Author's response to reviews

Title: Evaluation of the Widal tube agglutination test for the diagnosis of typhoid fever among children admitted to a rural hospital in Tanzania and a comparison with previous studies

Authors:

Benedikt Ley (bnley@ivi.int)
George Mtove (mtoveq2002@yahoo.co.uk)
Kamala Thriemer (kthriemer@ivi.int)
Ben Amos (bennsal@teule.org.uk)
Lorenz von Seidlein (lseidlein@ivi.int)
Ilse Hendriksen (icehendriksen@hotmail.com)
Abraham Mwambuli (mwambuli1978@yahoo.com)
Aikande Shoo (muhezaddh@kaributanga.com)
Rajabu Malahiyo (mallahiyo@hotmail.com)
Shaali M. Ame (shaaliame@yahoo.com)
Deok R. Kim (dkim@ivi.int)
Leon R. Ochiai (rochiai@ivi.int)
John D. Clemens (jclemens@ivi.int)
Hugh Reyburn (Hugh.Reyburn@lshtm.ac.uk)
Stephen Magesa (smagesa@hotmail.com)
Jacqueline L. Deen (jdeen@ivi.int)
Harald Wilfing (harald.wilfing@univie.ac.at)

Version: 3 Date: 25 May 2010

Author's response to reviews: see over
Miss Hayley Hewitt
The BioMed Central Editorial Team

Dear Miss Hewitt:

Thank you for your email. Below please find a point-by-point response to each of the concerns. Attached please find the revised manuscript with 'tracked changes'.

1. **Please could you also structure your abstract according to the guidelines at:**
   [http://www.biomedcentral.com/bmcinfectdis/ifora/#abstract](http://www.biomedcentral.com/bmcinfectdis/ifora/#abstract)

   We have structured the abstract accordingly.

2. **Authors’ contributions - Please include an Authors' contributions section before the Acknowledgements and Reference list.**

   We have added an Authors' contributions section as follows:

   “BL performed the Widal tests, analysed and compared results and wrote the manuscript; GM was in charge of the implementation and management of the study; KT performed Widal tests, analyzed results and contributed to the manuscript; BA supervised the laboratory where blood cultures were performed and contributed to the manuscript; LvS provided scientific support to study staff and manuscript and was involved in clinical care of participants; IH was involved in clinical care of participants; AM was in charge of data management; Aikande Shoo performed blood culture procedures, RM facilitated activities to make data collection possible, SA provided laboratory support; DRK performed the statistical analysis; RLO provided scientific support to the manuscript; JDC provided scientific support to the manuscript; HR provided major contributions to the manuscript, HW provided scientific support to the manuscript, SM facilitated activities to make data collection possible, JLD provided major scientific support to the manuscript and was involved in clinical care of participants.”

3. **Please also ensure that your revised manuscript conforms to the journal style (http://www.biomedcentral.com/info/ifora/medicine_journals ). It is important that your files are correctly formatted.**

   We have ensured that the manuscript conforms to the journal style.
Responses to referee number 2: Patrick Seed

1. On what basis was a titer \( >1:80 \) chosen as “optimal”? The PPV was higher when a \( 1:160 \) cutoff was applied without significantly compromising the other test performance parameters.

We interpreted the PPV, NPV and specificity in the primary analysis as more-or-less unchanged from the cut-off titers of \( \geq 1:80 \) to \( \geq 1:320 \), whereas the sensitivity was highest at the cut-off titer of \( \geq 1:80 \). We added the following sentence in the discussion (page 4, lines 70 -72):

“The PPV, NPV and specificity in the primary analysis was more-or-less unchanged from the cut-off titers of \( \geq 1:80 \) to \( \geq 1:320 \), whereas the sensitivity was highest at the cut-off titer of \( \geq 1:80 \).”

2. The introduction and discussion pass over a number of other studies where the performance of the Widal test has been examined in Africa and on other continents (Choo et al (1993), Parry et al (1999; only briefly cited in the discussion), Olsen et al (2004)). In Parry et al, the test sensitivity is reported as 92%. In Wilke et al in 2002 (a study from Turkey) a PPV of 76 and NPV 71 are reported. The discussion would be improved by describing differences between the studies and the authors’ comments as to why the test performance may have been different between the reported studies and their own. Many differences may arise from the choice of control group, and although the authors do a good job of showing differences in calculated test performance (PPV/NPV) depending on the choice of control group, a more explicit discussion of the control groups used in other studies should be performed.

a) We changed the Title to: “Evaluation of the Widal tube agglutination test for the diagnosis of typhoid fever among children admitted to a rural hospital in Tanzania and a comparison with previous studies” (please see page 1).

b) We added the following text in the Abstract (please see page 3):

“We compared our main findings with those from previous studies.” (page 3, lines 44 – 45)

c) In the Introduction, we changed the following last line

“We assessed the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the Widal tube agglutination test among Tanzanian children hospitalized with febrile illness.” to

“We assessed the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the Widal tube agglutination test among Tanzanian children hospitalized with febrile illness and compared our results with those from previous studies.” (please see page 4, lines 70 -72)
d) We added the following text to the Methods (please see page 8-9, lines 176-183):

"Literature review

We conducted a literature review to compare our main findings with those from previous studies. We included studies of the Widal test which were identified by direct searches of the MEDLINE database through PubMed. The searches were restricted to publications from 1993 to date. We also conducted supplementary searches of the references in retrieved articles. Abstracts were reviewed and if relevant, the article was included".

e) We added the following text to the Results (please see page 11, lines 233 - 236):

“We found 4 articles from 3 countries [. In this series, the age group included and prevalence of blood-culture confirmed typhoid fever varied considerably. The cut-off titre used ranged from ≥1:20 to ≥1:200 and the resulting sensitivity, specificity, PPV and NPV varied considerably (Table 6).”

f) We added the following text to the Discussion (please see page 14, lines 288 - 299):

“[The previous studies included in our review (Table 6) had not been performed in Africa different cut – off titers were applied, and the resulting sensitivity, specificity, PPV and NPV varied considerably. PPV as well as NPV are dependent on the prevalence of disease within the group of participants; the selection process of study participants has therefore direct influence on the results. The difficulty of choosing the correct control group has been noted earlier [8]. While the gold standard, blood culture, is applied in most studies, the true negatives may be defined as febrile patients with a non-typhi laboratory-confirmed diagnosis as done by Parry et al. and Olsen et al. [8, 19]. Alternatively, some studies use healthy controls. Choo et al [20]. considered all febrile cases with an S.typhi negative blood culture as the control group which is problematic as a number of blood culture-negative results are likely to be false-negative due to the poor sensitivity of the blood culture [15, 16, 17, 17] Furthermore, it is difficult to compare the different test kits, as varying antigens perform differently [21].

g) We added the following References (please see page 17ff):


h) We added a new Table reviewing facility-based studies of invasive salmonellosis (please see page -29)
<table>
<thead>
<tr>
<th>Authors</th>
<th>Date</th>
<th>Country</th>
<th>Sample Size</th>
<th>Age classes included</th>
<th>Prevalence of S. typhi in participants</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Cut Off Titre</th>
<th>Control Group(s)</th>
<th>Gold Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choo et al.</td>
<td>1993</td>
<td>Malaysia</td>
<td>2382</td>
<td>Children</td>
<td>6.1%</td>
<td>89%</td>
<td>89%</td>
<td>&lt;50%</td>
<td>99.2%</td>
<td>O or H ≥1:40</td>
<td>Non-typhoid febrile children admitted to hospital</td>
<td>Blood Culture</td>
</tr>
<tr>
<td>Parry et al.</td>
<td>1999</td>
<td>Vietnam</td>
<td>2000</td>
<td>Children &amp; Adults</td>
<td>30.8%</td>
<td>O: 49%; H: 67%; O or H ≥1:100:88%; O: 97%; H: 96%; O or H ≥1:100:87%; O: 88%; H: 88%; O or H ≥1:100:74%</td>
<td>O: 97%; H: 96%; O or H ≥1:100:87%; O: 88%; H: 88%; O or H ≥1:100:74%</td>
<td>O: ≥1:200 H: ≥1:100; O or H ≥1:100:94%</td>
<td>O: ≥1:200 H: ≥1:100; O or H ≥1:100; O or H ≥1:100:94%</td>
<td>O: ≥1:200 H: ≥1:100; O or H ≥1:100</td>
<td>Lab confirmed malaria, dengue or bacteriaemia</td>
<td>Blood Culture</td>
</tr>
<tr>
<td>Wilke et al.</td>
<td>2002</td>
<td>Turkey</td>
<td>410</td>
<td>≥18 y</td>
<td>13.2%</td>
<td>52% Post 7-10 d: 90%; 88% Post 7-10 d: 90%; 76% Post 7-10 d: 88%</td>
<td>71% Post 7-10 d: 93%</td>
<td>O: ≥1:200 H: ≥1:200</td>
<td>Healthy controls, nontyphoidal febrile patients, blood culture negative febrile cases</td>
<td>Blood Culture, Stool Culture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olsen et al.</td>
<td>2004</td>
<td>Vietnam</td>
<td>80</td>
<td>≥3y</td>
<td>73.8%</td>
<td>64% (field) 61% (lab); 76% (field) 100% (lab); 88% (field) 100% (lab); 43% (field) 48% (lab)</td>
<td>O or H ≥1:100</td>
<td>Lab confirmed bacteriaemia, AFB, dengue, malaria, pos. stool culture, pos. urine culture</td>
<td>Blood Culture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ley et al.</td>
<td>This study</td>
<td>Tanzania</td>
<td>1680</td>
<td>2m. – 14y</td>
<td>1%</td>
<td>75%</td>
<td>98%</td>
<td>26%</td>
<td>100%</td>
<td>H: ≥1:80</td>
<td>Non-typhoid febrile children admitted to hospital</td>
<td>Blood Culture</td>
</tr>
</tbody>
</table>
3. Given the very low prevalence of typhoid fever in the Tanzanian population, reporting and discussion the positive and negative likelihood ratios may prove useful, especially if there are additional clinical parameters that may increase the pre-test probability of disease to which the likelihood ratios may be applied.

Yes, certain clinical parameters may increase the pre-test probability of disease. We added the following in the discussion (Please see page 12, line 252 - 253):

“In a previous paper describing the clinical aspects of the children included in this study [13] older age and long duration of fever were predictive of typhoid fever in this group”.

4. With the very broad distribution of age in the study, it would be useful to present a table that shows the ages of the children falling into Groups A-D. Can the authors apply their test parameters across this whole age range or are certain age ranges underrepresented within some of these groups?

a) In the Results, we changed the following line:

“We compared the highest anti-TH and -TO titer by blood culture-confirmed diagnosis (Table 1).“ to

“We assessed the age distribution and highest anti-TH and -TO titer by blood culture-confirmed diagnosis (Table 1). Children with typhoid fever were significantly older compared to the other groups.“ (Please see page 10, lines 202 - 204)

b) In the Discussion, we changed the following line:

“ It is difficult to choose patients with febrile illness who are blood culture-negative and who definitely do not have typhoid fever.“ to

“ It is difficult to choose patients with febrile illness who are blood culture-negative and who definitely do not have typhoid fever. Furthermore, there were relatively few hospitalised children with no bacteraemia in the same age range as those with typhoid fever. Thus, the control children were significantly younger than the cases.“ (Please see page 13, lines 281 - 233)

c) We added a row in Table 1 as shown below:
### TABLE 1. Number and cumulative frequencies of anti-TH and anti-TO levels, overall and by blood culture isolate

<table>
<thead>
<tr>
<th>Highest titer reached; Number (%)</th>
<th>All (n=1,680)</th>
<th>Group 1: Culture-confirmed typhoid fever (n=16)</th>
<th>Group 2: Non-Typhi serotypes of <em>S. enterica</em> (n=49)</th>
<th>Group 3: Other pathogenic bacteria (n=113)</th>
<th>Group 4: No pathogenic bacteria isolated (n=1,502)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (Range)*</td>
<td>1.83 (14.81)</td>
<td>7.21 (11.96)</td>
<td>1.58 (6.73)</td>
<td>1.43 (11.83)</td>
<td>1.84 (14.81)</td>
</tr>
<tr>
<td>Anti-TH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥1:640</td>
<td>15 (0.9)</td>
<td>3 (18.8)</td>
<td>5 (10.2)</td>
<td>0 (0)</td>
<td>7 (0.5)</td>
</tr>
<tr>
<td>1:320</td>
<td>24 (1.4)</td>
<td>6 (37.5)</td>
<td>6 (12.2)</td>
<td>0 (0)</td>
<td>12 (0.8)</td>
</tr>
<tr>
<td>1:160</td>
<td>36 (2.1)</td>
<td>11 (68.8)</td>
<td>7 (14.3)</td>
<td>0 (0)</td>
<td>18 (1.2)</td>
</tr>
<tr>
<td>1:80</td>
<td>46 (2.7)</td>
<td>12 (75.0)</td>
<td>7 (14.3)</td>
<td>1 (0.9)</td>
<td>26 (1.7)</td>
</tr>
<tr>
<td>1:40</td>
<td>85 (5.1)</td>
<td>12 (75.0)</td>
<td>9 (18.4)</td>
<td>3 (2.7)</td>
<td>61 (4.1)</td>
</tr>
<tr>
<td>No agglutination</td>
<td>1,595 (94.9)</td>
<td>4 (25.0)</td>
<td>40 (81.6)</td>
<td>110 (97.3)</td>
<td>1,441 (95.9)</td>
</tr>
<tr>
<td>Anti-TO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥1:640</td>
<td>6 (0.4)</td>
<td>3 (18.8)</td>
<td>2 (4.1)</td>
<td>0 (0)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td>1:320</td>
<td>18 (1.1)</td>
<td>6 (37.5)</td>
<td>3 (6.1)</td>
<td>0 (0)</td>
<td>9 (0.6)</td>
</tr>
<tr>
<td>1:160</td>
<td>34 (2.0)</td>
<td>10 (62.5)</td>
<td>6 (12.2)</td>
<td>0 (0)</td>
<td>18 (1.2)</td>
</tr>
<tr>
<td>1:80</td>
<td>44 (2.6)</td>
<td>11 (68.8)</td>
<td>7 (14.3)</td>
<td>0 (0)</td>
<td>26 (1.7)</td>
</tr>
<tr>
<td>1:40</td>
<td>95 (5.7)</td>
<td>12 (75.0)</td>
<td>10 (20.4)</td>
<td>3 (2.7)</td>
<td>70 (4.7)</td>
</tr>
<tr>
<td>No agglutination</td>
<td>1,585 (94.3)</td>
<td>4 (25.0)</td>
<td>39 (79.6)</td>
<td>110 (97.3)</td>
<td>1,432 (95.3)</td>
</tr>
</tbody>
</table>

*P values: Group 1 vs 4 <0.001; Group 2 vs 4 = 0.209; Group 3 vs 4 = 0.013

**Minor points:**

5. **Line 48: Are there more current epidemiological data than 2000?**

To the knowledge of the authors no more recent epidemiological data exists on the topic.

6. **Lines 55-57: the Widal test is based on serum-mediated agglutination**

We changed the line:

“The test was first introduced by F. Widal in 1896 [2] and is based on a macroscopically visible agglutination reaction between *S. typhi* somatic lipopolysaccharide O antigens (TO) and flagellar H antigens (TH).”

“We the test was first introduced by F. Widal in 1896 [2] and is based on a macroscopically visible serum – mediated agglutination reaction between *S. typhi* somatic lipopolysaccharide O antigens (TO) and flagellar H antigens (TH).” (Please see page 4, line 61 - 63)

7. **Lines 182-184: “increased” suggests a statistical difference. From a statistical standpoint, the numbers are the same.**
We changed the line:

“The sensitivity of an anti-TH and -TO titer of 1:80 increased from 67% to 80% and 67% to 70%, respectively, with the longer duration of fever prior to admission (both p>0.05).” to

“The sensitivity of an anti-TH and -TO titer of 1:80 increased, however not significantly, from 67% to 80% and 67% to 70%, respectively, with the longer duration of fever prior to admission (both p>0.05).” (Please see page 10, line 222)

8. Lines 184-186: See comment above (3).

We changed the line:

“The PPV of an anti-TH and -TO titer of 1:80 increased from 21% to 30% and 19% to 30% (both p>0.05), respectively, with the longer duration of fever prior to admission.” to

“The PPV of an anti-TH and -TO titer of 1:80 rose from 21% to 30% and 19% to 30%, respectively, with the longer duration of fever prior to admission. But the change was also not statistically significant.” (Please see page 11, line 228 - 229)

9. Lines 83: Did all children have only a single blood culture? Were the groupings uniformly determined only on a single initial culture?

Yes. We changed the line:

“We collected 3 to 5 milliliters (ml) of blood (depending on body weight) from each eligible child for the Widal test and a blood culture.” to

“On admission, we collected 3 to 5 milliliters (ml) of blood (depending on body weight) from each eligible child for the Widal test and a single blood culture.” (Please see page 5-6, line103 + 112)

10. Figure 1: The lines in the flowchart are not properly end joined in several places.

Thank you. We corrected the formatting of the flowchart.

Responses to referee number 2: Timothy Barkham

1. The general preamble suggests the Widal does not have a long and well described past – it does! It would be more helpful to readers if the introduction stated this well worn past and told us why they continue to use the assay and why they wanted to do this work when many laboratories in the ‘west’ stopped using the assay many years ago. The reason is presumably because Typhoid is endemic in their area, unlike in the ‘west’ and that the Widal may be the only available / affordable assay they have access to. Most readers are unfamiliar with the difficult conditions in Tanzania so it
would help to set the scene in a little more detail to help us appreciate their work. The clinically naked presentation of this study makes it a little less interesting. This is largely old news as shown by the references quoted. However, for each new generation it may appear ‘new’, especially as many ‘western’ laboratories no longer offer the Widal test because of its known poor performance – see reference 2 in this manuscript!

Thank you for this very useful suggestion. We changed the following line in the introduction:

“The Widal test is widely used in Africa [3] despite there being very little information on its performance on the continent.” to

“Laboratories in industrialized countries have stopped performing the assay. In Africa the Widal test is still widely used [3] because typhoid fever is perceived to be endemic in the area [3] and the Widal test is the only rapid diagnostic assay that is available and affordable. The Widal test is commonly performed when children and adults present with fever to treatment centres, as few centres have the capacity to perform microbacterial culture [4]. Despite this widespread use, little has been published on its performance in Africa.”

(Please see page 3, line 63 - 69)

2. It would make fascinating reading to understand the problems faced in this setting, which I imagine are considerable, and to understand the value placed on the results by the clinical users. I would like to know more about the utility of the assay and consequence of results in their practice. How do they actually use the test?

The reviewer brings up a very interesting point but we believe this is beyond the scope of this paper. A qualitative research study exploring the perceptions and attitudes of African health care workers on the perceived utility of the test and the consequent results in clinical practice would shed light on this matter.

Major Compulsory Revisions:

3. Is this really the population that would be tested with the Widal in real life or was this only one part of a larger study that recruited all febrile children with the stated parameters to study febrile children in general? If so then it is unrealistic and less valuable. I have the impression that all children recruited were tested with the Widal irrespective of the clinical indication. Would they really expect to test over 100 cases to detect one real positive (and 3 false positives)? The test population is described as ‘febrile children’ so I can’t help wondering whether some could be excluded on clinical grounds which would increase the PPV and reduce the NPV. In the secondary analysis the overall performance increased dramatically as subjects with blood cultures positive for pathogens other than Salmonellae were excluded, going some way towards a more targeted testing algorithm. This may be a false group – it all depends on how the clinicians decide who to test with the Widal in real life in their practice. Do they await blood cultures before doing the Widal? I doubt it, so suspect this improved performance is probably unrealistic. The authors should clarify whether the various
populations tested in the analyses are really representative of the population that the clinicians would test in real life. If not then the data is not applicable to real life. If the authors don’t know the they should at least explain this reservation about the real life applicability of the data.

As stated on page ---, “Children aged 2 months to 14 years ... during study hours from 7am to 7pm, Monday to Sunday...with fever of 3 or more days prior to admission, or fever of less than 3 days but with at least one severity criteria” were included. This was the group to which the results are applicable.

4. **The data suggest to me that the test is unhelpful, so if it is used then I would like to hear more about how and why. The value of a test result may be best realized by asking how the result affects the probability of the disease and its management – the clinical applicability as specified in the STARD criteria.** We are told that the pre test probability is less than 1%. What is the post test probability? If it is still very low then the test should not be used. The low PPV is expected, simply because the pre test probability is <1%. Considering the low PPV, what is their next step after a positive Widal? Do they withhold therapy in cases with negative results and give therapy to all who are positive?

The reviewer brings up a very interesting point but goes beyond the scope of this study. In this study the Widal was performed in a retrospective manner several months after the blood culture result had been obtained. The result therefore did not have any impact on treatment. Therapy was based on clinical picture and blood culture results. Our objective was to determine the sensitivity, specificity, PPV and NPV, which we report in the results.

**Discretionary Revisions:**

5. **Considering the low sensitivity of blood cultures the apparent ‘false positive' Widal results may have been true positives. Did they follow up these ‘false positives’? What was their final diagnosis?**

No, the study design did not allow for this kind of follow up.

**Minor Essential Revisions:**

6. **Line 30 especially in developing countries. The Widal test is widely used in Africa but little information exists about its reliability. Comment: The Widal is old and well known to be unreliable. Perhaps you mean... “…little has been published on its performance in Africa’**.

We changed the following line in the introduction:

“The Widal test is widely used in Africa [3] despite there being very little information on its performance on the continent.” to
“Laboratories in industrialized countries have stopped performing the assay. In Africa the Widal test is still widely used [3] because typhoid fever is perceived to be endemic in the area [3] and the Widal test is the only rapid diagnostic assay that is available and affordable. The Widal test is commonly performed when children and adults present with fever to treatment centres, as few centres have the capacity to perform microbacterial culture [4]. Despite this widespread use, little has been published on its performance in Africa. .”

(Please see page 4, line 63 - 69)

7. Line 115 Definitions and analysis. Fever was defined as stated history or presence of fever of #37.5°C. Comment: I don’t understand ‘Fever was defined as stated history’. Does this mean that if the patient complained of fever then this was accepted as such? Please clarify.

Yes, we accepted the patients/guardians presenting complaint of fever. We included this definition because of the concern that we would exclude children with febrile illnesses because they were afebrile when initially assessed on admission (for example, because of paracetamol intake or severe illness with shock). In our data set, less than 1% of the participants had an auxillary temperature on admission <37.5°C.

Discretionary Revisions:

8. Table 1. Comment The columns are not well aligned. The words ‘Children with’ could be removed from each column to make it simpler.

Thank you. We have corrected the alignment and revised the headings in the table. (Please see page --)

We hope that we have addressed all the points of concern in a satisfactory manner and look forward to hearing from you.

Sincerely yours,

Benedikt Ley