Reply

Carlos M. Oliveira

Critical Health SA, Coimbra, Portugal

I am writing as the corresponding author of the article 'Improved Automated Screening of Diabetic Retinopathy' which received some comments raised by Fleming et al. I would like to start by stating my appreciation for the interest shown in the results we have presented, and I will in this letter offer clarification on the issues that have arisen from the analysis of the data included.

As stated in the paper, and observed by Fleming et al., the results presented come from a retrospective study applying the novel screening technology to patient data collected in the years 2008 and 2009. Contrary to the remarks by Fleming et al., the sensitivity was indeed calculated, as shown in the article and in the tables. Since the comments made by Fleming et al. show that the article was not fully understood, we will add the following clarifications.

In a diabetic screening context, an automated system indicates those patients who should be referred to a specialist (positives). The specialist then either corrects the indication (false positives) and calls the patients back in 1 year (the typical

screening interval) or confirms the indication (true positives) and refers the patients to treatment, thus excluding them from the screening program.

The proposed performance can be evaluated in a screening program, as demonstrated, and both in a retrospective study and in an experimental scenario a true positive never returns in the next screening interval since that patient is referred for treatment. The only patients in a given screening event who could have been previously classified as positive by the software are those for whom the specialist previously did not confirm the indication (false positives) as explained in the article. Further, by combining the exams of the two screening occasions, the novel software increases its specificity.

Fleming et al. also mention the issue of two graders not always agreeing with each other. Previous publications such as that of Bernardes et al. [1] have shown that while two graders may not always agree in the number of MAs present in an image, the agreement in the cases of patients who need to be referred to treatment is very

high. For this reason, and since the software identifies those cases, the comparison presented is the only way to compare methods published in the literature. One needs to be careful, though, when considering the cases of publications which refer to a small cohort of patients.

Regarding the issue of quality control, it should be noted that all comparisons presented in the tables excluded the cases considered ungradable, both for the novel software and for those already published in literature. Due to the retrospective nature of the study, patient exams considered ungradable were not available for the study, thus not enabling quality assessment. Again, to be clear, when comparing results the ungradable examinations were excluded from all of the studies.

Reference

- 1 Bernardes R, Nunes S, Pereira I, Torrent T, Rosa A, Coelho D, Cunha-Vaz J: Computer-assisted microaneurysm turnover in the early stages of diabetic retinopathy. *Ophthalmologica* 2009;223:284–291.