Author’s response to reviews

Title: Repeatability and reproducibility of Applanation Resonance Tonometry: a cross-sectional study

Authors:

Laura Ottobelli (laura.o@fastwebnet.it)
Paolo Fogagnolo (fogagnolopaolo@googlemail.com)
Paolo Frezzotti (frezzottip@unisi.it)
Stefano De Cillà (stefano_de_cilla@yahoo.co.uk)
Elena Vallenzasca (elena.vallenzasca@virgilio.it)
Maurizio Digiuni (mauriziodigiuni@gmail.com)
Ruggiero Paderni (doctorpaderni@email.it)
Ilaria Motoles (ilaria.motoles@libero.it)
Simone A Bagaglia (bagasimo@gmail.com)
Eduardo Motoles (motoles@unisi.it)
Luca Rossetti (luca.rossetti@unimi.it)

Version: 11
Date: 8 December 2014

Author’s response to reviews: see over
Point-by-point responses to the reviewer’s comments on the manuscript

"Repeatability and reproducibility of Applanation Resonance Tonometry: a cross-sectional study"

Reviewer: Dr. Luciana Alencar

Major Compulsory Revisions

1. The authors have correctly improved the description of repeatability on the text (“In the entire cohort, ART1 was 0.4±2.2 mmHg (-7.0 to 5.7 mmHg) higher than ART2 (p=0.03)“), the same complete sentence should be added to the abstract. In addition, the conclusion that “ART repeatability and reproducibility were almost perfect” is way too strong when differences between the two measurements may vary from -7.0 to 5.7 mmHg. Consider adding a Bland-Altman plot with limits of agreement for the repeatability (ART1 & ART2), as this is one of the main purposes of the study. The sentence added to the second paragraph of results (“Test-retest difference with ART fell within \( \pm 1 \) mmHg in 41\% of cases, within \( \pm 2 \) mmHg in 70\%, within \( \pm 3 \) mmHg in 85\%. 15\% had a test-retest difference higher than \( \pm 3 \) mmHg.”) is a very good way of describing the findings, and could be added to the abstract.

In this point the Reviewer correctly suggests that caution is required in defining ART repeatability and reproducibility almost perfect when differences between the two measurements may vary from -7.0 to 5.7 mmHg. In this version of the paper, we modified the sentence “ART repeatability and reproducibility were almost perfect” using higher caution: “In most cases ART repeatability and reproducibility were high” (see line 48). Bland-Altman plot for ART repeatability has been added as Figure 1; 95\% intervals of confidence were -3.9 and +4.6 mmHg (see lines 45-46; 206-207). The two sentences have been added to abstract as recommended (see lines 40-41; 43-45).

2. Detecting outliers is important, as one would not want results to be biased by typos or by failure of the instrument. The authors are to be commended for identifying and reporting the exclusion of a case in which GAT was missing. Nevertheless, excluding the other 4 cases is more controversial and needs more clarifying. One possible reason for artifacts and spurious measurements is bad quality. Although the authors report having assessed quality of ART measurements, and having repeated any measurements with \( Q \geq 3 \), they do not report how many eyes had \( Q > 3 \) and if they excluded any cases in which no measurements with good quality were obtained. For the clinician that would consider using such tonometer it is not clear why in the excluded cases the repeatability was so poor and how should they identify such potential errors.

Table B on the author’s point-by-point responses could be added to the final text (with the addition of quality score to each excluded case).

Table B has been added to the manuscript as recommended and Q scores for each ART measurement have been added as well (see table 7 in the new version of the manuscript).

In the Discussion we mentioned that “These cases had Q-values of 1 or 2 (as defined in Methods) and no procedural deviations or abnormal eye characteristics were found. One case was excluded as GAT measurement had not been reported on data sheet” (see lines 264-267).

3. The authors state on their review that “…Also, the difference between ART and GAT increased at higher IOP levels, thus confirming the findings of a previous paper [16]. According to Jóhannesson et al., ART failed to respect the ISO standards for patients with IOP > 23 mmHg, when compared to GAT [14].” Consequently, it is not clear why the authors
decided to further study the repeatability of this tonometer and to “suggest a possible use in clinical practice”. Should it be used as a screening tonometer? Should every result above 23 mmHg be discarded and the patient referred to GAT? Should two measurements be obtained at all times? How should this research translate into clinical use?

We inspected sources of difference between ART and GAT and found out that differences tended to increase at higher IOP values (Figure 2A and B in the new version of the manuscript) as previously described [16]. Yet, $R^2$ values were considerably low, a fact that may indicate that regression model may be a poor descriptor of study data, in particular only a minority of measurements was higher than normal. A properly designed study to explore the characteristics of ART in eyes with IOP higher than 21 mmHg is recommendable, also considering that in a previous paper ART failed to respect the ISO standards for patients with IOP > 23 mmHg, when compared to GAT [14] (see lines 272-278).