Letters to the Editor

Is automated screening for diabetic retinopathy indeed not yet ready as stated by Grauslund et al.?

Frank D. Verbraak,[1] Ursula Schmidt-Erfurth,[2] Andrzej Grzybowski,[3] Michael Abramoff[4] and Reinier Schlingemann[1]

[1]Department of Ophthalmology, Amsterdam University Medical Center, Amsterdam, The Netherlands; [2]Department of Ophthalmology, Medical University Vienna, Vienna, Austria; [3]Department of Ophthalmology, Uniwersytet Warminsko-Mazurski, Olsztyn, Poland; [4]Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA

Editor,

In a recent paper by Grauslund et al. (2018), the evidence-based guidelines for screening of diabetic retinopathy (DR) in Denmark were presented. The authors addressed several aspects of screening: classification of DR, examination techniques, screening interval and the possible role of automated screening.

The authors are a group of retinal specialists formed by the Danish Ophthalmological Society, but how the guideline was conceived, and whether as systematic approach was used, is not described in detail in the article. In short, they recommend the use of mydriatic two-field disc- and fovea-centred images, for grading the International Clinical Retinopathy Disease severity score, supplementary optical coherence tomography (OCT) in case of suspicion of macular oedema, and a flexible individualized screening interval.

However, even though they underscore the need for automated screening in light of the increasing burden of any screening program for DR, with numbers of patients with diabetes mellitus expected to rise considerably in the coming years, they considered automated screening not yet ready for implementation.

The authors acknowledge that the performance in sensitivity (reportedly up to 96%) and specificity (reportedly up to 98.5%) of the latest artificial intelligence-based algorithms for automated screening is high enough, but they have two arguments against the present use of automated devices for DR screening in Denmark. The first is based on assumed differences between study populations and the real-world Danish screening population, the second is the assumption that most of these algorithms use single-field non-mydriatic fundus images.

Four recent studies, validating different artificial intelligence-based devices for DR screening have included a wide variety of patients, all representing real-life populations, with a mix of multiple ethnicities (Gulshan et al. 2016; Ting et al. 2017; Abramoff et al. 2018; Verbraak et al. 2019). These studies showed that the tested devices have a high accuracy with high sensitivity and specificity, against various reference standards. Most importantly, one of these studies was a preregistered clinical trial overseen by FDA, and compared to AI-based system to a patient outcome proxy and was powered to measure equity, that is comparable accuracy for different races and ethnicities. Based on these studies, the conclusion can be drawn that any population studied appropriately can be reliably screened using these devices, irrespective of the kind of patients screened. Differences in the prevalence of DR, or in the degree of severity of DR encountered in screening populations, can also not be used as an argument against the use of these devices, as the distribution of DR severity in these studies mimics that in real-world populations (Verbraak et al. 2019).

With the exception of the study by Gulshan et al., the other three studies used fundus photographs as recommended by the authors, including one optic disc-centred and one fovea-centred fundus image, and not only a single fovea-centred image (Gulshan et al. 2016; Ting et al. 2017; Abramoff et al. 2018; Verbraak et al. 2019).

The pivotal trial conducted by Abramoff et al., which lead to FDA approval of the IDx-DR device was preregistered, and used analysis of the device on these two standard fundus images, and compared this with the only known patient outcome proxy for DR, the 4-widefield stereo images, covering the seven fields defined in the ETDRS classification studies, graded by the Wisconsin reading centre, combined with OCT-based diagnosis of centre-involved diabetic macular oedema as the gold standard (Abramoff et al. 2018). Even against this high level of accuracy of the gold standard, the sensitivity and specificity of this device were very high.

We would like to argue that at present, AI-based devices for screening of DR are indeed ready to be implemented into the care of our patients with diabetes mellitus. These devices will not only help to reduce the growing burden of screening for DR, but improve the accuracy of screening compared to single human graders (Verbraak et al. 2019).

In line with our conclusion, we have noted that in a more recent article, the first author of the Danish Guideline article, J. Grauslund, appears to agree, stating that 'Considering both strengths and limitations, as well as the high performance of deep learning-based algorithms, automated DR classification using deep learning could be feasible in a real-world screening scenario' (Nielsen et al. 2019).

References

Abramoff MD, Lavin PT, Birch M, Shah N & Folk JC (2018): Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NP J Digit Med 1: 9.

Grauslund J, Andersen N, Andresen J et al. (2018): Evidence-based Danish guidelines for screening of diabetic retinopathy. Acta Ophthalmol 96: 763–769.

Gulshan V, Peng L, Coram M et al. (2016): Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316: 2402–2410.

Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR & Grauslund J. (2019): Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol Retina 3: 294–304.

Ting DSW, Cheung CY, Lim G et al. (2017): Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images

from multi-ethnic populations with diabetes. JAMA **318**: 2211–2223.

Verbraak FD, Abramoff MD, Bausch GCF et al. (2019): Diagnostic accuracy of a device for automated detection of diabetic retinopathy in a primary care setting. Diabetes Care **42**: 651–656.

Reply: Is automated screening for DR indeed not yet ready as stated by Grauslund et al?

Jakob Grauslund,[1,2,3] Nis Andersen,[1,4] Jens Andresen,[1,4] Per Flesner,[1,4] Per Haamann,[1,4] Steffen Heegaard,[1,5] Michael Larsen,[1,5] Caroline Schmidt Laugesen,[1,6] Katja Schielke,[1,7] Jesper Skov[1,4] and Toke Bek[1,8]

[1]Danish Ophthalmological Society, Copenhagen, Denmark; [2]Department of Ophthalmology, Odense University Hospital, Odense, Denmark; [3]Department of Clinical Research, University of Southern Denmark, Odense, Denmark; [4]Organization of Danish Ophthalmologists, Copenhagen, Denmark; [5]Department of Ophthalmology, Rigshospitalet, Copenhagen, Denmark; [6]Department of Ophthalmology, Zealand University Hospital Roskilde, Roskilde, Denmark; [7]Department of Ophthalmology, Aalborg University Hospital, Aalborg, Denmark; [8]Department of Ophthalmology, Aarhus University Hospital, Aarhus, Denmark

Editor,

We welcome the correspondence by Verbraak et al. in response to the official Danish evidence-based consensus guidelines for diabetic retinopathy (DR) screening (Grauslund et al. 2018). The guidelines address the issue of whether or not automated methods of image analysis should be applied in DR screening as a means of limiting costs and promoting speed and efficacy. Advances in artificial intelligence (AI) based on deep learning (DL) algorithms reported after the guidelines were prepared have addressed some of the limitations that we listed, as stated by Verbraak et al. We remain concerned, however, about the accuracy of automated DR screening, its potential to limit healthcare expenses and the operational issues that notoriously arise when new computer systems are introduced with little assurance that they will improve the overall function of increasingly complex health care systems.

In Denmark, the retinopathy severity level that should lead to referral from a non-physician-staffed screening system to a physician-staffed echelon supervised by ophthalmologists is sight-threatening DR (STDR), which includes proliferative diabetic retinopathy (PDR) and clinically significant diabetic macular oedema (CSMO). The typical automated screening algorithm is trained, however, to select patients for referral at a much lower threshold of retinopathy (Gulshan et al. 2016; Ting et al. 2017; Abràmoff et al. 2018). We have not found documentation that they have been tested against the reference standards used in Denmark. If used as they are, they would lead to a drastic increase in referrals to the specialist level. To illustrate the issue, the DR reference standard of the Google algorithm was defined as moderate non-proliferative DR (NPDR) or worse (Gulshan et al. 2016). Last year, 7.4% of patients in the Danish screening system met this threshold (Dansk Diabetes Database National årsrapport 2017/2018), which is a stark contrast to the 0.7% who require referral for active PDR. The patients that Google would refer to the specialist level, whereas we service them at the screening level, typically in an outpatient clinic with comprehensive one-stop diabetes service, are the 3.8% of patients who have non-proliferative DR, the 0.7% with severe NPDR and the 2.2% with stabile proliferative DR (Grauslund et al. 2018). This practice saves 9 out of 10 patients with inconvenience of going to the ophthalmologists' office, which is likely to promote compliance and reduce costs.

We are not against automated fundus image analysis as a method of augmenting diabetes care, but a total procedural shift would have to be evaluated in a real-life setting against a state-of-the-art system with a documented ability to produce attractive long-term outcomes in terms of safety, efficacy and equity. Additionally, we expect automated screening systems to fill some of the gaps that may exists in screening systems operated by non-ophthalmologists, for instance in the advanced evaluation of the optic disc degenerative retinopathy.

References

Abràmoff MD, Lavin PT, Birch M, Shah N & Folk JC (2018): Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Med **1**: 39.

Dansk Diabetes Database National årsrapport (2017/2018): https://www.sundhed.dk/content/cms/87/4687_diabetes_2017_18_offentliggoerelse_v2.pdf

Grauslund J, Andersen N, Andresen J et al. (2018): Evidence-based Danish guidelines for screening of diabetic retinopathy. Acta Ophthalmol **96**: 763–769.

Gulshan V, Peng L, Coram M et al. (2016): Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**: 2402–2410.

Ting DSW, Cheung CY, Lim G et al. (2017): Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA **318**: 2211–2223.