Author's response to reviews

Title: Does the Swedish Interactive Threshold Algorithm (SITA) accurately map visual field loss attributed to Vigabatrin?

Authors:

Miriam L Conway Dr. (miriam.conway.1@city.ac.uk)
Sarah L Hosking Prof. (shosking@vsg.org.au)
Haogang Zhu Dr. (haogangzhu@gmail.com)
Robert P Cubbidge Dr. (r.p.cubbidge@aston.ac.uk)

Version: 5 Date: 17 July 2014

Author's response to reviews: see over
Dear Editor,

Thank you again for considering our paper on “Does the Swedish Interactive Threshold Algorithm (SITA) accurately map visual field loss attributed to Vigabatrin?” We would like to take this opportunity to thank the reviewer for their useful comments.

Yours sincerely,

Dr. Miriam Conway

The authors have addressed most issues raised in the first review, and this version of the manuscript is a definite improvement on the previous submission.

Unfortunately there are still several issues to be addressed (listed below).

1. A concern is the interpretation of statistical significance. Throughout the MS the authors seem to imply that, unless a difference is statistically significant, it does either not exist or is unimportant. This is a dangerous misconception in a small sample such as this. (For example, the observation that SITA-Standard / SITA-Fast have larger variability than Full Threshold seems to be discounted in statements such as this: “The only analysis where SITA was not comparable was when threshold agreement within each algorithm (RMSE) was calculated between Full Threshold and the SITA Fast for the outer ring (p=0.013).”)

Whether one algorithm is “comparable” to another is not a yes/no question that can be answered by statistical significance. The authors should simply state that
the test-retest variability was highest with SF followed by SS and FT, and make some quantitative comparison (for example, “… the RMSE with SS was XX% higher than with FT.”)

We agree with the reviewer and have removed all post hoc analysis and placed more emphasis discussing the findings as a whole rather than concentrating on those which are statistically significant. For example our MS now states.

“Both SITA algorithms had less threshold agreement (group mean RMSE) and larger confidence intervals across all visual field regions when compared against Full Threshold (Figure 2). Within the outer field region the RMSE with SITA Fast was 27% higher than Full Threshold and SITA Standard was 19% higher than Full Threshold (p =0.017). Similar differences were found across all field regions.”

There are other places in the MS where the focus on hypothesis testing is similarly troubling (given the small sample). For example, even though it is abundantly documented that SITA is faster than FT, the authors perform post-hoc testing to come up with a p-value.

We agree with the reviewer and have removed all post-hoc testing and p values associated with test duration.

Minor Essential Revisions
2. “Patients receiving VGB are normally monitored using a combination of peripheral kinetic perimetry…” I doubt that this is true. Provide a reference, or rewrite.

A reference has now been included and the sentence re-written and now states

“Patients receiving VGB are frequently monitored using either kinetic perimetry or automated static perimetry.”

3. “Vigabatrin recipients will have their visual field measured using both SITA strategies and the Full Threshold algorithm.” -> “Visual fields were measured with SITA-Standard, SITA-Fast, and the Full Threshold algorithm.” (Also, this sentence does not belong where it appears in the current MS. – please revise the
introduction to a more logical order).

The order has now been changed so that it is now in the methodology section and states:

“Vigabatrin recipients had their visual field measured using both SITA strategies and the Full Threshold algorithm. The Full Threshold algorithm does not use prior distributions of normal and abnormal visual field behaviour to estimate threshold sensitivity but instead employs a 4-2 dB staircase to estimate each threshold sensitivity. The design of the Full Threshold algorithm ensures that the visual field is not artificially influenced by prior models and therefore provided the gold standard for mapping VGB attributed field loss. At the first visit, all patients underwent a 30-2 visual field examination on both eyes using the Full Threshold algorithm. This visit served to reduce the learning effect observed in perimetry\textsuperscript{[20]} and the results were not used for data analysis. At the second and third visits, each patient underwent perimetry on one randomly assigned eye which remained constant for a given patient according to one of four randomly assigned protocols (Table 2). This unconventional order protocol was designed to induce similar degrees of fatigue within all three algorithms by ensuring that the first and second test sessions were of similar duration.”

4. “At the final visit compared to the Full Threshold algorithm the mean sensitivity of SITA Standard was 1.25 dB higher and 1.51 dB higher for the SITA Fast algorithm….” (**and the following sentences**). These differences are already well documented in the literature, and there is little doubt that they are real. (This is particularly obvious for the test durations.) It is therefore quite immaterial here whether they are “statistically significant” in this sample or not. Therefore the post-hoc analyses are inappropriate & should be omitted.

We agree with the reviewer and have removed all post-hoc testing.

5. The error bars on Fig 2 do not simply indicate “between-subject variability” – they reflect the larger scatter of the RMSE estimates from individual locations. The scatter in the RMSEs is expected to increase with the RMSE.

This statement has been removed and the now reads
“Both SITA algorithms had less threshold agreement (group mean RMSE) and larger confidence intervals across all visual field regions when compared against Full Threshold (Figure 2). Within the outer field region the RMSE with SITA Fast was 27% higher than Full Threshold and SITA Standard was 19% higher than Full Threshold (p =0.017). Similar differences were found across all field regions.”

6. Fig. 4: The defect scores should be compared with scatterplots, not by line plots. It does not make sense to connect with lines consecutive patients who have nothing to do with each other.

We agree with the reviewer and both figures have been changed to scatter plots. Both plots are shown below.

**Figure 4:** showing the sum of total deviation aggregate (Top) and pattern deviation aggregate (bottom) probability levels for every patient
7. What is meant by “shape probability analysis”? Pattern deviation probability?

Discretionary Revisions

We agree with the reviewer and have removed the incorrect phrase. We now state:

“Closer inspection of the “normal” Full Threshold visual field plot reveals that if those locations demonstrating a 2% loss on pattern deviation probability analysis were also included in the analysis then the patient would have been diagnosed with a VGB attributed defect”

8. Fig. 2: Unlike in Figure 4, for this Figure it would actually make sense to plot 3 lines (one for each strategy) on top of each other, to aid comparison.

We agree with the reviewer and this Figure has been changed to 3 lines as suggested. The figure is shown below

**Figure 2**: Root Mean Square Error (dB) and 95% Confidence Intervals as a function of algorithm for the whole field, outer ring, middle ring and inner ring.
9. It would be useful to also show, at least in supplementary material, visual fields of the patients who did not have vigabatrin-associated visual field damage according to this paper’s criteria.

We agree with the reviewer and have now submitted all grey scale plots of those who did not have vigabatrin-associated visual field damage as supplementary material. The grey scale plots are shown in the supplementary material.

Reviewer 1

Revision of this paper has done little to reassure me. I’m not a statistician, but it seems that the authors have jumped into analysis of the ‘results’ before they have considered whether these results are actually suitable for analysis at all.

We would like to thank the reviewer for pointing out that we have omitted important information which would re-assure the reader that the visual field plots were reliable and suitable for analysis. All participants included in this study were seen by the clinicians at Aston University for vision test/s, independent from the research study, to determine whether they had VGB attributed field loss or not. If it was a follow up visit the clinicians would decide whether it had progressed. The examination included visual field testing on both eyes and a proportion of patients also had investigations using electrophysiology. The report from the
vision test was then sent to their neurologist stating whether they had a VGB attributed field defect or not and whether the defect had progressed or not and whether the test was reliable. All patients diagnosed with either VGB-attributed field loss or no field loss therefore had this confirmed in their hospital notes, many of which had this confirmed on several occasions. Those with unreliable or questionable visual fields were not included in the study. In the MS we now state:

“Forty four percent of the patients had a confirmed clinical diagnosis of VGB attributed field loss from their medical records. The diagnosis was made independently from the research study for purely clinical purposes. Visual field defects exhibited a bilateral symmetrical defect showing concentric constriction of the peripheral visual field which was more pronounced nasally and typically characteristic of VGB attributed field loss. The other 66% of the sample had a confirmed clinical diagnosis of no visual field loss from their medical records.”

As mentioned in my previous review, one of the commonest reasons for abnormal visual field plots is learning/inattention artefact, and there are many patients who are simply unable to provide a reliable plot using the HFA. In clinical practice, many patients are found to provide ‘reproducibly unreliable’ visual field plots, and these plots can look very like VAVFL (indeed, when we first reported VAVFL there was initial scepticism regarding the veracity of our HFA plots, as many thought that it was simply learning/inattention artefact).

I was expecting the authors to state that a proportion of field plots were discarded as unreliable (and therefore ineligible for further analysis), or were repeated until a reliable plot was obtained, but this does not appear to have been done. To quote this very paper: “approximately 25% of patients (n=734) with epilepsy are unable to produce a conclusive visual field test at any visit [13].”

Again we would like to thank the reviewer for pointing out that we have omitted useful information which would re-assure our reader. We can confirm that four patients were removed from the study either due to a field defect not attributed to VGB (25%) another two participants were removed because of poor reliability (13%). Additionally at the first visit, all patients underwent a 30-2 visual field examination on both eyes using the Full Threshold algorithm. This visit served to reduce the learning effect observed in perimetry and the results were not used for data analysis. In our MS we now state:

“Twenty-Two participants: 12 females and 10 males (mean age 38.54 years, SD 13.45, range 16 to 61 years) who were undergoing or who had previously undergone treatment with VGB were invited to take part in the study. Four participants were removed from the study as they had a visual field defect
not attributed to VGB another two participants were removed because of poor reliability. The sample therefore consisted of 16 epilepsy patients; 10 females and 6 males (mean age 39.3, SD 14.52, range 18 to 61).

The ‘Methods’ section includes the sentence “All visual fields fell inside the criteria of <33% false positive and negative catch trials and <20% fixation losses” - was this actually a ‘result’ or was it an inclusion criterion? In clinical practice, an "unreliable" SITA visual field plot with 30% false negatives (or even 10% False neg) may often resemble VAVFL. By the authors’ own admission (their reference 13), they were expecting ‘approximately 25% of [subjects] unable to produce a conclusive visual field test at any visit’ - A plot with up to 32% False Negatives due to learning/inattention artefact may superficially resemble VAVFL: this referee suggests that much ‘tighter’ reliability criteria should be used for inclusion in the analysis. Which (if any) of the 6 sets of ‘VAVFL’ plots were actually considered reliable?

We thank the reviewer for their suggestion of including the reliability criteria for all patients. It is now included in supplementary Table 1 (See below). We acknowledge that false negative rate is slightly higher for those with significant visual field loss however it is now well recognised that higher the FN catch trial methods are inadequate for estimating patient attentiveness in perimetry in eyes with significant visual field loss. Instead the frequency of false-negative responses in eyes with visual field defects is associated with amount of field loss (Bengtsson & Heijl 2000; Heijl et al 2012). Because of this we now state in the MS:

All visual fields fell inside the criteria of less than 33% false positive, less than or equal to 33% false negative and 20% fixation losses (See supplementary Table 1). We acknowledge that the false negative rate was higher for the patients with significant visual field loss. However, it is now well recognised in perimetry that the false negative catch trial methods are not suitable for estimating patient attentiveness in eyes with significant visual field loss as the frequency of false-negative responses in eyes with visual field defects is associated with amount of field loss [21].
Supplementary Table 1: Reliability criteria for all patients (FT = Full Threshold; SS = SITA Standard; SF = SITA Fast; V2 = visit 2; V3 = Visit 3)

| Patient no. | FT Fixation losses V2 | FT False positive V2 | FT False Negative V2 | FT Fixation losses V3 | FT False positive V3 | FT False Negative V3 | SS Fixation losses V2 | SS False positive V2 | SS False Negative V2 | SS Fixation losses V3 | SS False positive V3 | SS False Negative V3 | SF Fixation losses V2 | SF False positive V2 | SF False Negative V2 | SF Fixation losses V3 | SF False positive V3 | SF False Negative V3 |
|-------------|------------------------|----------------------|----------------------|------------------------|----------------------|----------------------|------------------------|----------------------|----------------------|------------------------|----------------------|----------------------|------------------------|----------------------|----------------------|------------------------|----------------------|----------------------|----------------------|
| 1           | 0%                     | 0%                   | 18%                  | 7%                     | 0%                   | 21%                  | 0%                     | 3%                   | 21%                  | 0%                     | 4%                   | 6%                   | 13%                    | 3%                   | 33%                  | 0%                     | 2%                   | 33%                  |
| 2           | 3%                     | 0%                   | 5%                   | 8%                     | 5%                   | 0%                   | 0%                     | 8%                   | 5%                   | 1%                     | 9%                   | 0%                   | 0%                     | 14%                  | 6%                   | 2%                     | 16%                  |
| 3           | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 2%                     | 6%                   | 0%                   | 0%                     | 2%                   | 0%                   | 0%                     | 0%                   |
| 4           | 4%                     | 7%                   | 0%                   | 9%                     | 14%                  | 7%                   | 0%                     | 0%                   | 0%                   | 1%                     | 0%                   | 25%                  | 0%                     | 0%                   | 8%                   | 0%                     | 0%                   |
| 5           | 3%                     | 0%                   | 0%                   | 0%                     | 0%                   | 6%                   | 0%                     | 0%                   | 7%                   | 0%                     | 2%                   | 10%                  | 7%                     | 0%                   | 0%                   | 0%                     | 0%                   | 11%                  |
| 6           | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 11%                   | 1%                   | 0%                   | 0%                     | 5%                   | 0%                   | 0%                     | 7%                   | 0%                   | 11%                   | 8%                   | 0%                   | 5%                   |
| 7           | 4%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 11%                   | 1%                   | 0%                     | 5%                   | 0%                   | 6%                     | 17%                  | 7%                   | 0%                     | 0%                     | 0%                   | 3%                   | 10%                  |
| 8           | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 4%                   | 7%                     | 0%                   | 13%                  | 12%                   | 0%                   | 5%                   | 0%                     | 6%                     | 17%                  | 7%                   | 0%                     | 0%                   | 0%                   | 3%                   | 10%                  |
| 9           | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 6%                   | 1%                   | 1%                     | 0%                   | 0%                     | 0%                   | 0%                   | 14%                  | 0%                   | 0%                   |
| 10          | 0%                     | 0%                   | 7%                   | 4%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 6%                     | 1%                   | 1%                   | 0%                     | 0%                     | 0%                   | 0%                   | 14%                  | 0%                   | 0%                   |
| 11          | 0%                     | 0%                   | 18%                  | 0%                     | 0%                   | 12%                  | 0%                     | 7%                   | 7%                   | 6%                     | 5%                   | 7%                   | 8%                     | 5%                     | 1%                   | 8%                     | 8%                   | 15%                  |
| 12          | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 1%                   | 2%                   | 15%                    | 1%                   | 1%                   | 0%                     | 0%                   | 0%                   |
| 13          | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 0%                   | 0%                   | 0%                     | 1%                   | 0%                   | 1%                     | 0%                     | 1%                   | 0%                     | 0%                   |
| 14          | 3%                     | 0%                   | 32%                  | 0%                     | 0%                   | 5%                   | 0%                     | 0%                   | 20%                  | 5%                     | 0%                   | 6%                     | 7%                     | 4%                     | 20%                  | 0%                     | 0%                     | 0%                   |
| 15          | 4%                     | 0%                   | 11%                  | 0%                     | 0%                   | 10%                  | 0%                     | 0%                   | 10%                  | 1%                     | 4%                   | 26%                  | 0%                     | 0%                     | 6%                     | 0%                     | 0%                     | 10%                  |
| 16          | 7%                     | 0%                   | 0%                   | 18%                  | 0%                     | 0%                   | 11%                  | 2%                     | 5%                     | 18%                  | 2%                     | 0%                     | 0%                     | 4%                     | 0%                     | 4%                     | 10%                  |
I see that only 6 of the 16 patients ostensibly had VAVFL. It would have been useful to have mentioned this at the start of the ‘results’ section.

We thank the reviewer for their sensible suggestion and now state at the start of the results section:

“Forty four percent of the patients had a confirmed clinical diagnosis of VGB attributed field loss from their medical records. The diagnosis was made independently from the research study for purely clinical purposes. Visual field defects exhibited a bilateral symmetrical defect showing concentric constriction of the peripheral visual field which was more pronounced nasally and typically characteristic of VGB attributed field loss. The other 66% of the sample had a confirmed clinical diagnosis of no visual field loss from their medical records.”

When determining the false positive and false negative rate (clinical status) we have included a new table (Table 4) and state:

“All visual fields were categorised using the classification defined by Wild et al (2009) [19]. Visual defects ranged from mild to severe based on the number and position of stimulus locations exhibiting an abnormality at either $p<0.01$ or $p<0.005$ out to 30 degrees eccentricity for static threshold perimetry and were present in 38% after Full threshold was assigned the gold standard.”

Finally we highlight the difference in the prevalence of defect figures (44% versus 38%) by stating:

“Closer inspection of the “normal” Full Threshold visual field plot (patient 15 Figure 3) reveals that if those locations demonstrating a 2 % loss on pattern probability analysis were also included in the analysis then the patient would have been diagnosed with a VGB attributed defect. Additionally, information from their medical records shows that this patient had a confirmed clinical diagnosis of a VGB attributed field loss with previous visual field testing.

Please can the severity of VAVFL (Wild criteria) for each patient be mentioned too, if it is applicable at all.

The Wild et al 1999 criteria is a specific set of peer reviewed criteria which is designed to classify VGB attributed field loss into four categories. At the reviewers suggestion we have now included the severity of VAVL into Table 4 below.
Figure 3 does not show the reliability indices: what were the reliability indices for these plots? (Are these data analyisable at all? – see previous paragraph). As far as I can tell from the data presented, it could be that all 6 patients in figure 3 might be people who are “unable to produce a conclusive visual field test at any visit”. maybe these patients were all experienced at doing VF and all had a clinical diagnosis of definite VAVFL, confirmed by OCT, or maybe (as i suspect) none of the 16 had ever done VF before: the paper does not make this at all clear.

Again we would like to confirm that all patients with VGB attributed field loss had a clinical diagnosis of VGB attributed field loss that was made independent from the study. Additionally, all reliability criteria are included in supplementary Table 1 (see above)

This clause is not true, it should be omitted: “Patients with compressive optic neuropathy and optic neuritis have nerve fibre defects similar to glaucoma”

The statement has now been removed