii) possible lack of quality of informed consent; and (iv) data analysis that may have been influenced by commercial considerations. [but see below HD may want to delete this section.]

*** Authors’ Response: We have completely changed this section in the spirit of what you suggest. ***

25. Page 10, line -6. taking novices off the street is very strong critical language. Please soften. And on page 11, line 4, I would find words other than compromised its ethical standing.

*** Authors’ Response: The wording has been changed. ***

26. Page 11, par 3, last line. Try to recast to modify the ethical concern wording.

*** Authors’ Response: Done. ***

27. Page 11, line -3. Delete the work ethically. The rest of the paragraph still is strong, but it doesn’t impugn the integrity of Gomes et al.

*** Authors’ Response: Done. ***
28. Page 11, the section on Quality of Consent. I found this section fine; I would delete the word on after behooves on line 2 of the first full paragraph.

*** Authors’ Response: Done. The section has been removed. See letter to the editor. ***

29. The section on Commercial Influence (p 12-13). I found this section extremely problematic because it questions the motivation of the Gomes et al. It is possible that Gomes et al. believed their analysis was correct. The criticisms that HD level against the Gomes article are serious enough without questioning their integrity. I would delete this section entirely. (Or modify it considerably, focusing on the difference between US regulatory standards and the need to deal with presumptive therapy in the countries where the trial took place. Interestingly, at the FDA Advisory Committee, there was considerable discussion about the appropriateness of excluding all the children who were excluded.)

*** Authors’ Response: We have reframed the issue as you suggest and taken care not to appear to be questioning the motivations of the authors. Thank you for informing us that the issue of exclusions was given a lot of attention at the FDA Advisory Committee meeting. ***

30. Page 13. First paragraph under Conclusions I would delete the sentence on ethical concerns or rewrite to say, Other serious concerns were detected as well.

*** Authors’ Response: Sentence rewritten in the spirit of what you suggest. ***

31. I found the reanalysis very useful and sobering.

*** Authors’ Response: Thank you. ***

32. In conclusion, this is a very instructive paper. The authors have convinced this reader than the Gomes study is so flawed as to render its results problematic. If HD tone their paper down to be more scientific and not to question the motivation and the ethics of the investigators, this will be a very useful paper.

*** Authors’ Response: Thank you. We have toned down the paper and taken care in the revised version to not convey the impression that we are questioning the motivations of the investigators. ***

33. One addition that HD might consider is to describe why doing such a study is very difficult. Some of the decisions Gomes made may well have been forced by the exigencies of the situation in the field. I am impressed by the magnitude of the effort of Gomes and the team I think the HD paper would be improved if the authors acknowledged some of the challenges this type of trial poses. Giving Gomes credit
for carrying out such a trial would make the paper sound less angry and hostile. It would not detract from the serious criticisms HD has leveled at the paper.

*** Authors’ Response: We address the challenges entailed in conducting a study of the type done by Gomes et al. by a detailed comparison with a large trial done under similar circumstances. ***

Level of interest: An article of outstanding merit and interest in its field
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I declare that I have no competing interests.

*** Authors’ Response: Thank you for your detailed comments. ***