Reanalysis of visual grading characteristics (VGC) data using VGC Analyzer

J Hansson[1,2], L G Månsson[1,2] and M Båth[1,2]

[1]Department of Medical Physics and Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg, Sweden
[2]Department of Radiation Physics, the Sahlgrenska Academy, University of Gothenburg, Sweden

Purpose: Visual grading studies have a weak point in that the statistical evaluation of collected data is often performed in a questionable manner. The introduction of visual grading characteristics (VGC) analysis in 2007 aimed at an improvement by presenting a non-parametric rank-invariant method of comparing visual grading data from two modalities. The resulting figure-of-merit, the area under the VGC curve (AUC_{VGC}), was initially proposed to be determined using standard software for receiver operating characteristics (ROC) analysis. However, due to different properties of ROC and VGC data, it can be assumed that the uncertainty of the AUC_{VGC} is not determined correctly using ROC software. A novel developed software tool (VGC Analyzer) for statistical analysis of VGC data using non-parametric resampling methods has recently been verified on simulated data. The purpose of the present work was to reanalyse data from previously published VGC studies using VGC Analyzer in order to evaluate the validity of the reported results of the previous studies and to investigate the behaviour of VGC Analyzer on real data.

Method: Visual grading data from previously performed studies on optimisation of x-ray examinations were reanalysed using VGC Analyzer. The outcome (the mean and 95% confidence interval (CI) of the AUC_{VGC} and the p-value) was compared with previously reported data from the studies where single reader adapted ROC software had been used and rating data from multiple readers had been pooled. The studies included both paired and non-paired data and were analysed using both fixed readers and random readers.

Major findings: The results showed good agreement between the AUC_{VGC} determined with VGC Analyzer and the previously used methods. However, on non-paired data, narrower CIs were reported by previous studies compared to VGC Analyzer whereas in one study with paired data the reported CIs were similar or even broader using ROC software. Significance testing based on the CI and the p-value from VGC Analyzer in most cases gave the same result, but sometimes led to different results in statistically weak studies when the random reader approach was used. In these cases a p-value < 0.05 was obtained although the CI of the AUC_{VGC} included 0.5.

Conclusions: The results of the present work indicate that studies using single reader adapted ROC software for analysing VGC data may, especially in non-paired data studies, underestimate the uncertainty of the obtained AUC_{VGC}, leading to an increased risk of Type I errors. On the other hand, incorrect use of ROC software for analysis of paired data may overestimate the uncertainty of the obtained AUC_{VGC}, leading to increased risk of Type II errors.