Original Investigation | Health Policy

Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy

Jesus Gomez Rossi, DMD, MSc; Natalia Rojas-Perilla, PhD; Joachim Krois, PhD; Falk Schwendicke, DMD, MDPH

Abstract

OBJECTIVE To assess the cost-effectiveness of artificial intelligence (AI) for supporting clinicians in detecting and grading diseases in dermatology, dentistry, and ophthalmology.

IMPORTANCE AI has been referred to as a facilitator for more precise, personalized, and safer health care, and AI algorithms have been reported to have diagnostic accuracies at or above the average physician in dermatology, dentistry, and ophthalmology.

DESIGN, SETTING, AND PARTICIPANTS This economic evaluation analyzed data from 3 Markov models used in previous cost-effectiveness studies that were adapted to compare AI vs standard of care to detect melanoma on skin photographs, dental caries on radiographs, and diabetic retinopathy on retina fundus imaging. The general US and German population aged 50 and 12 years, respectively, as well as individuals with diabetes in Brazil aged 40 years were modeled over their lifetime. Monte Carlo microsimulations and sensitivity analyses were used to capture lifetime efficacy and costs. An annual cycle length was chosen. Data were analyzed between February 2021 and August 2021.

EXPOSURE AI vs standard of care.

MAIN OUTCOMES AND MEASURES Association of AI with tooth retention–years for dentistry and quality-adjusted life-years (QALYs) for individuals in dermatology and ophthalmology; diagnostic costs.

RESULTS In 1000 microsimulations with 1000 random samples, AI as a diagnostic-support system showed limited cost-savings and gains in tooth retention–years and QALYs. In dermatology, AI showed mean costs of $750 (95% CI, $608-$970) and was associated with 86.5 QALYs (95% CI, 84.9-87.9 QALYs), while the control showed higher costs $759 (95% CI, $618-$970) with similar QALY outcome. In dentistry, AI accumulated costs of €320 (95% CI, €299-€341) (purchasing power parity [PPP] conversion, $429 [95% CI, $400-$458]) with 62.4 years per tooth retention (95% CI, 60.7-65.1 years). The control was associated with higher cost, €342 (95% CI, €318-€368) (PPP, $458; 95% CI, $426-$493) and fewer tooth retention–years (60.9 years; 95% CI, 60.5-63.1 years). In ophthalmology, AI accrued costs of R $1321 (95% CI, R $1283-R $1364) (PPP, $559; 95% CI, $543-$577) at 8.4 QALYs (95% CI, 8.0-8.7 QALYs), while the control was less expensive (R $1260; 95% CI, R $1222-R $1303) (PPP, $533; 95% CI, $517-$551) and associated with similar QALYs. Dominance in favor of AI was dependent on small differences in the fee paid for the service and the treatment assumed after diagnosis. The fee paid for AI was a factor in patient preferences in cost-effectiveness between strategies.

CONCLUSIONS AND RELEVANCE The findings of this study suggest that marginal improvements in diagnostic accuracy when using AI may translate into a marginal improvement in outcomes. The current evidence supporting AI as decision support from a cost-effectiveness perspective is limited;

(continued)

Key Points

Question Are existing artificial intelligence (AI) algorithms cost-effective for use as a decision-support system in dermatology, dentistry, and ophthalmology?

Findings In this economic evaluation analyzing data from 3 Markov models used in previous cost-effectiveness studies, the use of AI was associated with a modest improvement in outcomes. All benefits were highly dependent on treatment effects assumed after diagnosis and were very sensitive to the fee paid for the use of AI.

Meaning These results suggest that even when AI can achieve better diagnostic capacities than the average physician, this may not directly translate to better or cheaper care, and that analysis using this technology should be used on a case-by-case basis.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

AI should be evaluated on a case-specific basis to capture not only differences in costs and payment mechanisms but also treatment after diagnosis.

Introduction

Artificial intelligence (AI) is frequently referred to as a facilitator for more precise, personalized, and safer health care.[1,2] A major use of AI is decision support (ie, to help physicians detecting and grading diseases, such as through image analysis of skin photographs).[3] AI algorithms with diagnostic accuracies at or above the average physician have been reported in dermatology,[4] dentistry,[5] and ophthalmology,[6] among others.

Although US regulatory bodies (including the Food and Drug Administration [FDA]) approved the first AI diagnostic solution for the detection of diabetic retinopathy in 2018,[7] the benefits that this technology could generate on existing treatment paths have not been thoroughly assessed.[8,9] AI diagnostic solutions are currently under study in real-world settings in the US, India, Thailand, China, Australia,[10] and Singapore.[11] Importantly, these studies frequently take a third-party perspective and do not extrapolate over patient lifetime. Furthermore, differences between the setting in which an AI solution is deployed and where it is developed could open new questions of cost-effectiveness relevant to discussions of ever-rising health care costs.[12] New research is necessary to determine if AI can reduce costs and improve outcomes on its own, or if it may even increase pressure on existing resources.[13] An informed understanding can help decide possible reimbursement for the use of AI in diagnosis and to steer research and development to where most health and economic benefits can be expected.[14]

It is likely that the cost-effectiveness of AI depends on its diagnostic accuracy for the use case assumed (ie, Is it helping doctors or patients? What is the current standard of screening for the disease?), the patient population (What is the prevalence and costs of treatment for the disease studied?), and factors specific to the health care setting (What is the frequency of testing? What treatments do patients receive at each stage of the disease after being diagnosed?).

To the best of our knowledge, no previous study has modeled cost-effectiveness of existing AI algorithms for different use cases in different settings.[15] We aimed to evaluate AI's cost-effectiveness as a diagnostic support system in dermatology, dentistry, and ophthalmology in different countries using health economic modeling via Markov models with a lifetime horizon. We decided to account for AI as fee-for-service and explored how it factored into cost-effectiveness (per-person) through sensitivity analysis. Our research goal was to test the assumption that an AI with superior diagnostic accuracy used as a decision-support system would always clearly reduce costs and improve outcomes. Better understanding these aspects is particularly important for decision-makers assessing AI solutions, as well as for developers deciding to invest resources in decision-support systems using AI.

Methods

Study Design

Three model-based cost-effectiveness analyses were performed from the payer perspective for 3 diagnostic procedures in different medical disciplines—melanoma detection in dermatology, caries detection in dentistry, and detection of diabetic retinopathy in ophthalmology. AI as a diagnostic support system has been used previously to help detect and/or grade melanoma lesions on skin photography[4]; dental caries lesions on radiographs[16]; and diabetic retinopathy on fundus photography.[17] Our economic evaluations used data and models of previously published studies that

had performed cost-effectiveness analyses on each use case without involving AI (**Table**). In all cases, the sensitivity and specificity of AI as a diagnostic support system were compared with those of the standard of care.

The 3 use cases, AI applications, and health economic models are summarized in the Table in line with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) reporting guideline. Transitions between states and transition probabilities are explained in detail in eAppendices 1 to 3 in the Supplement. The settings of the different studies were the US for melanoma, Germany for caries detection, and Brazil for ophthalmology, with all parameters such as prevalence and life expectancy adjusted to these settings. Only 1 study considered the research and development costs of the AI application, which we extrapolated to the other 2 use cases, as is common practice in pharmacoeconomics.[23] We explored in a sensitivity analysis the effects of price variation. All economic models were constructed using Markov chains with simulations at discrete yearly intervals under a lifetime horizon. No approval by an ethics committee was requested as we performed a modeling exercise; data were deidentified and no original data were used.

Setting and Population

All 3 analyses adopted a payer perspective in 3 different health care settings. The US health care system is ranked as first in health care expenditure worldwide.[24] Expenditures are financed by a combination of voluntary health insurance, employer insurance, and out-of-pocket expenditures,

Table. Comparative Summary of Included Models

	Dermatology	Dentistry	Ophthalmology
Model characteristics			
Economic model source	Losina et al[18]	Schwendicke et al[19]	Ben et al[20]
AI accuracy model	Brinker et al[4]	Cantu et al[21]	Abramoff et al[22]
Target population	General population, age 50 y	Children, age 12 y	Individuals with diabetes, age >40 y
Perspective of payer	OOP	Third-party plus OOP	Third-party
AI use-case assumption	Decision support	Decision support	Decision support
Comparator	Standard dermatological screening	Standard dental screening	Standard ophthalmological screening
Setting and location	US	Germany	Brazil
Model utilized	Markov	Markov	Markov
AI development team location	Germany	Germany	US
Fee-for-use of AI[a]	US $8	€8	R $8
Measurement of outcomes	QALY/survival	Tooth-retention	QALY
Discounting[a]	3%	3%	3%
Study perspective	Lifetime	Lifetime	Lifetime
Currency and conversion	US$	Euro (€)	R$ transformed via PPP to US$
Opportunity costs	Not considered	Not considered	Not considered
Results (1000 microsimulations with 1000 random samples)			
AI			
Mean cost (95% CI)	US $750.35 ($608.77-$970.95)	€320.40 (€299-€341)	R $1321 (R $1283-R $1364)
2020 PPP (95% CI)	NA	$429.49 ($400.80-$458.76)	$559 ($543-$577)
QALYs (95% CI)	86.6 (84.9-88.0)	62.4 (61.6-63.1)[c]	8.42 (8.33-8.51)
Standard			
Mean cost (95% CI)	US $759.03 ($617.64-$980.73)	€342.24 (€318-€368)	R $1260.28 (R $1222-R $1303)
2020 PPP (95% CI)	NA	$458 ($426-$493)	$533 ($517-$551)
QALYs (95% CI)	86.6 (84.9-88.0)	60.9 (60.0-61.8)[c]	8.42 (8.33-8.51)

Abbreviations: AI, artificial intelligence; NA, not applicable; OOP, out-of-pocket; PPP, purchasing-power-parity; QALY, quality-adjusted life years; R$, Brazilian real.

[a] Explored in sensitivity analysis.

[b] 95% CIs ranged from 2.5% to 97.5% percentiles.

[c] Measured in tooth retention years as equivalent of QALYs.

with exceptions controlled by the government for older, disabled, and low-income populations.[25] In Germany, medical insurance, including dentistry, is 2-tiered, with most individuals (ie, over 87%) being publicly insured and only a minority being privately insured.[26] For members using public insurance, nearly all procedures are fully covered, while only some treatments are partially or fully paid out-of-pocket.[27] Brazil's universal public health care system is tax-funded by federal, state, and municipal governments and, despite limitations, offers comprehensive health coverage to the majority of its population.[28]

For the dermatological use case, direct costs to the health care payer in the US system (ie, health care system costs and patients' copayments combined) that would arise in the detection step, possible histological validation, and possible treatments and follow-up treatments were considered. Two cohorts of patients (AI vs control) entered the model to calculate morbidity, mortality, and costs. Individuals in both cohorts were in full health initially. The model assessed their risk of developing, being diagnosed, and being treated for melanoma by dermatologists, with the only difference between groups being the assistance of AI support.

For the dental use case, costs arising in the statutory German insurance as well as copayments by private insurance or out-of-pocket costs were considered, including detection costs with and without AI support and lifetime treatment and re-treatment costs. The unit of analysis was the tooth; both teeth that were sound or with an initial or advanced caries lesion were included, according to prevalence data drawn from a previous study.[29]

For the ophthalmological use case, a Brazilian taxpayer's perspective was taken. All costs accrued by the economic model, including treatment, were covered by the Brazilian National Health Service. We included in our model a group of patients with type II diabetes at risk of developing some form of diabetic retinopathy. Participants were tested biannually.

Comparators

For dermatology, the control group (ie, without AI) received the standard evaluation by dermatologists using a dermatoscope; accuracy for this group was extracted from previous studies.[30] Included treatments were derived from the health economic model that was used as a reference in our study.[18] The test group (AI) consisted of a convolutional neural network (CNN) for classifying skin photographs trained on 12 378 dermoscopic images labeled by 145 dermatologists.[4]

For dentistry, the control group was the detection of proximal caries lesion using biannual visual-tactile assessment and bitewing radiographs taken twice annually by dentists,[19] following to the health economic model that was used as reference.[31] In the test group, radiographic caries detection was assumed to be AI-assisted using a CNN that had been trained on 3293 images, validated on 252 images, and tested on 141 images (each of which had been labeled by 4 experts).[11]

For ophthalmology, the control group was the standard screening of diabetic retinopathy undertaken by ophthalmologists in Brazil,[20] in line with the economic model used as a reference for the study.[32] Diagnostic accuracy was modeled on the analysis of digital fundus photography previously used in the economic evaluation used as our data source.[33] The test group was a CNN trained on over 1 million lesions labeled according to a framework for automated lesion detection in retinal images.[34]

Models and Assumptions

For all 3 Markov models, initial and follow-up health states were included, with costs and utilities accrued for each transition. In the dermatological model, patients entered the model at age 48 years. In the case of dentistry, patients entered the model at age 12 years under the assumption that their permanent dentition is fully developed by then. In the case of ophthalmology, a population of individuals with diabetes entered the model at age 40 years, because according to US Centers for Disease Control and Prevention guidelines expanded screening strategies appear to be justified at that age.[35]

All models took a lifetime horizon according to their setting. In the case of melanoma, we differentiated between the risk of death to melanoma and the overall risk of death. In the case of dentistry, we followed tooth retention over average life expectancy, as tooth loss is an event that can be almost completely averted throughout a lifetime. In the case of ophthalmology, we reflected the utility derived from each stage, as diabetic retinopathy is a nonlethal disease that has a high impact on quality of life. In all cases, the development of a disease and its progression were modeled according to probabilities extracted from meta-analyses reflected in previously published peer-reviewed models. After diagnosis and treatment, the models transitioned patients to another stage, where they either remained stable or continued down the natural progression of the disease or transitioned to an absorbent state of death, tooth loss, or blindness.

When the model allowed it, we also included outcomes of choosing different treatment pathways after detecting a lesion. In all cases, model validation was performed internally by varying key parameters to check how they may be associated with results and performing univariate and multivariate sensitivity analyses. All results were then compared with available research in their fields.

Input Variables

Input variables were extracted from previous research used by the authors of the meta-analyses to construct their models. Diagnostic accuracies were also extracted from previous research. The references for the economic models and the diagnostic accuracy studies reporting on the different AI applications are summarized in **Figure 1**[4,19,22,29,33,36-54] and the Table.[4,18-22] Probabilities in prevalence rates, as well as sources, are described in eAppendices 1 through 3 in the Supplement.

Health Outcomes, Costs, and Discounting

Health outcomes were expressed as quality-adjusted life years (QALYs) for the dermatology and ophthalmology use case and the mean time a tooth was retained (in years) for the dental use case. Cost calculations from a payer perspective were built on costs estimated out-of-pocket (OOP) from a patient perspective in the US case (ie, dermatological use case), a combination of prices extracted from the public catalog of services paid by statutory insurance and a catalog of private services in the German case (dental use case), and payer perspective in the Brazilian case (ophthalmological use case).

Costs for the application of AI were charged as a fee for service. For the dental use case, an €8 fee per application had been assumed in the original publication based on direct costs for research, development, operation, and overhead. We proceeded to charge the same amount in local currency in the other cases and then performed univariate sensitivity analysis.

Costs and tooth retention–years were discounted at 3% per annum in all 3 cases and varied in a univariate sensitivity analysis between 0% and 10%.[55] Given our study's perspective, opportunity costs were not accounted for.

Statistical Analysis

We performed Monte Carlo microsimulations with 1000 independent individuals or teeth. Incremental cost-effectiveness ratios (ICERs) were used to express cost differences per QALY or mean year of tooth retention when comparing the 2 strategies. Results after performing 1000 Monte Carlo microsimulations with 1000 random samples in all 3 models can be found in the Table. To introduce parameter uncertainty, we randomly sampled transition probabilities from distributions reported in the original models and calculated 95% CIs or the range of parameters.[56] In the case of caries progression, we used uniform distributions.

Using estimates for costs (in US dollars, euros, and Brazilian real) and years for dentistry and QALY for dermatology and ophthalmology, the net benefit of each strategy combination was calculated as a mean average of each cohort using the formula: individual net benefit = WTP × change in QALYs or tooth retention–years − change in cost, where WTP indicates

the ceiling threshold of willingness to pay, ie, the additional costs a decision-maker is willing to bear for gaining an additional QALY or tooth retention–year.[32] If WTP was greater than change in cost divided by the change in QALYs or tooth retention–years, an alternative intervention was considered more cost-effective than the comparator despite possibly being more costly.[56] We used the net-benefit approach to calculate the probability of each intervention being acceptable regarding its cost-effectiveness for payers with different WTP ceiling thresholds. One-way sensitivity analyses were additionally performed to assess which strategy is associated with lowest cost or greatest increase in QALYs or tooth retention–years if key input parameters were changed to extreme values, thus exploring the impact of uncertainty and heterogeneity. Euros and reales were converted using 2020 Organisation for Economic Cooperation and Development purchasing power parities (PPP)[57] at €0.746 and R $2.362 per US $1, respectively. Significant results were determined using 95% CIs with percentiles 2.5% and 97.5%. All analyses were undertaken using R2 Healthcare version 2.1 (TreeAge).

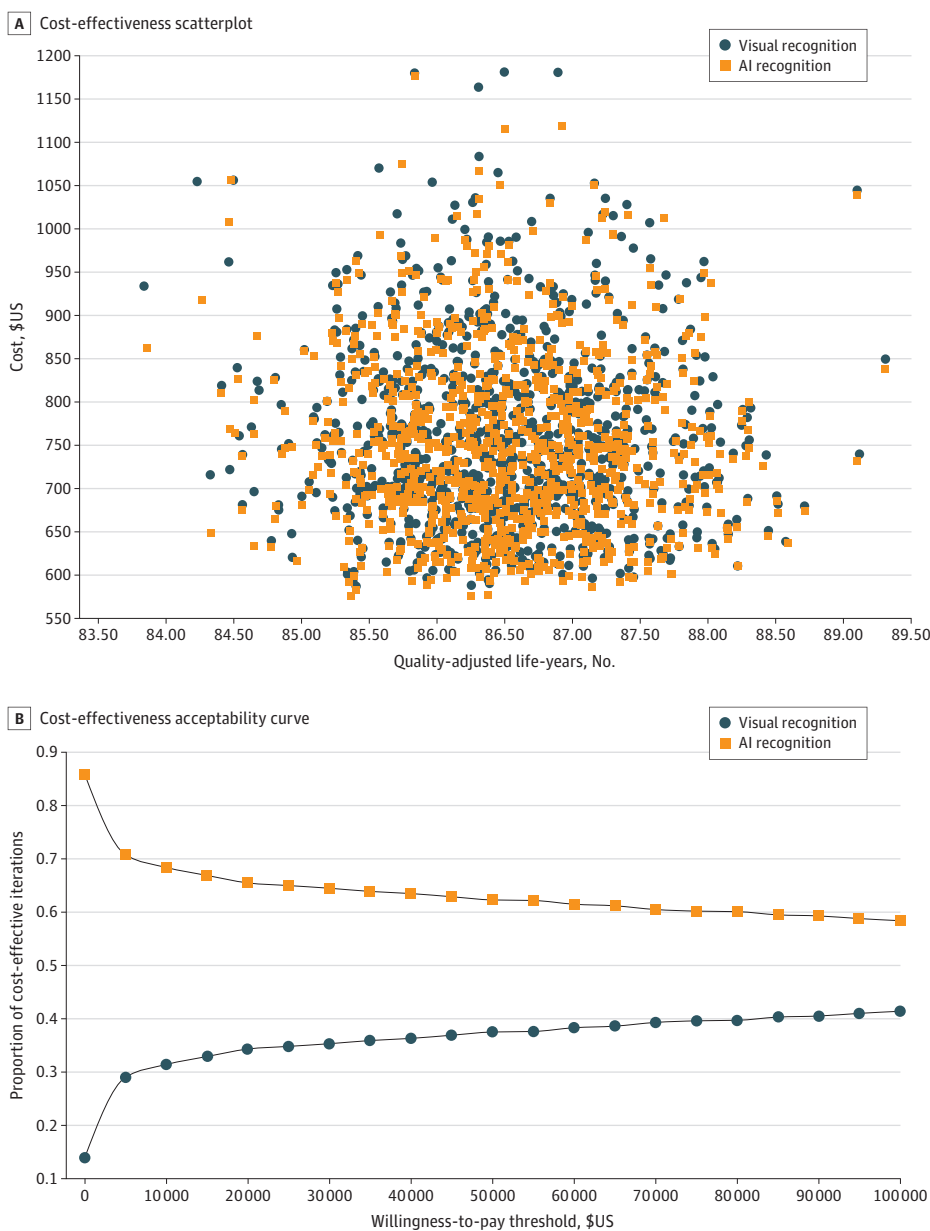Figure 1. Visual Summary of the Different Models Included in the Study
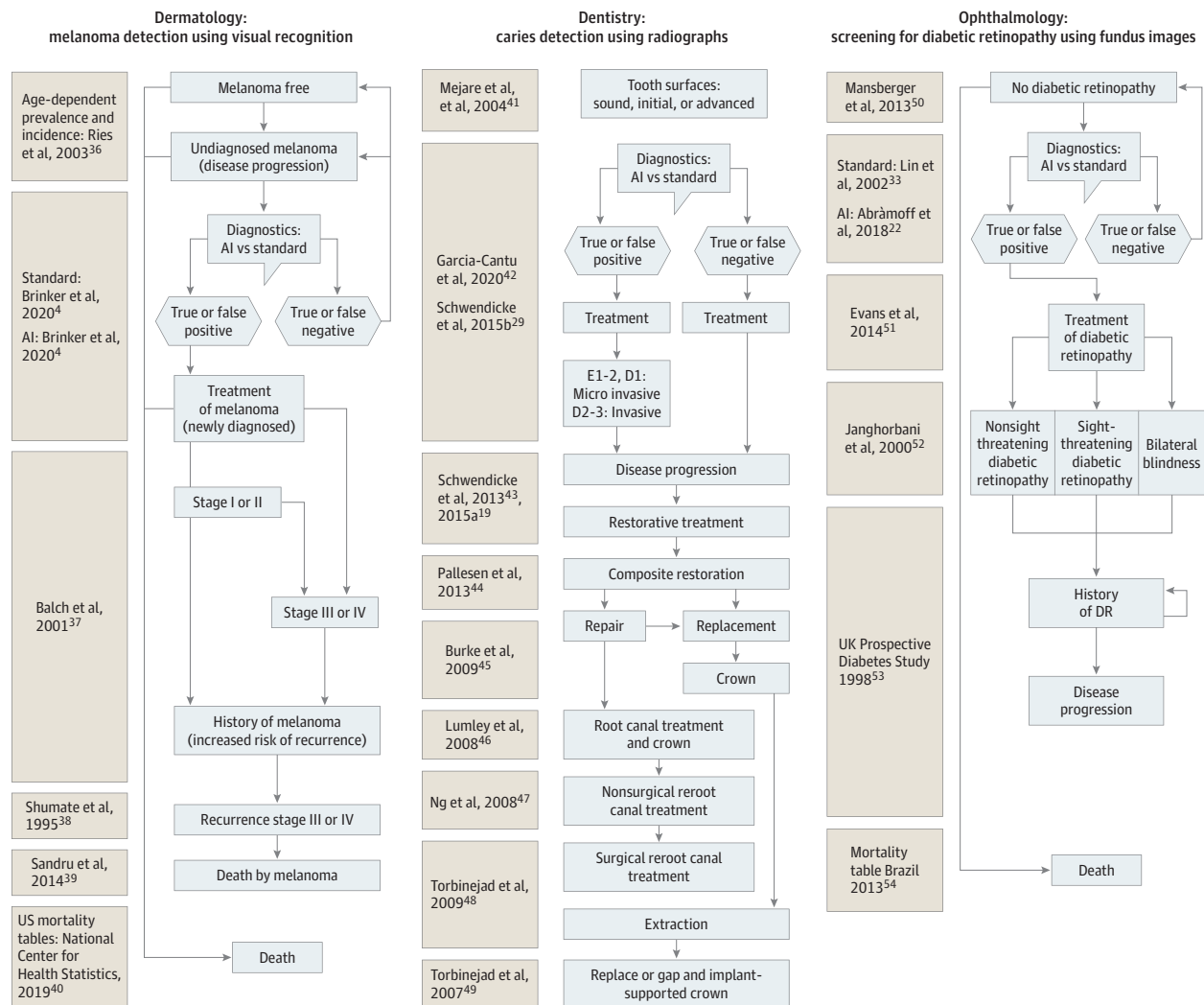
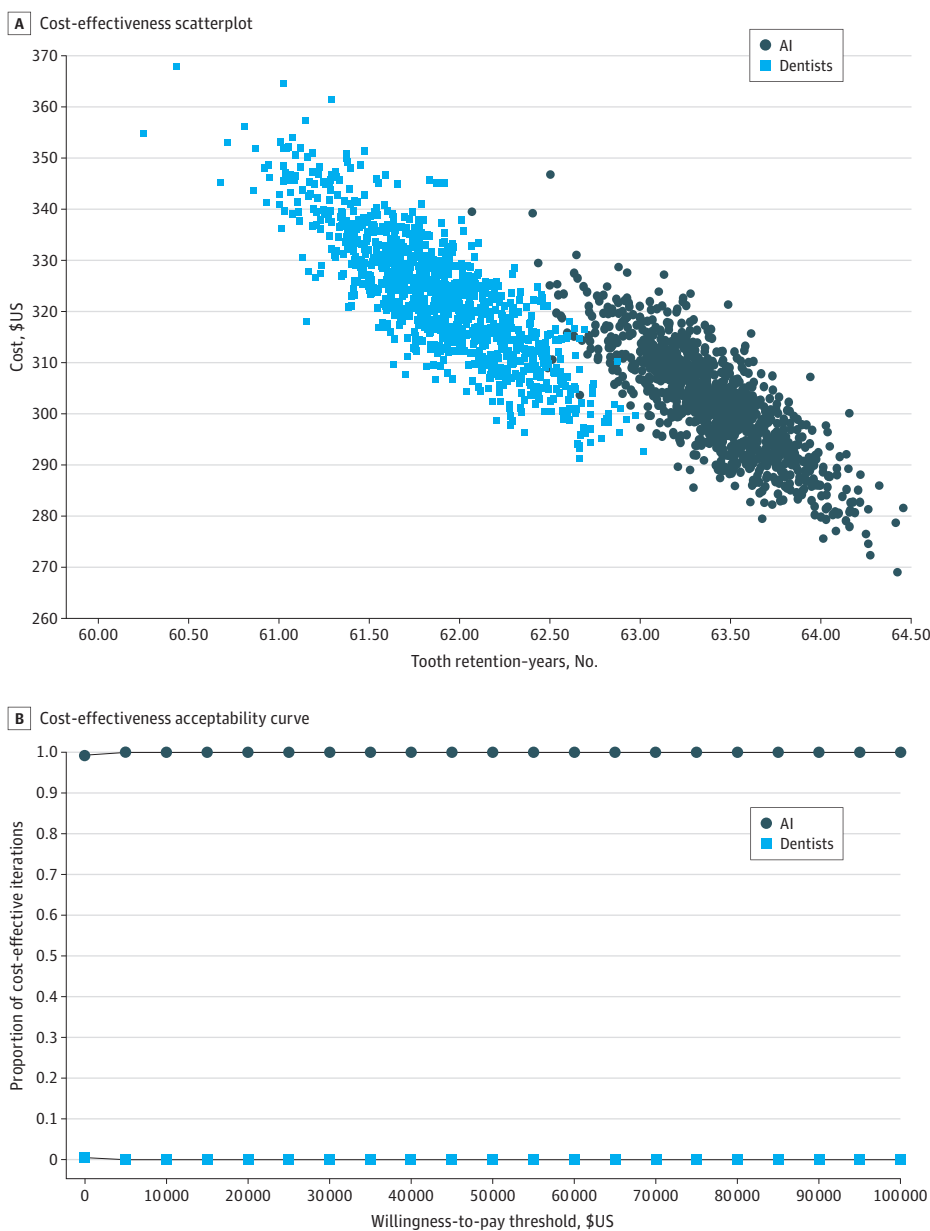AI indicates artificial intelligence; DR, diabetic retinopathy.

Results

In dermatology, the mean costs were $750 (95% CI, $608-$970) for AI and $759 (95% CI, $618-$980) for dermatologists without AI with similar health outcomes (AI, 86.6 QALYs; 95% CI, 84.9-88.0 QALYs; standard visual recognition, 86.6 QALYs; 95% CI, 84.9-88.0 QALYs). The ICER was −$27 580 per QALY (**Figure 2**A). The acceptability curve (Figure 2B) showed that AI was more likely to be more cost-effective at lower WTP; increasing WTP progressively increased the uncertainty (Figure 2B, **Figure 3**B, **Figure 4**B). Univariate sensitivity analysis on the discounting rates between 0% and 10% did not significantly affect results (eAppendix 4 in the Supplement). Univariate sensitivity analysis on the fee paid for the use of AI demonstrated that AI became the dominated strategy when the fee-for-service exceeded $16.

Figure 2. Cost-effectiveness of AI vs Standard of Care in Dermatology

A Cost-effectiveness scatterplot

B Cost-effectiveness acceptability curve

AI indicates artificial intelligence. In panel A, each dot and square represents an individual's lifetime costs accrued (in US$) when receiving either standard of care (ie, visual recognition) or AI-assisted screening. In panel B, although AI is more likely to be cost-effective at a lower willingness to pay (WTP), these results show high sensitivity to WTP.

In dentistry, AI was associated with increased tooth retention (mean tooth retention, 62.4 years; 95% CI, 61.6-63.1 years) and less costly (€320; 95% CI, €299-€341) (US $429; 95% CI, $400-$458) than caries lesion detection without AI (mean tooth retention, 60.9 years; 95% CI, 61.5-63.1 years; cost, €342.24; €318-€368). The ICER was −€15.01 per year (US $20.12) (Figure 3A). The results were very sensitive to the treatment path modeled after diagnosis; when an invasive approach for detected lesions was considered, AI was associated with fewer years of tooth retention and higher cost. The acceptability curve shows that AI was more likely to be more cost-effective independent of the cost-effectiveness studied (Figure 3B). Univariate sensitivity analysis on discounting rates between 0% and 10% showed a dominance of AI over standard diagnostic methods when discounted rates remained below 6% (eAppendix 4 in the Supplement). Univariate sensitivity analysis on the fee paid for the use of AI demonstrated that AI became the dominated strategy when fee-for-service costs were above €16 (US $21.44).

Figure 3. Cost-effectiveness of AI vs Standard of Care in Dentistry

A Cost-effectiveness scatterplot

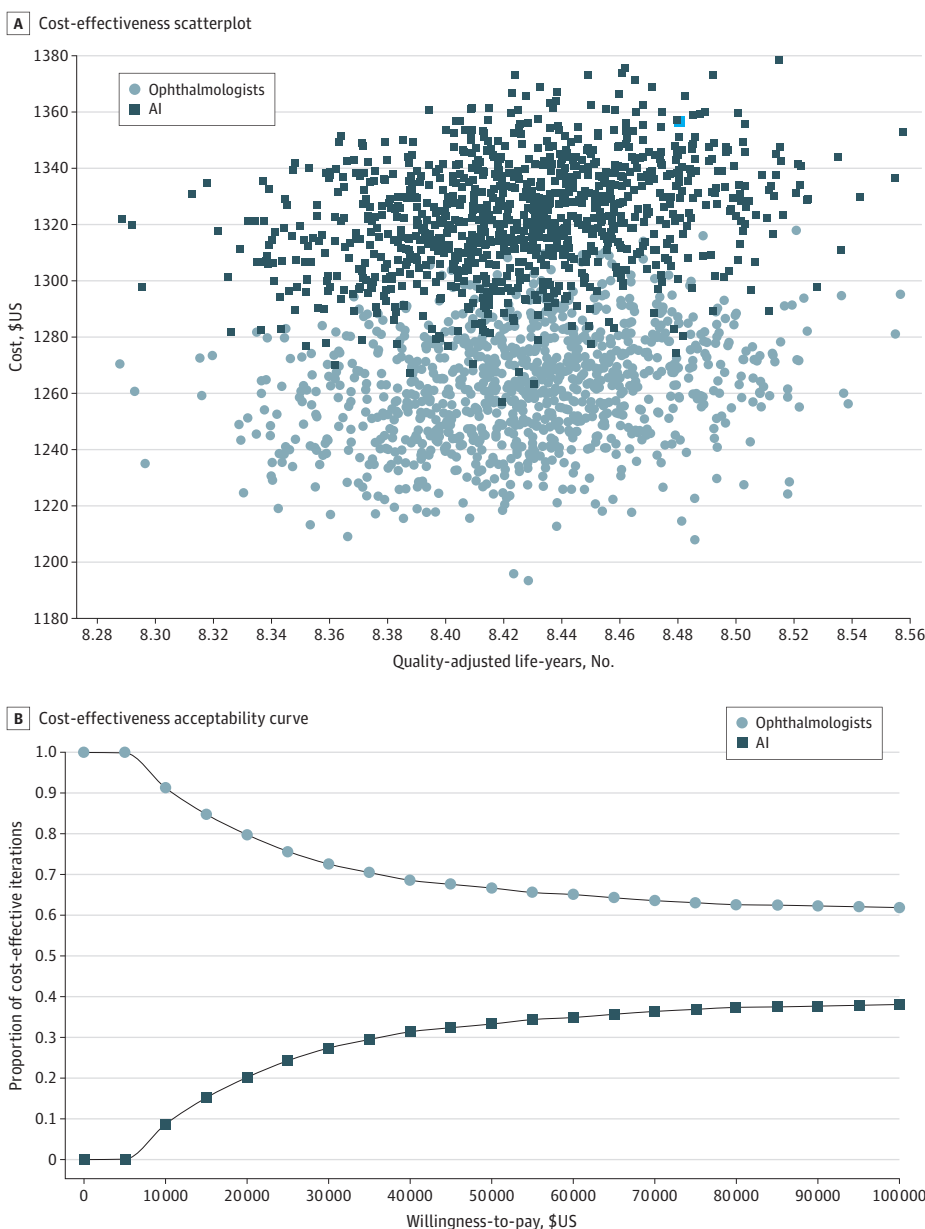B Cost-effectiveness acceptability curve

AI indicates artificial intelligence. In panel A, each dot and square represents a single tooth's lifetime costs accrued (in euros) after receiving either standard of care diagnostics or AI-assisted screening. In panel B, AI is more likely to be cost-effective at a lower willingness to pay, yet these results do not seem to be altered when one assumes higher willingness to pay (WTP).

In ophthalmology, the mean cost was R $1321 (95% CI, R $1283-R $1364) (US $559; 95% CI, US $543-$577) for AI and R $1260 (95% CI, R $1222-R $1303;) (US $533; 95% CI, $517-$551) for diagnosis without AI. Both strategies yielded a very similar mean (SD) utility of 8.4 (0.04) QALYs; however, AI increased costs by R $61 (US $25.82). The ICER was R −$91 760 (US −$38 848) (Figure 4A). The acceptability curve showed that standard of care was more likely to be more cost-effective, although higher WTP increased the uncertainty about the optimal strategy (Figure 4B).

Our results indicate that the incremental (per-person) cost per QALY would be R $39 705 (US $16 809); for reference, Brazilian GDP per capita PPP in 2020 was R $14 563 (US $6165). According to the thresholds recommended by the World Health Organization (WHO),[58] the maximum cost paid per QALY gained could be up to 3 times the GDP per capita (in our example, R $43 689 [US $18 496]) to be considered cost-effective in these settings. The dominance of standard of care was not affected

Figure 4. Cost-effectiveness of AI vs Standard of Care in Ophthalmology

AI indicates artificial intelligence. In panel A, each dot and square represents a single individual's lifetime costs accrued (Brazilian reales, R$) after receiving either standard of care diagnostics or AI-assisted screening. In panel B, AI was more likely to be less cost-effective at a lower willingness to pay based on study assumptions, although this certainty was sensitive to the willingness to pay (WTP) for additional quality-adjusted life years (QALYs).

by a sensitivity analysis on the discounting rates (eAppendix 4 in the Supplement) nor by the price charged for the use of AI support (eAppendix 5 in the Supplement).

Discussion

The cost-effectiveness of AI has been broadly studied and discussed for its potential to improve diagnosis,[14,59] facilitate screening,[10,60] and optimize laboratory tests and surgical appointments,[61,62] among other use-cases.[63-66] Our findings corroborate calls for solid economic evaluations of AI for health applications when AI is used to help determine care options for patients.[67]

To the best of our knowledge, this is the first study modeling several AI solutions against the standard of care. The main strength of this study was its design, which allowed comparisons of the same use case for the same technology used to detect different diseases. Our results suggest that the cost-effectiveness of AI vs standard of care should be evaluated specifically for each setting and use case, not only to consider the underlying costs generated by the AI application itself but also the treatments following diagnosis.

All AI solutions used as decision-support systems showed only moderate cost-effectiveness improvement. It can be assumed that if further improvements in AI are to be expected, its cost-effectiveness may improve too, as the accuracy of practitioner diagnosis without AI support is unlikely to increase. Moreover, regulation around AI, incentives for following AI recommendations, or differences in the efficiency and the diagnostic process when using AI or not should be explored further to come to a more realistic picture about the cost-effectiveness of AI in diagnostic support systems. Our results further indicate that AI may not necessarily have its biggest benefit in the hands of medical experts (where its advantages are limited) but could facilitate screening of patients in nonspecialist settings to allow targeted referral, as has been suggested in ophthalmology, for example.[59] Evaluating these differences would require building new models and methods of evaluation, where higher magnitudes of effect may be expected.

The models included in our analysis were sensitive to the fee paid for the AI and only moderately affected by discounting rates. Our study suggests that small changes in the price can alter the dominance between strategies in this use case, making the economic impact of these digital tools sensitive to aspects of implementation, settings, payers perspectives, and use cases assumed. More research on different payment methods for AI will be necessary to allow robust comparisons and draw definitive conclusions on the health economic outcomes associated with AI technology as well as to determine the role AI could play in improving value-based care.

Limitations

This study had several limitations. First, the limited information available on the research, operation and overhead costs, and payment mechanisms involved in incorporating AI did not allow for generating detailed comparisons. Aspects such as costs related to the hardware necessary for data acquisition were unknown and could potentially drastically alter our results. This uncertainty complicates establishing optimal pricing for AI services from a third-party payer perspective and is deserving of further scientific analysis. Regulations around subsequent treatment steps will also heavily affect overall cost-effectiveness and should be reflected in models. Regulators and decision-makers play an important role in making sure that developed AI solutions remain safe for patients and help to improve outcomes, while also sufficiently incentivizing further development so that digital health can accomplish some of the expectations it has generated.[57,58] Analyzing real-world evidence after improvements in diagnostic technology enter the market seems a judicious approach to prioritize patient and clinical cost-effectiveness, and can clarify how improvements in diagnostic accuracy can impact the cost-effectiveness of AI. Future studies could consider the expected value of information analysis to assess the relevance of uncertainty of a range of parameters, including diagnostic accuracy, and steer research and development accordingly.

Second, it is important to recognize that differences in outcomes across our models could be due to inconsistencies in the use of AI between different income settings. Epidemiological factors and lower fee-for-services paid in low- and middle-income countries should be studied to avoid that AI does not worsen existing health inequalities. This fact calls for a better understanding of how epidemiological differences such as incidence and morbidity of a certain disease can factor into decisions to reimburse AI services. Because of that, future research could focus on developing analytical frameworks to facilitate comparisons of AI from different perspectives, in different settings, and for different outcomes. This could allow for more targeted development of AI solutions for use cases where they are most impactful and cost-effective.

Third, we assumed physicians would act according to the AI-detection results, ie, in perfect congruence. However, this is not a given—physicians may disagree with AI diagnoses and make decisions that alter the resulting diagnostic accuracy (both to the benefit or detriment of the resulting composite accuracy). The same applies for the resulting therapies. We therefore invite readers to consider our results as a base case scenario, as in practice deviations from our findings are likely. New studies assessing how physicians interact with software would be fundamental for understanding how AI could best synergize with medical practitioners.

Conclusions

In this economic evaluation, AI used as a decision-support system came with limited and use case–specific cost-effectiveness advantages, which were sensitive not only to the costs assigned to AI but also the subsequent therapy paths assumed after the diagnosis. AI developers need to work jointly with regulators and the medical community to make sure that new AI solutions are deployed where they best improve outcomes. Developing appropriate payment mechanisms seems fundamental to incentivize new cost-effective therapies with this technology.

Corresponding Author: Jesus Gomez Rossi, DMD, MSc, Department of Oral Diagnostics, Digital Health and Health Services Research, Charité–Universitätsmedizin Berlin, Aßmannshauser Str 4-6, 14197 Berlin, Germany (jesus.gomez-rossi@charite.de).

Author Affiliations: Department of Oral Diagnostics, Digital Health and Health Services Research, Charité–Universitätsmedizin Berlin, Berlin, Germany (Gomez Rossi, Krois, Schwendicke); Department of Economics, Freie Universität Berlin, Germany (Rojas-Perilla); Department of Analytics in the Digital Era, United Arab Emirates University, Al Ain, United Arab Emirates (Rojas-Perilla).

REFERENCES

1. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395(10236):1579-1586. doi:10.1016/S0140-6736(20)30226-9

2. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230-243. doi:10.1136/svn-2017-000101

3. Montani S, Striani M. Artificial intelligence in clinical decision support: a focused literature survey. *Yearb Med Inform*. 2019;28(1):120-127. doi:10.1055/s-0039-1677911

4. Brinker TJ, Hekler A, Enk AH, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer*. 2019;111:148-154. doi:10.1016/j.ejca.2019.02.005

5. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020;99(7):769-774. doi:10.1177/0022034520915714

6. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol. 2019;137(12):1353–1360. doi:10.1001/jamaophthalmol.2019.3501

7. Hand S. IDx-DR Becomes First FDA-Approved AI-Based Diagnostic for Diabetic Retinopathy. *Xtalks*. Published August 30, 2018. Accessed October 24, 2019. https://xtalks.com/idx-dr-becomes-first-fda-approved-ai-based-diagnostic-for-diabetic-retinopathy-1274/

8. Royle P, Mistry H, Auguste P, Shyangdan D, Freeman K, Lois N, et al. Model for assessing cost-effectiveness of pan-retinal photocoagulation for non-proliferative diabetic retinopathy. In: Pan-retinal photocoagulation and other forms of laser treatment and drug therapies for non-proliferative diabetic retinopathy: systematic review and economic evaluation. NIHR Journals Library database. Published July 2015. Accessed April 1, 2021. https://www.ncbi.nlm.nih.gov/books/NBK305085/

9. Crabb DP, Russell RA, Malik R, Anand N, Baker H, Boodhna T, et al. Health economic modelling of different monitoring intervals in glaucoma patients. In: Frequency of visual field testing when monitoring patients newly diagnosed with glaucoma: mixed methods and modelling. NIHR Journals Library database. Published August 2014. Accessed July 24, 2020. https://www.ncbi.nlm.nih.gov/books/NBK259977/

10. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye (Lond)*. 2020;34(3):451-460. doi:10.1038/s41433-019-0566-0

11. Xie Y, Nguyen QD, Hamzah H, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*. 2020;2(5):e240-e249. doi:10.1016/S2589-7500(20)30060-1

12. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health*. 2018;3(4):e000798. doi:10.1136/bmjgh-2018-000798

13. Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, Roy D, Fleet R, Ahmed M, Fortin JP. Artificial intelligence and health technology assessment: anticipating a new level of complexity. *J Med Internet Res*. 2020;22(7):e17707. doi:10.2196/17707

14. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. *J Med Internet Res*. 2020;22(10):e17298. doi:10.2196/17298

15. Dismuke C. Progress in examining cost-effectiveness of AI in diabetic retinopathy screening. *The Lancet Digital Health*. 2020;2(5):E212-E213. doi:10.1016/S2589-7500(20)30077-7

16. Schwendicke F, Rossi JG, Göstemeyer G, Elhennawy K, Cantu AG, Gaudin R, Chaurasia A, Gehrung S, Krois J. Cost-effectiveness of artificial intelligence for proximal caries detection. *J Dent Res*. 2021;100(4):369-376. doi:10.1177/0022034520972335

17. Olvera-Barrios A, Heeren TF, Balaskas K, et al. Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol*. 2021;105(2):265-270. doi:10.1136/bjophthalmol-2019-315394

18. Losina E, Walensky RP, Geller A, et al. Visual screening for malignant melanoma: a cost-effectiveness analysis. *Arch Dermatol*. 2007;143(1):21-28. doi:10.1001/archderm.143.1.21

19. Schwendicke F, Stolpe M, Meyer-Lueckel H, Paris S. Detecting and treating occlusal caries lesions: a cost-effectiveness analysis. *J Dent Res*. 2015;94(2):272-280. doi:10.1177/0022034514561260

20. Ben ÂJ, Neyeloff JL, de Souza CF, et al. Cost-utility analysis of opportunistic and systematic diabetic retinopathy screening strategies from the perspective of the Brazilian public healthcare system. *Appl Health Econ Health Policy*. 2020;18(1):57-68. doi:10.1007/s40258-019-00528-w

21. Cantu AG, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*. 2020;100:103425. doi:10.1016/j.jdent.2020.103425

22. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1(1):39. doi:10.1038/s41746-018-0040-6

23. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ*. 2016;47:20-33. doi:10.1016/j.jhealeco.2016.01.012

24. Organisation for Economic Cooperation and Development. Health at a Glance 2019—OECD Indicators. Published November 7, 2019. Accessed February 4, 2021. https://www.oecd.org/health/health-systems/Health-at-a-Glance-2019-Chartset.pdf

25. Ridic G, Gleason S, Ridic O. Comparisons of health care systems in the United States, Germany and Canada. *Mater Sociomed*. 2012;24(2):112-120. doi:10.5455/msm.2012.24.112-120

26. Institute for Quality and Efficiency in Health Care. Health care in Germany: the German health care system. InformedHealth.org. Updated February 8, 2018. Accessed June 14, 2021. https://www.ncbi.nlm.nih.gov/books/NBK298834/

27. Massuda A, Hone T, Leles FAG, de Castro MC, Atun R. The Brazilian health system at crossroads: progress, crisis and resilience. *BMJ Glob Health*. 2018;3(4):e000829. doi:10.1136/bmjgh-2018-000829

28. Organisation for Economic Cooperation and Development. OECD Reviews of Health Systems: Brazil 2021. Published December 14, 2021. Accessed February 4, 2022. https://www.oecd.org/health/oecd-reviews-of-health-systems-brazil-2021-146d0dea-en.htm

29. Schwendicke F, Paris S, Stolpe M. Detection and treatment of proximal caries lesions: milieu-specific cost-effectiveness analysis. *J Dent*. 2015;43(6):647-655. doi:10.1016/j.jdent.2015.03.009

30. Grin CM, Kopf AW, Welkovich B, Bart RS, Levenstein MJ. Accuracy in the clinical diagnosis of malignant melanoma. *Arch Dermatol*. 1990;126(6):763-766. doi:10.1001/archderm.1990.01670300063008

31. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res*. 2020;22(2):e16866. doi:10.2196/16866

32. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; 2005.

33. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134(2):204-213. doi:10.1016/s0002-9394(02)01522-2

34. Niemeijer M, Abramoff MD, van Ginneken B. Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Trans Med Imaging*. 2009;28(5):775-785. doi:10.1109/TMI.2008.2012029

35. Beaser RS, Turell WA, Howson A. Strategies to improve prevention and management in diabetic retinopathy: qualitative insights from a mixed-methods study. *Diabetes Spectr*. 2018;31(1):65-74. doi:10.2337/ds16-0043

36. Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Edwards BK, eds. SEER Cancer Statistics Review, 1973-1999. National Cancer Institute. 2003. Accessed September 10, 2021. https://seer.cancer.gov/archive/csr/1973_1999/

37. Balch CM, Buzaid AC, Soong SJ, et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol*. 2001;19(16):3635-3648. doi:10.1200/JCO.2001.19.16.3635

38. Shumate CR, Urist MM, Maddox WA. Melanoma recurrence surveillance: patient or physician based? *Ann Surg*. 1995;221(5):566-569. doi:10.1097/00000658-199505000-00014

39. Sandru A, Voinea S, Panaitescu E, Blidaru A. Survival rates of patients with metastatic malignant melanoma. *J Med Life*. 2014;7(4):572-576.

40. National Center for Health Statistics. National Vital Statistics System—Mortality Tables. Updated December 16, 2016. Accessed February 25, 2019. https://www.cdc.gov/nchs/nvss/mortality_tables.htm

41. Mejàre I, Stenlund H, Zelezny-Holmlund C. Caries incidence and lesion progression from adolescence to young adulthood: a prospective 15-year cohort study in Sweden. *Caries Res*. 2004;38(2):130–141. doi:10.1159/000075937

42. Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*. 2020;100:103425. doi:10.1016/j.jdent.2020.103425

43. Schwendicke F, Stolpe M, Meyer-Lueckel H, Paris S, Dörfer CE. Cost-effectiveness of one- and two-step incomplete and complete excavations. *J Dent Res*. 2013;92(10):880-887. doi:10.1177/0022034513500792

44. Pallesen U, van Dijken JWV, Halken J, Hallonsten A-L, Höigaard R. Longevity of posterior resin composite restorations in permanent teeth in Public Dental Health Service: a prospective 8 years follow up. *J Dent*. 2013;41 (4):297–306. doi:10.1016/j.jdent.2012.11.021

45. Burke FJT, Lucarotti PSK. Ten-year outcome of crowns placed within the General Dental Services in England and Wales. *J Dent*. 2009;37(1):12-24. doi:10.1016/j.jdent.2008.03.017

46. Lumley PJ, Lucarotti PSK, Burke FJT. Ten-year outcome of root fillings in the General Dental Services in England and Wales. *Int Endod J*. 2008;41(7):577-585. doi:10.1111/j.1365-2591.2008.01402.x

47. Ng Y-L, Mann V, Gulabivala K. Outcome of secondary root canal treatment: a systematic review of the literature. *Int Endod J*. 2008;41(12):1026-1046. doi:10.1111/j.1365-2591.2008.01484.x

48. Torabinejad M, Corr R, Handysides R, Shabahang S. Outcomes of nonsurgical retreatment and endodontic surgery: a systematic review. *J Endod*. 2009;35(7):930-937. doi:10.1016/j.joen.2009.04.023

49. Torabinejad M, Anderson P, Bader J, et al. Outcomes of root canal treatment and restoration, implant-supported single crowns, fixed partial dentures, and extraction without replacement: a systematic review. *J Prosthet Dent*. 2007;98(4):285-311. doi:10.1016/S0022-3913(07)60102-4

50. Mansberger SL, Gleitsmann K, Gardiner S, et al. Comparing the effectiveness of telemedicine and traditional surveillance in providing diabetic retinopathy screening examinations: a randomized controlled trial. *Telemed J E Health*. 2013;19(12):942-948. doi:10.1089/tmj.2012.0313

51. Evans JR, Michelessi M, Virgili G. Laser photocoagulation for proliferative diabetic retinopathy. *Cochrane Database Syst Rev*. 2014;11:CD011234. doi:10.1002/14651858.CD011234.pub2

52. Janghorbani M, Jones RB, Allison SP. Incidence of and risk factors for proliferative retinopathy and its association with blindness among diabetes clinic attenders. *Ophthalmic Epidemiol*. 2000;7(4):225-241. doi:10.1076/opep.7.4.225.4171

53. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352(9131):837-853. doi:10.1016/S0140-6736(98)07019-6

54. Brazilian Institute of Geography and Statistics. Tábua completa de mortalidade para o Brasil—2013. Published 2014. Accessed April 28, 2021. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=73097

55. Institute for Quality and Efficiency in Health Care. Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung—Version 2.0. Updated March 16, 2009. Accessed February 4, 2022. https://www.iqwig.de/methoden/09-03-18_entwurf_methoden_kosten-nutzen-bewert.pdf

56. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making*. 2002;22(4):290-308. doi:10.1177/027298902400448867

57. Organisation for Economic Cooperation and Development. OECD Data—Purchasing power parities (PPP), 2000-2020. OECD.org. Accessed April 29, 2021. https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

58. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost–effectiveness of interventions: alternative approaches. *Bull World Health Organ*. 2015;93(2):118-124. doi:10.2471/BLT.14.138206

59. Hwang J-J, Jung Y-H, Cho B-H, Heo M-S. An overview of deep learning in the field of dentistry. *Imaging Sci Dent*. 2019;49(1):1-7. doi:10.5624/isd.2019.49.1.1

60. Wolf RM, Channa R, Abramoff MD, Lehmann HP. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol*. 2020;138(10):1063-1069. doi:10.1001/jamaophthalmol.2020.3190

61. Gönel A. Clinical biochemistry test eliminator providing cost-effectiveness with five algorithms. *Acta Clin Belg*. 2020;75(2):123-127. doi:10.1080/17843286.2018.1563324

62. Rozario N, Rozario D. Can machine learning optimize the efficiency of the operating room in the era of COVID-19? *Can J Surg*. 2020;63(6):E527-E529. doi:10.1503/cjs.016520

63. Grover D, Bauhoff S, Friedman J. Using supervised learning to select audit targets in performance-based financing in health: an example from Zambia. *PLoS One*. 2019;14(1):e0211262. doi:10.1371/journal.pone.0211262

64. Lee I, Monahan S, Serban N, Griffin PM, Tomar SL. Estimating the cost savings of preventive dental services delivered to Medicaid-enrolled children in six southeastern states. *Health Serv Res*. 2018;53(5):3592-3616. doi:10.1111/1475-6773.12811

65. Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak*. 2018;18(1):44. doi:10.1186/s12911-018-0620-z

66. US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. FDA news release. Published April 11, 2018. Accessed July 27, 2020. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye

67. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res*. 2020;22(2):e16866. doi:10.2196/16866

SUPPLEMENT.
eTable 1. Input Parameters—Dermatology
eTable 2. Input Parameters—Dentistry
eTable 3. Input Parameters—Ophthalmology
eTable 4. Sensitivity Analysis—Discounting Rate
eTable 5. Sensitivity Analysis—Cost of AI
eReferences.