Progress in examining cost-effectiveness of AI in diabetic retinopathy screening

Artificial intelligence (AI) technology using deep learning systems (DLS) is becoming increasingly important in the provision of medical care in ophthalmology.1 A major ophthalmological complication of diabetes is diabetic retinopathy, for which regular screening is recommended;2 however, given the large number of individuals diagnosed with diabetes worldwide,3 screening for diabetic retinopathy can be costly.4 A DLS developed in Singapore has been shown to have comparable diagnostic performance to human assessors.5 In India, an automated AI algorithm for diabetic retinopathy and vision-threatening diabetic retinopathy detection using Remidio Fundus on a phone imaging device tested on 296 patients, was shown to have 96% sensitivity and 80% specificity in detecting any diabetic retinopathy, and 99% sensitivity and 80% specificity in detecting vision-threatening diabetic retinopathy. In Thailand, a Google AI model was evaluated for 7517 individuals with 25326 gradeable images. Referable diabetic retinopathy was detected with a sensitivity of 97% by the DLS and 75% by human assessment; however, specificity was lower for the DLS model than human assessment. In China, a DLS model (Inception V3) was evaluated for the detection of vision-threatening diabetic retinopathy (preproliferative) or worse, and for diabetic maculae oedema, tested on 19900 images. Referable diabetic retinopathy showed a sensitivity of 97% and specificity of 91%. In Kenya, a machine language software tested on 3460 individuals with 6788 fundus photos showed a 91% sensitivity and 70% specificity. In Zambia, an ensemble DLS model tested on 1547 individuals with 4504 retinal fundus images found a sensitivity of 92% and a specificity of 89% for referable diabetic retinopathy.4

However, until the study by Yuchen Xie and colleagues in The Lancet Digital Health,7 the cost implications of this DLS implementation in Singapore had not been evaluated. Xie and colleagues di a model-based cost-minimisation analysis of three diabetic retinopathy screening models in Singapore comparing current human assessment with a semi-automated DLS model with a triage before secondary human assessment for confirmatory diagnoses, and a fully automated DLS model without human assessment. The authors found that the least-expensive model was the semi-automated DLS model with an annual cost of US$62 per person per year, compared with $66 for the fully automated DLS model, and $77 for the human assessment. On the basis of estimated diabetes prevalence in 2050, the authors estimate potential annual cost savings for Singapore to be as high as $15 million.

Xie and colleagues state that these results are due to a higher proportion of false positives under the fully automated model, which leads to unnecessary specialist visits and their associated costs.

The cost-effectiveness of screening for diabetic retinopathy, especially vision-threatening retinopathy has been investigated in several studies from other countries. Cost-effectiveness of population-based screening programmes has been shown to vary depending on the frequency of retinal examinations and retinal imaging. Extending the screening interval from once a year to every two or three years in patients with diabetes without evidence of retinopathy at first eye examination, has been shown to be cost-effective in a number of European studies. However, it is important to separate patients into low-risk and high-risk groups to further improve cost-effectiveness.8

Xie and colleagues state that they assumed equivalent effectiveness across all potential screening techniques, and therefore do not include treatment costs in the model. They also stated that labour costs between countries could affect the results if the same study was done outside of Singapore. It is important to note that in countries where labour costs are high compared with technology costs, these results would be valid, and perhaps DLS would be even more cost-effective relative to human assessment. However, in countries where labour costs are low compared with technology costs, the human assessment model could be more financially beneficial. For this reason, it will be important to replicate the study by Xie and colleagues in other countries.

In addition to the potential contribution of false positives to the difference in the cost savings between the fully automated and semi-automated models, the CIs surrounding these false positive and false negative estimates could also be important, because very high or very low values can affect cost-effectiveness. High
values of false positives will increase unnecessary follow-up specialty care costs, whereas high values of false negatives will increase the costs of undetected diabetic retinopathy because of resulting complications.

Despite the limitations, Xie and colleagues have made a very important contribution to the AI and DLS medical literature, by showing the increasing potential AI and DLS have to improve medical care delivery to greater numbers of patients worldwide in multiple medical specialties in a financially beneficial modality.

I declare no competing interests. This work is the opinion of the author only, and does not represent any position or opinion of the US Department of Veterans Affairs.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Clara Dismuke
clama.dismuke@va.gov