

PERSPECTIVE OPEN



Effectiveness of artificial intelligence screening in preventing vision loss from diabetes: a policy model

Roomasa Channa^{1✉}, Risa M. Wolf², Michael D. Abràmoff³ and Harold P. Lehmann⁴

The effectiveness of using artificial intelligence (AI) systems to perform diabetic retinal exams ('screening') on preventing vision loss is not known. We designed the Care Process for Preventing Vision Loss from Diabetes (CAREVL), as a Markov model to compare the effectiveness of point-of-care autonomous AI-based screening with in-office clinical exam by an eye care provider (ECP), on preventing vision loss among patients with diabetes. The estimated incidence of vision loss at 5 years was 1535 per 100,000 in the AI-screened group compared to 1625 per 100,000 in the ECP group, leading to a modelled risk difference of 90 per 100,000. The base-case CAREVL model estimated that an autonomous AI-based screening strategy would result in 27,000 fewer Americans with vision loss at 5 years compared with ECP. Vision loss at 5 years remained lower in the AI-screened group compared to the ECP group, in a wide range of parameters including optimistic estimates biased toward ECP. Real-world modifiable factors associated with processes of care could further increase its effectiveness. Of these factors, increased adherence with treatment was estimated to have the greatest impact.

npj Digital Medicine (2023)6:53; <https://doi.org/10.1038/s41746-023-00785-z>

INTRODUCTION

Digital Technology, including autonomous Artificial Intelligence (AI), has the potential to improve patient outcomes, reduce health disparities, improve access to care, and lower health-care costs^{1–5}. Typical metrics for evaluation of new technology focus on efficacy^{6,7}. In the case of diagnostic AI systems these efficacy metrics translate into diagnostic-accuracy measures, such as sensitivity and specificity, compared to an agreed upon reference standard⁸. Multiple AI systems have been shown to be safe and efficacious using such metrics, resulting in FDA De Novo clearance and clinical use^{9–11}. While these diagnostic-accuracy metrics correctly estimate the efficacy of the diagnostic AI system, they do not give information on the overall impact (effectiveness) of the AI system on patient outcomes⁶.

Instead, the impact of implementing AI on patient outcome is dependent on many factors beyond the diagnostic accuracy of AI⁷. These factors include characteristics of the disease, such as prevalence, and natural history, as well as potential frictions in the care process, including access to care, adherence with a recommended referral, and adherence with treatment and management recommendations. In addition, treatment itself where indicated is unlikely to be perfect, and may itself lead to imperfect outcomes. Therefore, even if a diagnostic AI with perfect accuracy is implemented, outcomes will be affected by these frictions associated with processes of care, as will potential efficiency gains, and differential effects on health inequities. These processes of care frictions/imperfections may be less obvious, as they cannot be determined from inspection of the standalone AI system, but instead depend greatly on how the AI system is integrated into the care process as well as the health delivery network. While some AI systems, such as those used in the critical care environment may affect patient outcome in real time, in many cases, AI systems are designed for chronic conditions, where

a difference in outcome may take years or even decades to manifest. Thus, process-of-care metrics need to be considered in addition to outcomes to determine whether it is worth designing, developing, validating, implementing, regulating and reimbursing such AI systems⁶.

An example of an AI system that has the potential to affect real-world outcomes is the first diagnostic autonomous AI (IDx-DR, Digital Diagnostics Inc, Coralville, Iowa). It received US FDA De Novo clearance in 2018 to autonomously, that is without human oversight, diagnose diabetic retinopathy and macular edema—Diabetic Retinal Disease (DRD)¹². Clearance was based on efficacy, as determined in a preregistered clinical trial¹¹, which provided information on the diagnostic-accuracy metrics of sensitivity and specificity, but not on effectiveness or impact on patient outcomes. As diabetes is a chronic disease, it will take years to determine this impact, requiring following each patient that interacted with the AI system to a disease endpoint for years. Given the lack of such empirical data of the impact on patient outcome (vision loss), we modeled screening strategies and the downstream care process, as the Care Process for Preventing Vision Loss from Diabetes (CAREVL) policy model, to estimate the impact on patient outcome (vision loss).

The primary purpose of this study was to develop the CAREVL model and leverage it to determine the differential impact of autonomous AI-based diabetic retinal exams ('screening') vs screening performed in the clinic by an eye care provider (ECP). Secondly, its purpose was to explore how processes of care modulate the effectiveness of screening strategies.

RESULTS

All analytical inputs are listed in Table 1 and detailed in the Supplementary.

¹Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, WI, USA. ²Department of Pediatrics, Division of Endocrinology, Johns Hopkins Medicine, Baltimore, MD, USA. ³Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA, USA. ⁴Department of Medicine, Section on Biomedical Informatics and Data Science, Johns Hopkins University, Baltimore, MD, USA. ✉email: rchanna@wisc.edu

Table 1. Parameters for the decision model.

Parameter Description	Base-case estimate	For sensitivity analysis	
		Low	High
<i>Population-metrics: Prevalence and natural history of disease</i>			
Prevalence of Metabolic DRD ¹¹	0.22	0	0.40
Prevalence of Ophthalmic DRD ¹¹	0.0088 ^a	0	0.10
Prevalence of DRD with Vision Loss ^{42,43}	0.01	0	0.05
No DRD to metabolic DRD ⁴⁴	0.05	0	0.15
Metabolic DRD to ophthalmic DRD ⁴⁴	0.02	0	0.20
Ophthalmic DRD to vision loss ^{39,40}	0.075	0	0.20
Vision loss to irreversible vision loss ⁴⁵	0.37	0	0.50
<i>Diagnostic-accuracy metrics: Sensitivity and Specificity of Screening Strategies</i>			
Sensitivity of screening for DRD with AI ^{11,13,17}	0.87	0	1
Sensitivity of screening for DRD with ECP ⁴⁶	0.33	0	1
Specificity of screening for DRD with AI ^{11,13,17}	0.91	0	1
Specificity of screening for DRD with ECP ⁴⁶	0.99	0	1
<i>Process-of-care metrics: Screening and Referral for Appropriate Care</i>			
Probability that patient follows up for eye care after AI screen positive ^{4,47–49}	0.75	0	0.95
Probability that patient follows up for eye exams after ECP screen positive ⁴⁷	0.29	0	1
Probability of patient Accepting Screening by AI ⁵⁰	0.95	0	1
Probability of patient Accepting Screening by ECP ^{4,5,27,47,51,52}	0.20	0	0.80
Probability that patient with Vision Loss Accepts referral to ECP ²⁸	0.58	0	0.75
<i>Process-of-care metrics: Effectiveness of treatments for DRD(Progression of treated disease)</i>			
Metabolic DRD to ophthalmic DRD ⁴⁴	0.01	0	0.05
Ophthalmic DRD to vision loss ^{39,41,53}	0.02	0	0.50
Vision loss to irreversible vision loss ⁵⁴	0.034	0	0.05
<i>Process-of-care metrics: Probability of Adherence to Treatment</i>			
Adhering to metabolic management ^{24,25,55}	0.24	0	1
Adhering to ophthalmic management ^{26,28}	0.26 ^a	0	1
Adhering to DRD vision loss management ²⁸	0.41 ^a	0	1

^aCalculated values: see Supplementary.

Base-case and sensitivity analysis

For the base case, in the no-screening strategy the proportion of adults with DM who are estimated to develop any vision loss at 5 years is 1637/100,000; it is 1625/100,000 for the ECP screening strategy, and 1535/100,000 for the AI screening strategy. Thus, the proportion of DM participants who develop vision loss in the model with AI-based screening (1535/100,000) is estimated to be 102/100,000 lower compared with no-screening (1637/100,000) and 90/100,000 lower compared with ECP-based screening (1625/100,000). The difference between no-screening and ECP is 12/100,000. Thus, CAREVL suggests that introduction of an AI-based screening strategy is 8.6 times more effective at preventing vision loss than ECP, under base-case assumptions. No meaningful

thresholds were found in one-way or two-way sensitivity analyses (See Table 2, supplementary Table 1 and supplementary Fig. 2 in Supplementary). The results of the two-way sensitivity analysis (supplementary Table 1) showed that across the broad range of sensitivity values AI dominated over ECP across all ranges for the following two-way comparisons: sensitivity and specificity of AI vs ECP screening, and accepting AI vs ECP screening. For the comparison regarding accepting referral after AI vs ECP screening, AI dominated except in the unlikely scenario of low probability of accepting referral after AI and a high probability of accepting referral after ECP. This scenario is far from the base-case, for further clarification, the output of this two-sensitivity analysis is shown in supplemental Fig. 3. In 2019, an estimated 37.3 million Americans had diabetes. If we use a conservative estimate of 30 million Americans with diabetes, based on the numbers above we anticipate that AI-based screening strategy is expected to prevent vision loss in over 27,000 more Americans at 5 years as compared to ECP-based strategy, under base-case assumptions.

Maximal scenarios

The scenario analyses show that, if adherence to recommended metabolic and ophthalmic treatment were maximized, the estimated total number with any vision loss at 5 years would be lower for both AI and ECP strategies when compared to the base case, with a higher reduction noted for the AI strategy. In the scenario that maximizes adherence to recommended treatment, the number with any vision loss by 5 years is estimated to be 1167/100,000 for the AI strategy, an additional reduction of 367/100,000 from the AI base case. In this scenario, the estimated total number with any vision loss for the ECP strategy instead is 1488/100,000, an additional reduction of 137/100,000 from the ECP base case. In all scenarios tested, the number with vision loss per 100,000 is lower with the AI strategy compared with the ECP strategy. Figure 1 shows the relative impact of increasing the probability of adhering with metabolic and ophthalmic treatments on projected vision loss for each screening strategy. Figure 2 shows the impact of maximizing diagnostic and process-of-care metrics on vision loss when using the AI screening strategy. The largest impacts are when adherence with recommended metabolic or ophthalmic treatments is maximized. The scenario with maximum adherence to metabolic treatment (100%) results in 110/100,000 fewer patients progressing to vision loss. The scenario with maximum adherence to ophthalmic treatment (100%), results in 294/100,000, fewer patients progressing, and maximizing both results in 367/100,000 fewer patients progressing. These numbers suggest an accretive effect of adherence to both metabolic and ophthalmic treatment. Using a conservative estimate of 30 million Americans with diabetes this translates into vision loss prevented in over 110,000 additional Americans with diabetes when AI-based screening is introduced, and treatment adherence is maximized. Maximizing the effectiveness of metabolic and ophthalmic treatments themselves, namely more effective drugs or procedures—does have a marginal impact of 25–28/100,000 fewer progressing, but this is only 6.8–7.6% of the benefit achieved by maximizing adherence to therapies that are currently available.

DISCUSSION

Using CAREVL, we conclude that autonomous AI is expected to be more effective than ECP-based screening at preventing vision loss among patients with diabetes. This effectiveness can be maximized by improving processes of care, particularly adherence with recommended treatments. Under base-case assumptions, introducing AI in a no-DRD-screening scenario is estimated to be 8.6 times more effective at preventing vision loss from DRD compared with introducing ECP-based screening. The expected

Table 2. One-way sensitivity analyses.

Population, diagnostic-accuracy, and process-of-care metrics	Type of metric	Base Case (minimum, maximum value for sensitivity analysis) ^a		Threshold ^d
		ECP	AI	
Prevalence of Metabolic DRD	Population	0.22 (0, 0.4)	0.22 (0,0.4)	AI dominates
Offered and accepts screening for DRD	Process of care	0.2 (0,0.8)	0.95 (0,1)	AI dominates
Sensitivity of the DRD screening strategy	Diagnostic accuracy	0.33 (0,1)	0.87 (0,1)	AI dominates
Specificity of DRD screening strategy	Diagnostic accuracy	0.99 (0,1)	0.91 (0,1)	AI dominates
Accepts referral for eye care after a positive screen	Process of care	0.29 (0,1)	0.75 (0, 0.95)	0.08 ^b
Probability of adhering with treatment	Process of care	0.24 (0,1)	0.24 (0,1)	AI dominates
Effectiveness of ophthalmic treatments	Process of care	0.02 (0,0.5)	0.02 (0,0.5)	0.1 ^c
Effectiveness of metabolic treatments	Process of care	0.01 (0,0.05)	0.01 (0,0.05)	AI dominates

DRD diabetic retinal disease, ECP eye care provider, AI artificial intelligence.

^aThe no-screening option is dominated by AI or ECP in all scenarios and has therefore not been included in this table.

^bFar from base case value of 0.75.

^cFar from base case value of 0.02.

^dAI dominates refers to the finding that AI is the preferred strategy on each of the one-way sensitivity analyses (across the range of the minimum and maximum values for the parameter specified in parenthesis next to the base-case value).

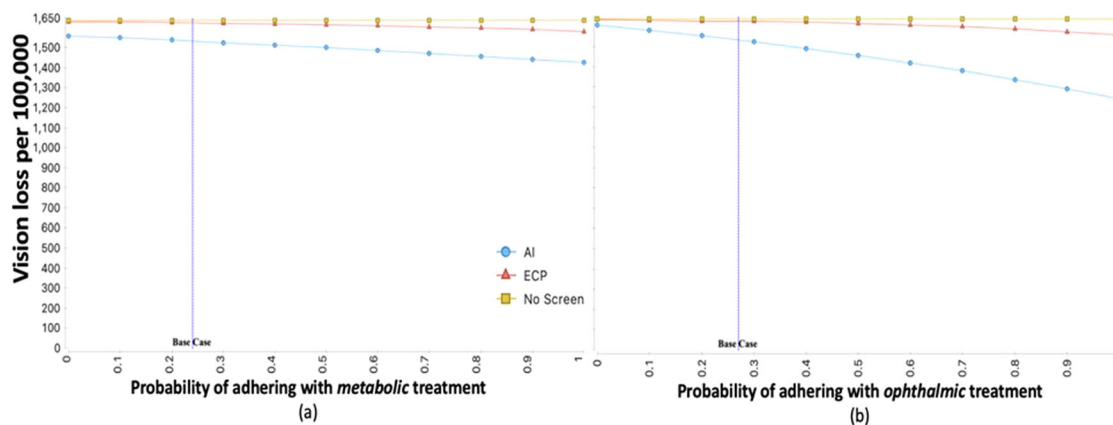


Fig. 1 Expected vision loss per 100,000 vs probability of adhering with treatment for each screening strategy. **a, b** Show that as the adherence with recommended metabolic and ophthalmic treatments increases the number of patients with vision loss per 100,000 decreases for both the eye care provider (ECP) and artificial intelligence (AI) screening strategies. However, the decrease in number with vision loss is more marked for the AI vs ECP screening strategy.

differences between ECP and autonomous AI screening strategies are likely due to a combination of factors. While the *efficacy* of AI screening systems in detecting referable DRD has been established in prospective trials^{11,13}, the main drivers of higher *effectiveness* are likely the point-of-care availability of AI and immediate diagnostic output which make it more likely for a patient to accept screening and the recommended referral^{4,14}.

The CAREVL model is novel in that it allows evaluating the effectiveness of AI algorithms within the context of real-world patient workflow and the health-care system. Our approach is based on standards for performing and reporting modeling of expected impact of digital health technologies^{7,15,16}. Studies evaluating AI have traditionally focused on its diagnostic-accuracy metrics for a given task^{11,17–19}. However, as we have shown, to evaluate the effectiveness of AI in the real-world we need to know not only its diagnostic accuracy or how it performs in controlled research settings, but its impact on patient outcomes⁷. This is important because many digital and non-digital health interventions may work well in an ideal ‘model’ setting, but real-world evaluation often reveals outcomes that are

much less compelling compared to what can be achieved in a clinical trial setting^{6,20–22}.

The CAREVL model further allows us to study the expected impact of adjusting process-of-care metrics on patient outcomes and to identify which metrics may be most important in maximizing the effectiveness of AI within a health-care system, given a chosen strategy. This effort has immediate real-world implications for the implementation of AI-enabled patient-centered care. The CAREVL model suggests that the full potential of AI algorithms in preventing vision loss can be achieved by optimizing processes of care. Among the process-of-care metrics evaluated in the model, adherence with recommended metabolic and ophthalmic treatments had the largest impact on preventing vision loss. Prior studies such as the one by Rohan, et al., have estimated that screening and early treatment of DRD can prevent vision loss and reduce risk of blindness by an estimated 56%²³. However, this estimate is predicated on perfect-world assumptions of 100% of patients accepting screening, high sensitivity of detecting referable disease (88%), and 100% complying with recommended treatments. Real-world data from the US regarding adherence to metabolic management show that, on average, only

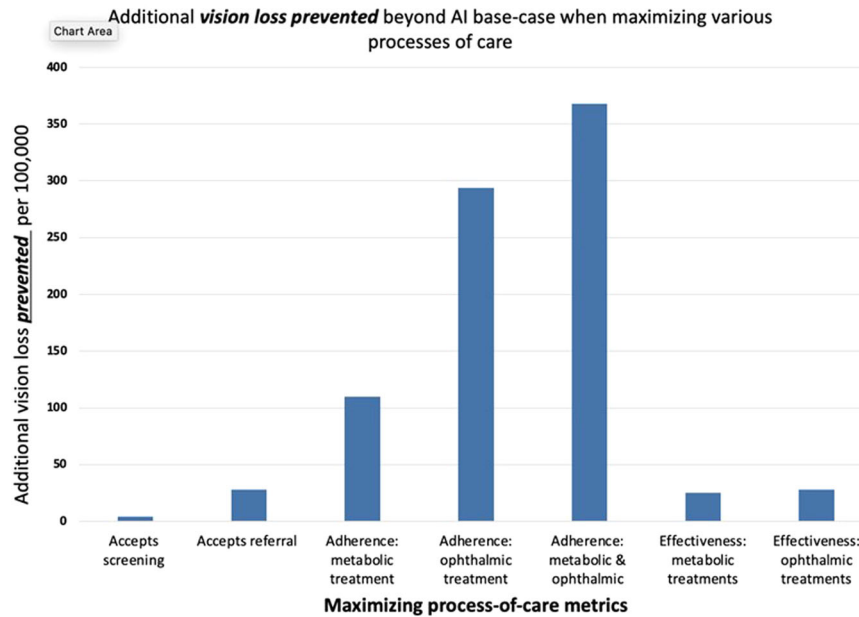


Fig. 2 Additional vision loss prevented beyond AI base-case when maximizing processes of care. Figure 2 shows the additional impact on vision loss prevented beyond the base-case scenario when each of the processes of care are maximized. The largest impact on vision loss is estimated to be from maximizing adherence with ophthalmic treatment, followed by adherence with metabolic treatments. Maximizing effectiveness of current metabolic and ophthalmic treatments has a lower impact.

about 22% of patients with diabetes achieve the recommended lipid, blood pressure and glucose control and only about 24% of patients with Type 2 DM achieve a glycosylated-hemoglobin level of <8%^{24,25}. Adherence with recommended screening eye exams for DRD and follow-up eye care is similarly low²⁶. Analysis of insured patients with diabetes showed that only about 15%²⁷ met the American Diabetes Association's recommendation for annual DRD screening and data from the National Health Interview Survey showed that only about a third of insured adults in the US followed up for eye care in the absence of visual impairment²⁸. These low rates are concerning, as DRD is asymptomatic until late stages, hence the existence of Healthcare Effectiveness Data and Information Set (HEDIS) and Merit-based Incentive Payment System quality measures that incentivize diabetic retinal exams to be performed early and regularly²⁹.

The CAREVL model confirms that improving adherence with both the current metabolic and ophthalmic treatments is key to maximizing the success of implementing DRD screening strategies. The model suggests that when autonomous AI is used as a screening strategy, maximizing adherence with metabolic and ophthalmic treatments prevents vision loss in an additional 367 patients/100,000. This reduction is ~4 times more than just introducing AI without improving the process of care. While it remains important to develop increasingly effective treatments for metabolic and ophthalmic DRD, CAREVL suggests that its population impact on vision loss is much lower (~one-tenth) than that of maximizing adherence with existing treatments. This projected impact has important clinical and public health implications. Diabetes is a chronic disease that currently affects almost 37 million adults in the US^{30,31}, thus introducing AI-based screening could potentially prevent vision loss in an additional 27,000 patients with diabetes over the current ECP-based standard of care. Introducing AI and optimizing processes of care, particularly adherence with recommended treatment, could potentially prevent vision loss in an additional 110,000 patients. These benefits are expected to accrue as the prevalence of diabetes continues to rise. A more nuanced estimate would require further modeling to account for age distributions, annual incidence of diabetes and patient mortality.

The strength of our study is that we developed a real-world model, CAREVL, defining how to evaluate the effectiveness of an AI-based technology on patient outcomes. CAREVL is a relatively novel and more patient-centered approach to evaluating AI technologies as opposed to the overwhelming focus on evaluating diagnostic accuracy. Evaluating the impact of AI and digital health technologies on patient outcomes is an evolving area of research and we have made the model publicly available and invite others to contribute to it. The CAREVL model and this study have limitations. We relied on available, published and peer-reviewed data for the various metrics. It is important to collect real-world data over time, particularly with regards to ECP parameters, to further validate this model. This model does not address costs or utilities. As we are focused on effectiveness, we have considered vision loss in either eye. In future analyses focused on cost and disability benefits it may be better to consider vision loss in the worse seeing eye^{32,33}. We did not compare the effectiveness of AI-based screening strategies with telemedicine programs as there is considerable variation between programs but once the relevant metrics identified in the model are collected, the relative effectiveness of telemedicine programs can be determined. One of the limitations of the study is that we did not model the benefit of ophthalmic encounters with an ECP as opposed to AI in potentially detecting diseases other than DRD (e.g., cataracts, macular degeneration, glaucoma). Our rationales for this decision were that (a) patients with visually significant cataract will have vision impairment (by definition) and will visit eye care instead of entering the screening pathway; (b) the number of potential missed cases is small: the pivotal trial that led to FDA approval of autonomous AI estimated that 0.2% of participants with glaucoma and 1.6% with non-exudative age-related macular degeneration may be missed by AI-based screening, no cases of neovascular age-related macular degeneration were noted¹¹. Furthermore, the United States Preventive Services Task Force has determined that there is insufficient evidence to recommend screening for impaired vision from age-related changes such as age-related macular degeneration and glaucoma at this point^{34,35}. We have analyzed the overall impact of autonomous AI in the US, and not the outcomes of specific

sub-groups within a population—these require more sophisticated models that we are currently developing. Nevertheless, we expect that CAREVL can be extended in well-established ways to help in answering questions regarding the impact of new technology on real-world outcomes from multiple perspectives (healthcare system, payor, or society).

In summary, our novel CAREVL model suggests that AI-based DRD screening is more effective at preventing vision loss from diabetes than ECP-based screening, and that this effectiveness can be further enhanced by optimizing processes of care. As use of digital health technology and AI increases in the healthcare system, this comprehensive model may serve as a framework for evaluating and estimating the real-world impact of digital technologies on patient outcomes in other chronic disease scenarios.

METHODS

Model structure

CAREVL is implemented as a computer-simulation based on a state-transition Markov model decision tree (one Markov model for each screening strategy). The model considers *population-metrics*, namely, community prevalence of the multiple severity stages of DRD, natural history of disease; *diagnostic-accuracy metrics* namely sensitivity and specificity; and *process-of-care metrics*, namely, the probability of accepting screening, of follow-up in case of a positive screen, of adherence with metabolic and ophthalmic management of DRD, and the effectiveness of metabolic and ophthalmic treatment. Figure 3 shows the states considered and the transitions

between states permitted in the Markov model. The parameters of the model are probabilities and related quantities for states and transitions defined by the structure of the decision tree. The parameter values were derived from peer-reviewed published literature, and the base-case estimates are presented. Where choices for the base-case values were required, we biased the model against autonomous AI. Where relevant, the probabilities extracted from the literature were converted to transition probabilities³⁶. The 12 model assumptions are detailed in the Supplementary. The models were built in TreeAge software (TreeAge Pro Healthcare version 2021 R1.1, Williamstown, MA), and we have made a spreadsheet version of the model available in the supplementary materials via Figshare (<https://figshare.com/s/ad7809b8f7010fdf83c9>). The parameters used in CAREVL are summarized in Table 1 and detailed in the Supplement. Fig. 3 shows the Markov-model structure used for each screening strategy. The full decision tree is in the Supplementary Fig. 1.

Target population

The target population is adults with Type 1 or 2 DM (age > 21 years) under regular care by a primary care physician, endocrinologist or other licensed provider. The base-case assumption of prevalence of DRD and its stages in primary care was estimated from a representative sample, as this population was drawn from adult patients with diabetes presenting to primary care settings with a racial and ethnic distribution that is representative of the 37 million people with diabetes in the US^{11,30,31}. People with diabetes eligible for screening were categorized into three states: no DRD, metabolic DRD or ophthalmic DRD, defined as follows:

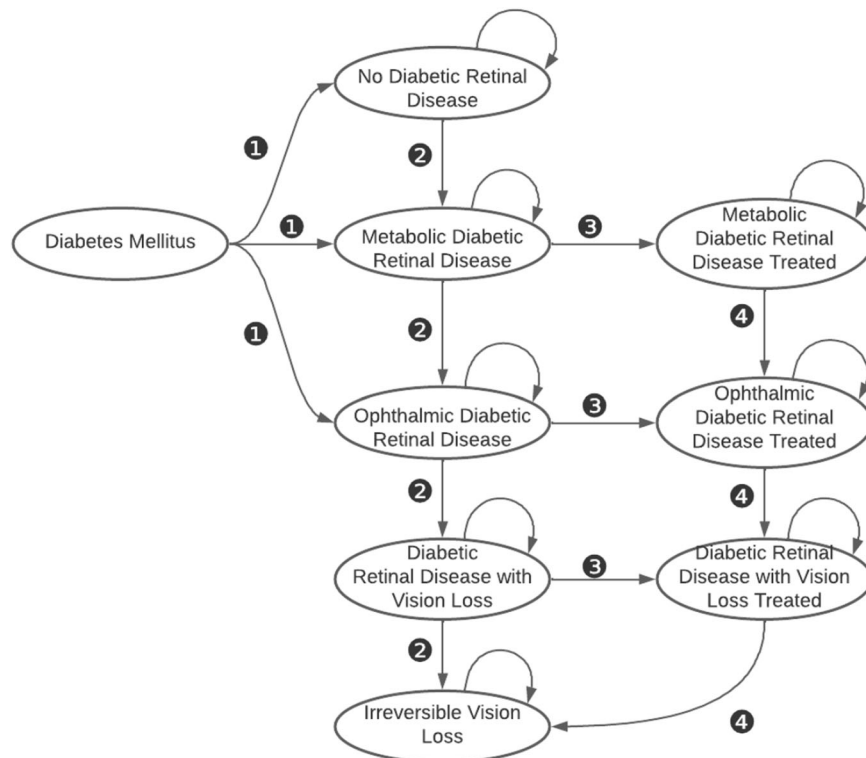


Fig. 3 Markov model showing the states and transitions relevant to diabetic retinal disease used in the current analysis. ①Patients with diabetes mellitus presenting to the primary care or endocrine clinic with each of the following states: No Diabetic Retinal Disease, Metabolic Diabetic Retinal Disease, or Ophthalmic Diabetic Retinal Disease. ②Natural history transitions of diabetic retinal disease. ③Transitions from untreated to treated diabetic retinal disease. ④Transitions of treated diabetic retinal disease. The transitions take into account process-of-care metrics i.e., probability of accepting screening and referral in case of a positive screen, probability of disease progression, probability of adhering with recommended treatments. The structure of the Markov model is the same for both screening strategies. Table 1 shows the base-case probabilities and limits of sensitivity analysis for each parameter that are specific to the AI and ECP screening strategies. The details of the transitions specific to each strategy are represented in the decision tree in the supplement (Fig. 3 is preserved and shared on Figshare (<https://figshare.com/s/ad7809b8f7010fdf83c9>)).

mild, moderate and severe non-proliferative DRD (ETDRS levels 35–53) primarily require metabolic control and are categorized as “metabolic DRD”; “ophthalmic DRD” is defined as DRD requiring ophthalmic treatment in addition to metabolic treatment and is taken as equivalent to ETDRS level 60 and higher (i.e., proliferative diabetic retinopathy (PDR)) or having clinically significant macular edema or center-involved macular edema, without symptoms of vision loss. The prevalences of these states were varied in sensitivity analyses. Patients with known vision loss, are recommended to go directly for eye care as opposed to first going through a screening exam and were not included in the model^{37,38}.

Screening strategies

The CAREVL model is designed and built from the patient’s perspective. Three alternative strategies for the diabetic eye exam were modeled: (1) no screening; (2) ECP strategy, where all patients are referred by the diabetes or primary care provider to an ophthalmologist or optometrist—referred to as ECP—for dilated diabetic eye exams in the clinic; (3) autonomous AI strategy, where the a digital fundus photograph is acquired and an autonomous artificial intelligence (AI) algorithm is used to analyze the image, real time result is provided, and only those diagnosed as having diabetic retinopathy or diabetic macular edema (DME) are referred for further management to ECP. The no-screening strategy was included to assess the relative impact of the other two screening strategies on expected visual outcome.

Main outcomes and measures

The model is focused on clinical outcome—any vision loss experienced by the patient. Specifically, outcome is quantified as the probability of severe vision loss by 5 years. Because of the established benefit of treatments for PDR and DME, today, it is impossible to ethically collect natural history outcome data on untreated PDR or DME. Therefore, we used the most recent data from landmark randomized clinical trials for treatment that still had natural-history arms for PDR and DME^{39,40}. In these studies, severe vision loss was defined as worse than 5/200³⁹ and loss of 15 or more letters⁴⁰ on a standardized visual acuity chart. For visual outcomes of treated DME we used data from the anti-vascular endothelial growth factor (VEGF) treated arm of diabetic retinopathy clinical research network’s protocol I⁴¹, a landmark clinical trial which established the effectiveness of anti-VEGF agents for DME treatment. In that study, vision loss was defined as visual acuity of 20/200 or worse. Irreversible vision loss was the probability of visual acuity of 20/200 or worse at 2 years with or without treatment in either eye.

Sensitivity and scenario analyses

To evaluate the outcome under varying scenarios, and to account for uncertainty in base-case estimates, one-way sensitivity analyses were performed by varying one parameter at a time, while holding the others constant at their base-case estimates. Sensitivity analysis in decision analysis, to address uncertainty in model parameters, plays the same role as confidence intervals do in empirical statistical studies, to address uncertainty due to sampling. For those parameters that had different base-case values for AI vs ECP i.e., sensitivity and specificity of both strategies; accepting screening and accepting referral after screening, multiple two-way sensitivity analyses were conducted to determine if there were any scenarios where a strategy other than that identified in the base-case would be preferred. The relative impact of varying the values of key parameters on vision loss using either one of the screening strategies was evaluated. The key parameters included (1) population metrics; (2) diagnostic-accuracy metrics; (3) process-of-care metrics (see Table 1). We additionally created a series of maximal scenarios where

individual parameters of process of care were set to each one’s maximum value and assessed the marginal impact of each maximum scenario over the base-case dominant strategy. The goal of the maximal-scenario analysis was to estimate the potential impact of maximizing process-of-care metrics on expected vision loss.

The manuscript complies with the Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS) checklist¹⁵; this checklist was chosen because, while ours is not an economic evaluation, this checklist comes closest to governing the type of study we present.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The online [supplementary material](#) includes a macro-enabled excel sheet that provides all data that were used to build the model.

CODE AVAILABILITY

The model was built using the TreeAge software. No separate code was written to run the Markov model.

Received: 12 May 2022; Accepted: 24 February 2023;

Published online: 27 March 2023

REFERENCES

- Xie, Y. et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health* **2**, e240–e249 (2020).
- Wolff, J., Pauling, J., Keck, A. & Baumbach, J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. *J. Med. Internet Res.* **22**, e16866 (2020).
- Wolf, R. M., Channa, R., Abramoff, M. D. & Lehmann, H. P. Cost-effectiveness of Autonomous Point-of-Care Diabetic Retinopathy Screening for Pediatric Patients With Diabetes. *JAMA Ophthalmol.* **138**, 1063–9 (2020).
- Liu, J. et al. Diabetic Retinopathy Screening with Automated Retinal Image Analysis in a Primary Care Setting Improves Adherence to Ophthalmic Care. *Ophthalmol. Retin.* **5**, 71–77 (2020).
- Fuller, S. D. et al. Five-Year Cost-Effectiveness Modeling of Primary Care-Based, Nonmydriatic Automated Retinal Image Analysis Screening Among Low-Income Patients with Diabetes. *J. Diabetes Sci. Technol.* **16**, 415–27 (2020). 1932296820967011.
- Haynes, B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* **319**, 652–653 (1999).
- World Health Organization (2016) Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. <https://apps.who.int/iris/handle/10665/252183>. License: CC BY-NC-SA 3.0 IGO
- Abramoff, M. D. et al. Foundational Considerations for Artificial Intelligence Using Ophthalmic Images. *Ophthalmology* **129**, e14–e32 (2022).
- Thomasian, N. M., Eickhoff, C. & Adashi, E. Y. Advancing health equity with artificial intelligence. *J. Public Health Policy* **42**, 602–611 (2021).
- FDA-approved A.I.-based algorithms. <https://medicalfuturist.com/fda-approved-ai-based-algorithms/> Accessed March, 2022.
- Abramoff, M. D., Lavin, P. T., Birch, M., Shah, N. & Folk, J. C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digit. Med.* **1**, 39 (2018).
- FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>. Updated 04/12/2018. Accessed March 2022.
- Ipp, E. et al. Pivotal Evaluation of an Artificial Intelligence System for Autonomous Detection of Referrable and Vision-Threatening Diabetic Retinopathy. *JAMA Netw. Open.* **4**, e2134254–e2134254 (2021).
- Wolf, R. M. et al. The SEE Study: Safety, Efficacy, and Equity of Implementing Autonomous Artificial Intelligence for Diagnosing Diabetic Retinopathy in Youth. *Diabetes Care.* **44**, 781–787 (2021).

15. Huserau, D. et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Int. J. Technol. Assess. Health Care*. **38**, e13 (2022).
16. Philips, Z., Bojke, L., Sculpher, M., Claxton, K. & Golder, S. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment. *Pharmacoeconomics* **24**, 355–371 (2006).
17. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* **316**, 2402–2410 (2016).
18. Kim, H.-E. et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit. Health* **2**, e138–e148 (2020).
19. Ström, P. et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* **21**, 222–232 (2020).
20. Kwan, J. L. et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *Bmj* **370**, m3216 (2020).
21. Ciulla, T. A., Bracha, P., Pollack, J. & Williams, D. F. Real-world Outcomes of Anti-Vascular Endothelial Growth Factor Therapy in Diabetic Macular Edema in the United States. *Ophthalmol. Retin.* **2**, 1179–1187 (2018).
22. Ciulla, T. A. et al. Real-world Outcomes of Anti-Vascular Endothelial Growth Factor Therapy in Neovascular Age-Related Macular Degeneration in the United States. *Ophthalmol. Retin.* **2**, 645–653 (2018).
23. Rohan, T. E., Frost, C. D. & Wald, N. J. Prevention of blindness by screening for diabetic retinopathy: a quantitative assessment. *Br. Med. J.* **299**, 1198–1201 (1989).
24. Pantalone, K. M. et al. The Probability of A1C Goal Attainment in Patients With Uncontrolled Type 2 Diabetes in a Large Integrated Delivery System: A Prediction Model. *Diabetes Care*. **43**, 1910–1919 (2020).
25. Fang, M., Wang, D., Coresh, J. & Selvin, E. Trends in diabetes treatment and control in US adults, 1999–2018. *N. Engl. J. Med.* **384**, 2219–2228 (2021).
26. Bresnick, G. et al. Adherence to ophthalmology referral, treatment and follow-up after diabetic retinopathy screening in the primary care setting. *BMJ Open Diabetes Res. Care*. **8**, e001154 (2020).
27. Benoit, S. R., Swenor, B. & Geiss, L. S. Eye Care Utilization Among Insured People With Diabetes in the U.S., 2010–2014. *Diabetes Care* **42**, 427–433 (2019).
28. Lee, D. J. et al. Reported Eye Care Utilization and Health Insurance Status Among US Adults. *Arch. Ophthalmol.-Chic.* **127**, 303–310 (2009).
29. (NCQA) NCFQA. *HEDIS Measurement Year 2020 and Measurement Year 2021. Volume 2L Technical Specifications for Health Plans*. (National Committee for Quality Assurance (NCQA), Washington DC, 2020).
30. American Diabetes Association: Statistics About Diabetes. <https://www.diabetes.org/about-us/statistics/about-diabetes>. Updated 2/4/2022. Accessed April 2022.
31. Type 2 Diabetes. Centers for Disease Control and Prevention. <https://www.cdc.gov/diabetes/basics/type2.html>. Published 2021. Accessed April 2022.
32. If You're Blind or Have Low Vision — How We Can Help. Social Security Administration. <https://www.ssa.gov/pubs/EN-05-10052.pdf>. Published 2022. Accessed April 2022.
33. Hirneiss, C. The impact of a better-seeing eye and a worse-seeing eye on vision-related quality of life. *Clin. Ophthalmol. (Auckl., NZ)*. **8**, 1703–1709 (2014).
34. Primary Open-Angle Glaucoma: Screening. <https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/primary-open-angle-glaucoma-screening> Published 2022. Accessed Oct 2022.
35. Mangione, C. M. et al. Screening for Impaired Visual Acuity in Older Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* **327**, 2123–2128 (2022).
36. Naimark, D., Krahn, M. D., Naglie, G., Redelmeier, D. A. & Detsky, A. S. Primer on medical decision analysis: Part 5-Working with Markov processes. *Med Decis. Mak.* **17**, 152–159 (1997).
37. Flaxel, C. J. et al. Diabetic Retinopathy Preferred Practice Pattern®. *Ophthalmology* **127**, P66–p145 (2020).
38. Association AD. 11. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes—2021. *Diabetes Care*. **44**(Supplement_1), S151–S167 (2020).
39. Diabetic Retinopathy Study G. Photocoagulation treatment of proliferative diabetic retinopathy: clinical application of DRS findings: DRS report 8. *Ophthalmology* **88**, 583–600 (1981).
40. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch. Ophthalmol.* **103**, 1796–1806 (1985).
41. Elman, M. J. et al. Intravitreal Ranibizumab for diabetic macular edema with prompt versus deferred laser treatment: 5-year randomized trial results. *Ophthalmology* **122**, 375–381 (2015).
42. Prasad, S., Kamath, G. G., Jones, K., Clearkin, L. G. & Phillips, R. P. Prevalence of blindness and visual impairment in a population of people with diabetes. *Eye (Lond.)*. **15**, 640–643 (2001). (Pt 5).
43. de Fine Olivarius, N., Siersma, V., Almind, G. J. & Nielsen, N. V. Prevalence and progression of visual impairment in patients newly diagnosed with clinical type 2 diabetes: a 6-year follow up study. *BMC Public Health* **11**, 80 (2011).
44. Diabetes Control and Complications Trial Research Group, Nathan, D. M., et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986 (1993).
45. Early vitrectomy for severe vitreous hemorrhage in diabetic retinopathy. Two-year results of a randomized trial. Diabetic Retinopathy Vitrectomy Study report 2. The Diabetic Retinopathy Vitrectomy Study Research Group. *Arch. Ophthalmol.* **103**, 1644–1652 (1985).
46. Pugh, J. A. et al. Screening for diabetic retinopathy: the wide-angle retinal camera. *Diabetes Care*. **16**, 889–895 (1993).
47. Crossland, L. et al. Diabetic Retinopathy Screening and Monitoring of Early Stage Disease in Australian General Practice: Tackling Preventable Blindness within a Chronic Care Model. *J. Diabetes Res.* **2016**, 8405395 (2016).
48. Stebbins, K., Kieleyka, S. & Chaum, E. Follow-Up Compliance for Patients Diagnosed with Diabetic Retinopathy After Tele-retinal Imaging in Primary Care. *Telemed. e-Health* **27**, 303–307 (2021).
49. Jani, P. D. et al. Evaluation of Diabetic Retinal Screening and Factors for Ophthalmology Referral in a Telemedicine Network. *JAMA Ophthalmol.* **135**, 706–714 (2017).
50. Wolf, R. M. et al. The SEE Study: Safety, Efficacy, and Equity of Implementing Autonomous Artificial Intelligence for Diagnosing Diabetic Retinopathy in Youth. *Diabetes Care*. **44**, 781–787 (2021).
51. Mansberger, S. L. et al. Comparing the effectiveness of telemedicine and traditional surveillance in providing diabetic retinopathy screening examinations: a randomized controlled trial. *Telemed. e-Health* **19**, 942–948 (2013).
52. An, J., Niu, F., Turpcu, A., Rajput, Y. & Cheetham, T. C. Adherence to the American Diabetes Association retinal screening guidelines for population with diabetes in the United States. *Ophthalmic Epidemiol.* **25**, 257–265 (2018).
53. Wykoff, C. C. et al. Risk of Blindness Among Patients With Diabetes and Newly Diagnosed Diabetic Retinopathy. *Diabetes Care*. **44**, 748–756 (2021).
54. Antoszyk, A. N. et al. Effect of intravitreal aflibercept vs vitrectomy with pan-retinal photocoagulation on visual acuity in patients with vitreous hemorrhage from proliferative diabetic retinopathy: a randomized clinical trial. *Jama* **324**, 2383–2395 (2020).
55. Foster, N. C. et al. State of Type 1 Diabetes Management and Outcomes from the T1D Exchange in 2016–2018. *Diabetes Technol. Ther.* **21**, 66–72 (2019).

ACKNOWLEDGEMENTS

National Eye Institute: 7K23EY030911-02; R01EY032323. This work was supported in part by an Unrestricted Grant from Research to Prevent Blindness to the Department of Ophthalmology and Visual Sciences at University of Wisconsin Madison. The funders did not take any role in the design, collection, analysis, decision to publish or preparation of the paper.

AUTHOR CONTRIBUTIONS

All authors (R.C., R.W., M.D.A., H.P.L.) contributed to the conception, design and writing of the paper. R.C. and H.P.L. performed the analysis of the study.

COMPETING INTERESTS

R.C.: none. R.W.: Research support Dexcom. M.D.A.: reports the following competing interests: Investor, Director, Consultant of Digital Diagnostics Inc, Coralville, IA; patents and patent applications assigned to the University of Iowa and Digital Diagnostics that are relevant to the subject matter of this paper. H.P.L.: None.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-023-00785-z>.

Correspondence and requests for materials should be addressed to Roomasa Channa.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023