

Program Number: 1002

Presentation Time: 3:45 PM–4:00 PM

Automated diabetic retinopathy image assessment softwares: large scale, real world evaluation of diagnostic accuracy and cost-effectiveness compared to human graders

Catherine A. Egan^{1,2}, Alicja Rudnicka³, Christopher Owen³,

Caroline Rudisill⁴, Sebastian Salas-Vega⁴, Paul Taylor⁵,

Gerald Liew¹, Aaron Lee⁷, Clare Bailey⁶, John Anderson⁸,

Adnan Tufail^{1,2}. ¹Medical Retina, Moorfields Eye Hospital, London,

United Kingdom; ²University College London, London, United

Kingdom; ³St George's University, London, United Kingdom;

⁴Department of Social Policy, London School of Economics and

Political Science, London, United Kingdom; ⁵Institute of Health

Informatics, University College London, London, United Kingdom;

⁶Bristol Eye Hospital, Bristol, United Kingdom; ⁷Department of

Ophthalmology, University of Washington Seattle, Seattle, WA;

⁸Homerton University Hospital, London, United Kingdom.

Purpose: Diabetic retinopathy screening involves labour intensive manual grading of retinal images. Automated Retinal Image Analysis (ARIA) software can determine whether diabetic retinal disease is present as an alternative to human graders. We aim to evaluate the clinical effectiveness/ cost-effectiveness of three CE marked ARIA softwares in a large number of patient images acquired as part of routine diabetic retinopathy screening in a NHS (National Health Service) setting in the United Kingdom.

Methods: Observational measurement comparison study of 20,258 consecutive patients and a decision analytic model was undertaken to determine the effectiveness and cost-effectiveness of three ARIA systems (Retmarker, iGrading, and EyeART) in replacing one or more steps of human grading in a NHS Diabetic Eye Screening Programme (DESP). Images were graded by human graders as well as the ARIA systems, before being sent for an arbitration. Secondary analysis explored the influence of patients' ethnicity, age, sex and camera on screening performance.

Results: The sensitivity point estimates (95% confidence interval) of the ARIA systems are as follows: EyeArt 94.7% (94.2-95.2) for any retinopathy, 93.8% (92.9-94.6) for referable retinopathy 99.6% (97.0-99.9) for R3 proliferative retinopathy; Retmarker 73.0% (72.0-74.0) for any retinopathy, 85.0% (83.6-86.2) for referable retinopathy 97.9% (94.9-99.1) for R3 proliferative retinopathy. iGradingM classified all images as either having disease or being ungradeable, this limited further analyses for iGradingM. Both EyeArt and Retmarker are cost saving relative to manual grading either as a replacement for Level 1 human grading or as a filter prior to Level 1 human grading.

Conclusions: Retmarker and EyeArt achieved acceptable sensitivity for referable retinopathy when compared with a quality-assured, real world human grader working in a high volume clinical setting as a reference standard and had specificity sufficient to make them cost effective alternatives to a purely manual grading approach.

Commercial Relationships: Catherine A. Egan; Alicja Rudnicka, None; Christopher Owen, None; Caroline Rudisill, None; Sebastian Salas-Vega, None; Paul Taylor, None; Gerald Liew, None; Aaron Lee, None; Clare Bailey, None; John Anderson, None; Adnan Tufail, None

Support: This project was funded by the National Institute for Health Research HTA programme (project no. 11/21/02); Fight for Sight Hirsch grant award; and the Department of Health's NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and UCL Institute of Ophthalmology. The views expressed are those of the authors, not necessarily those of the Department of Health