Cost-Utility Analysis of Deep Learning and Trained Human Graders for Diabetic Retinopathy Screening in a Nationwide Program

Attasit Srisubat · Kankamon Kittrongsiri · Sermsiri Sangroongruangsri ·
Chalida Khemvaranan · Jacqueline Baras Shreibati · Jack Ching ·
John Hernandez · Richa Tiwari · Fred Hersch · Yun Liu ·
Prut Hanutsaha · Varis Ruamviboonsuk · Saowalak Turongkaravee ·
Rajiv Raman · Paisan Ruamviboonsuk [ORCID]

ABSTRACT

Introduction: Deep learning (DL) for screening diabetic retinopathy (DR) has the potential to address limited healthcare resources by enabling expanded access to healthcare. However, there is still limited health economic evaluation, particularly in low- and middle-income countries, on this subject to aid decision-making for DL adoption.

Methods: In the context of a middle-income country (MIC), using Thailand as a model, we constructed a decision tree-Markov hybrid model to estimate lifetime costs and outcomes of Thailand's national DR screening program via DL and trained human graders (HG). We calculated the incremental cost-effectiveness ratio (ICER) between the two strategies. Sensitivity analyses were performed to probe the influence of modeling parameters.

A. Srisubat
Department of Medical Services, Ministry of Public Health, Nonthaburi, Thailand

K. Kittrongsiri
Social, Economic and Administrative Pharmacy (SEAP) Graduate Program, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand

S. Sangroongruangsri (✉) · S. Turongkaravee
Social and Administrative Pharmacy Division, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand
e-mail: ningsermsiri@gmail.com

C. Khemvaranan
Department of Research and Technology Assessment, Lerdsin Hospital, Bangkok, Thailand

J. B. Shreibati · J. Ching · J. Hernandez · R. Tiwari · F. Hersch · Y. Liu
Google LLC, Mountain View, CA, USA

P. Hanutsaha
Department of Ophthalmology, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

V. Ruamviboonsuk
Department of Biochemistry, Chulalongkorn University, Bangkok, Thailand

R. Raman
Sri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, Chennai, Tamil Nadu, India

P. Ruamviboonsuk (✉)
Department of Ophthalmology, College of Medicine, Rajavithi Hospital, Rangsit University, Bangkok, Thailand
e-mail: paisan.trs@gmail.com

Results: From a societal perspective, screening with DL was associated with a reduction in costs of ∼ US\$ 2.70, similar quality-adjusted life-years (QALY) of + 0.0043, and an incremental net monetary benefit of ∼ US\$ 24.10 in the base case. In sensitivity analysis, DL remained cost-effective even with a price increase from US\$ 1.00 to US\$ 4.00 per patient at a Thai willingness-to-pay threshold of ∼ US\$ 4.997 per QALY gained. When further incorporating recent findings suggesting improved compliance to treatment referral with DL, our analysis models effectiveness benefits of ∼ US\$ 20 to US\$ 50 depending on compliance.

Conclusion: DR screening using DL in an MIC using Thailand as a model may result in societal cost-savings and similar health outcomes compared with HG. This study may provide an economic rationale to expand DL-based DR screening in MICs as an alternative solution for limited availability of skilled human resources for primary screening, particularly in MICs with similar prevalence of diabetes and low compliance to referrals for treatment.

Keywords: Cost-utility analysis; Health economics; Diabetic retinopathy; Artificial intelligence; Public health

---

Key Summary Points

Why carry out this study?

A limited number of studies on economic evaluation of deep learning (DL) for diabetic retinopathy (DR) screening in high-income countries have demonstrated the potential for cost-saving with DL from a healthcare provider perspective. However, very few studies evaluated DL for DR screening in a middle-income country (MIC).

This study was conducted to analyze the cost-utility of DL compared to trained non-physician human graders (HG) to screen DR over a lifetime horizon of patients from both societal and healthcare provider perspectives.

What was learned from this study?

Our cost-utility analysis showed that in the context of MICs, such as Thailand, screening for DR using DL, compared to HG, may cause a high incremental cost-effectiveness ratio (ICER) over a lifetime horizon of patients from a healthcare provider perspective. This was due to much higher sensitivity of DL compared to that of HG and led to a higher treatment cost. However, screening using DL caused less bilateral blindness, which in turn caused more cost-saving than HG from a societal perspective.

Our data provide an economic rationale for decision makers to expand DL-based DR screening in MICs with similar prevalence of diabetes, shortages of skilled human resources for primary screening, and low compliance to referrals for treatment. In general, DR screening using DL or HG could achieve greater cost-savings with higher compliance to referrals for treatment.

Policy makers should be aware of budget impact of treating more patients with sight-threatening DR (STDR) with clinical deployment of DL.

INTRODUCTION

Deep learning (DL), a type of artificial intelligence (AI), has been retrospectively and prospectively validated as an effective tool for screening of diabetic retinopathy (DR), one of the leading causes of blindness worldwide, with levels of performance on par with expert ophthalmologists [1–3].

To build further evidence toward real-world deployment of DL for DR, rigorous health economic evaluations are essential to ensure appropriate allocation of healthcare budgets. While a few economic evaluations conducted in high-income countries found a potential for DL to be cost-effective for DR screening, few studies have been conducted for resource-limited settings, such as middle-income countries [4].

Thailand, a middle-income country (MIC) with a population of 70 million people and a gross domestic product of US\$ 7189.04 per capita in 2020, is one of the few countries where a nationwide screening program for DR has been well established. The program, supported

by the Ministry of Public Health, has been conducted across all provinces and administered by provincial health officers. Most of the approximately 700 screening units throughout the country are at the primary care level and use grading of color fundus photographs (CFP) for identification of referrals to ophthalmologists. Given the shortage of ophthalmologists, the grading is performed by trained, non-physician healthcare personnel (HG).

To reduce demands on limited human resources and expand access to screening in the country, DL approaches have been evaluated [5]. We previously conducted a retrospective validation study of a DL model for detecting referable DR in Thai patients with diabetes. Using a dataset of > 10,000 CFPs from those in the Thai diabetes registry, we found the model to have a high sensitivity (0.97) and specificity (0.96) for the detection of referrable DR [6]. Following this, a prospective interventional study in nine primary health centers validated the DL model's feasibility for real-time DR screening and confirmed the high sensitivity (0.91) and specificity (0.95) [7]. Human-centered evaluative research was also conducted, reporting that patients and health care professionals were receptive to deploying DL in the real world [5].

The cost burden of patients with DR without screening to detect referrals for timely intervention was demonstrated in a recent study on cost-effectiveness analysis of DR screening using DL in low-income patients at primary care centers in the US. The estimated cost incurred to patients who received routine care by ophthalmologists at referral centers without screening was 2082.91 USD per patient compared to 1596.99 USD when they received screening by DL [8].

Thus, to inform further implementation of DL for DR screening, we conducted this health economic evaluation to inform decision makers regarding the deployment of DL compared with HG in the context of MIC. The results of this study can be applied not only to Thailand but also to other MICs with similar prevalence of diabetes and limitations in resources.

METHODS

We constructed a model for cost-utility analysis (CUA) using data derived from previously published validation studies of the DL system for DR screening in Thailand [6]. All the studies were approved by the Ethics Committee of Rajavithi Hospital; all patients provided written informed consent. This economic evaluation was conducted and reported according to Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Checklist provided in Supplementary Material. This study was approved by the Institutional Review Board of Rajavithi Hospital, Thailand (COA: 087/2562) and conducted according to the Declaration of Helsinki.

Target Population

A hypothetical cohort of patients with type 2 diabetes, aged 40 years, not previously screened for DR, was modeled for this CUA. This cohort was selected because the national program predominantly screens type 2 diabetes patients > age 40 years and the model accounts for yearly screenings for non-referred patients. The patients were scheduled for annual DR screening and were referred for treatment if diagnosed with sight-threatening DR (STDR). STDR was defined as patients with either eye having diabetic macular edema (DME), severe non-proliferative DR (NPDR), or proliferative DR (PDR). These definitions are based on the International Clinical Classification of DR (ICCDR) [9, 10].

Model Structure

The CUA using a Markov model with an embedded decision tree was undertaken from both societal and provider perspectives. Costs and health outcomes were evaluated over a lifetime horizon of patients with a discount rate of

(A)

(B)
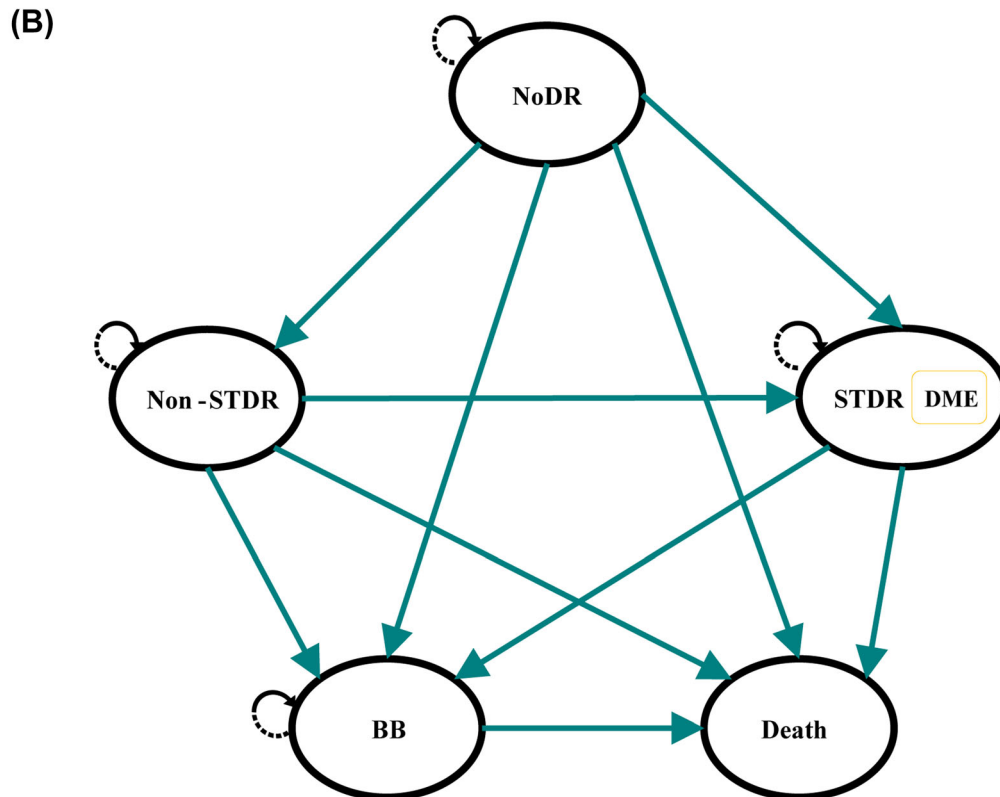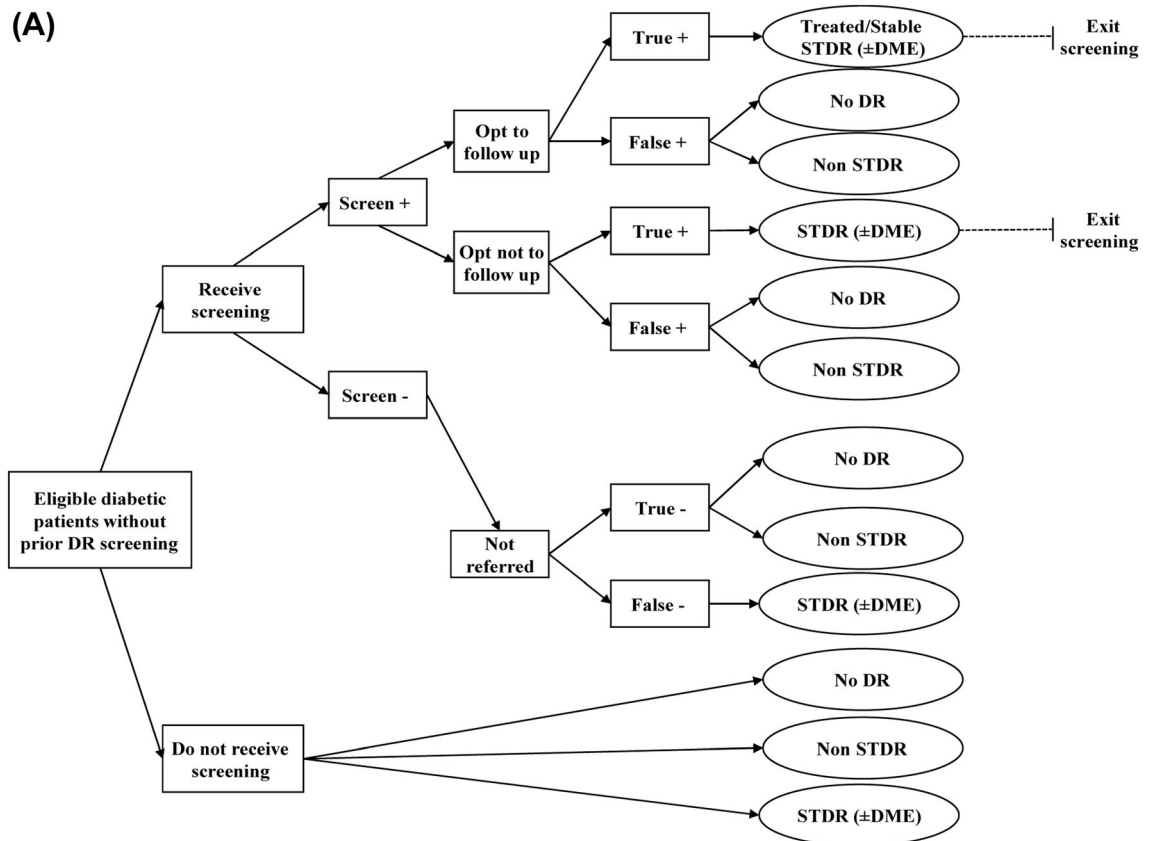
◀**Fig. 1** Structure of the cost utility analysis model. **A** Decision tree representing diabetic retinopathy screening options for patients with diabetes; **B** Markov model structure representing the clinical progression of diabetic retinopathy (Adapted from the Markov model of Ben et al. [12]). *BB* bilateral blindness, *DME* diabetic macular edema, *DR* diabetic retinopathy, *STDR* sight-threatening diabetic retinopathy. Non-STDR includes mild and moderate non-proliferative diabetic retinopathy (NPDR) without DME. Severe NPDR and proliferative diabetic retinopathy (PDR) including DME were defined as the STDR health state

3% a year for future costs and outcomes. The decision tree (Fig. 1A) illustrates the screening processes for patients. They received annual screening at primary health centers by grading of CFP by either HG or DL until they exited. An exit was defined as when a patient was graded as referral, ungradable, or with death. The ungradable patient was defined when CFP of both eyes was ungradable or when CFP of either eye was ungradable and that of the fellow eye was graded as non-STDR. This proportion of ungradable patients was found to be similar in both DL and HG arms in our previous study [11].

All patients with STDR in both arms received additional confirmatory grading by retinal specialists. Patients confirmed to be true positives were referred for treatment. Those confirmed as false positives and those with negative screening results were re-screened in the following year.

A Markov model was developed using Microsoft Excel (Microsoft Corp., Redmond, WA) comprising five health states based on the natural history of the disease and a 1-year cycle length (Fig. 1B). Transitional probabilities in this model are shown in Table 1. According to the progression rates of DR, patients could remain stable or progress towards bilateral blindness or death. A half-cycle correction was applied to account for events and transitions that might occur at any point during the cycle.

Screening Strategies

This model reflected differences in the diagnostic accuracy of DL as compared to HG. DL had higher sensitivity (0.950 for DL and 0.737 for HG) but slightly lower specificity than HG (0.980 for DL and 0.986 for HG) [11]. Another difference, not captured in this model, was that DL enabled immediate results, whereas HG results could be delayed by 1–2 weeks if primary gradings could not be completed at the point of care [5]. Given that the timing of results can lead to differences in the proportion of patients who adhere to follow-up at the referral centers for treatment [27, 28], for example, a prospective study of DR screening using DL to provide immediate results showed the adherence rate to referrals was 55.4% compared with the historical adherence rate of 18.7% at 1 year [27], a scenario analysis was performed to understand the potential impact of this differential compliance.

Model Input Parameters

Transitional Probabilities and Screening Performance

The baseline prevalence of the disease severity was obtained from our previous studies [11]. The transition probabilities were derived from the literature [11, 12]. The proportion of patients with bilateral DR was estimated at 75% in the model [29]. Since the relative risk of mortality of patients with diabetes and blindness had not yet been reported and the main causes of blindness in patients with DR were from STDR [30], we therefore applied the relative risk of mortality in patients with STDR. The compliance to follow-up treatment after true positive screening results had not yet been systematically quantified in the Thai health system. We therefore used 60% for both screening modalities at baseline based on random sampling from a few screening sites.

We used 50% as the screening uptake according to data from the Ministry of Public Health in the base case analysis. The age-specific all-cause mortality, derived from the Thai population's life table data, was used and integrated with relative risks of mortality among patients in the different health states.

Costs

Direct medical costs of the screenings included the costs of fundus cameras, CFP capture, and

Table 1 Input parameters used in the base case analysis

Parameters	Distribution	Mean	SE	Source
Transitional probability				
No DR to non-STDR	Beta	0.0524	0.0010	[11]
No DR to STDR	Beta	0.0086	0.0010	[11]
Non-STDR to STDR	Beta	0.0734	0.0010	[11]
No DR to BB	Beta	0.0016	0.0002	[13, 14]
Non-STDR to BB	Beta	0.0040	0.0004	[13, 14]
STDR to BB	Beta	0.0080	0.0006	[13, 14]
Efficacy of treatment				
Relative risk of DR progression when patients were treated	Beta	0.4900	0.0689	[15]
Prevalence				
Prevalence of diabetic patients without DR	Dirichlet	0.7236	0.7236	[11]
Prevalence of diabetic patients with non-STDR	Dirichlet	0.1537	0.1537	[11]
Prevalence of diabetic patients with STDR	Dirichlet	0.1227	0.1227	[11]
Proportion of DME among STDR	Beta	0.9247	0.0925	[11]
Relative risk of mortality in diabetic patients	Gamma	1.8500	0.1850	[16]
Relative risk of mortality in diabetic patients with non-STDR	Gamma	1.1300	0.1990	[17]
Relative risk of mortality in diabetic patients with STDR	Gamma	2.2600	0.6633	[17]
Relative risk of mortality in diabetic patients with blindness	Gamma	2.2600	0.6633	Assumed: using the risk of patients with STDR as stated in text
Efficacy of screening				
Sensitivity of screening by HG	Beta	73.72%	1.66%	[11]
Specificity of screening by HG	Beta	98.59%	0.17%	[11]
Sensitivity of screening by DL	Beta	95.03%	0.82%	[11]
Specificity of screening by DL	Beta	97.97%	0.20%	[11]
Uptake screening, HG	Beta	50.00%	5.00%	Primary data from Health Data Center (HDC) of Ministry of Public Health in the Fiscal year 2020

Table 1 continued

Parameters	Distribution	Mean	SE	Source
Compliance to follow-up confirmation after positive screening result, HG	Beta	60.00%	8.00%	Data from random sampling
Uptake screening, DL	Beta	50.00%	5.00%	Assumed: equal to HG based on model purpose
Compliance to follow-up confirmation after positive screening result, DL	Beta	60.00%	8.00%	Assumed: equal to HG based on model purpose
Direct medical cost				
Cost of screening				
Unit cost of image grading by DL per person (THB)	Gamma	32	3	Set for this study
Unit cost of image grading by HG per person (THB)*	Gamma	56.11	5.6	[18]
Unit cost of image grading (confirmation) by retina specialist per person (THB)	Gamma	112.22	11.22	[18]
Unit cost of colour fundus image capture (both eyes) (THB)	Gamma	194	19	Primary data
Unit cost of visual acuity measurement (THB)	Gamma	40	4	[19]
Unit cost of mydriasis (THB)	Gamma	73	7	[19]
Unit cost of OPD service per visit in a primary facility (THB)	Gamma	109	11	[19]
Unit cost of OPD service per visit in a tertiary facility (THB)	Gamma	309.50	30.95	[19]
Unit cost of IPD service per day of admission (THB)	Gamma	1329	133	[19]
Cost of treatment				
Cost of PRP (THB)	Gamma	7000	700	Primary data
Cost of PPV in 2020 (2-day admission or one-day surgery) (THB)	Gamma	30,000	2968	Primary data
Unit cost of bevacizumab per one injection (THB)**	Gamma	769	107	Primary data

Table 1 continued

Parameters	Distribution	Mean	SE	Source
Unit cost of an intravitreal administration and topical antibiotic (THB)	Gamma	1719	172	Primary data
Unit cost of OCT (THB)	Gamma	700	70	Primary data
Frequency of treatment using PPV	Gamma	1.00	0.10	[20]
Proportion of the patients with PDR who undergo PPV	Beta	0.33	0.03	[20]
Frequency of treatment using PRP***	Gamma	2.00	0.20	[21]
Frequency of OP visit (per year) for PRP in STDR without DME: 1st year	Gamma	2.00	0.20	Assumed: according to screening and treatment frequencies
2nd–5th year	Gamma	1.00	0.10	
Subsequent year	Gamma	1.00	0.10	
Frequency of IVB (per year) for DME; 1st year	Gamma	10.00	1.00	[21, 22]
Frequency of IVB (per year) for DME; 2nd–5th year	Gamma	1.75	0.18	[21, 22]
Frequency of OP visit (per year) for DME; 1st year	Gamma	12.00	1.20	Assumed: according to screening and treatment frequencies
Frequency of OP visit (per year) for DME; 2nd–5th year	Gamma	5.25	0.53	
Frequency of OP visit (per year) for DME; Subsequent year	Gamma	2.00	0.20	
Direct non-medical cost				
Cost of travel for OPD per visit, primary facility (THB)	Gamma	59	12	[19]
Cost of travel for OPD per visit, tertiary facility (THB)	Gamma	156	31	[19]
Cost of food for OPD per visit, primary facility (THB)	Gamma	15	3	[19]
Cost of food for OPD per visit, tertiary facility (THB)	Gamma	57	11	[19]
Patient time spent for OP visit at primary facility (min)	Gamma	69	14	[19]

Table 1 continued

Parameters	Distribution	Mean	SE	Source
Patient travel time spent for OP visit at primary facility (min)	Gamma	18	4	[19]
Patient travel time spent for OP visit at tertiary facility (min)	Gamma	60	12	[19]
Caregiver time spent for care to patient with BB (hours/week)	Gamma	28	6	[13]
Gross National Income per capita (THB/year)	Gamma	240,586	48,117	[23]
Utilities				
Utility of patients with no DR	Beta	0.748	0.026	[24]
Utility of patients with non-STDR (mild to moderate NPDR)	Beta	0.752	0.037	[24]
Utility of patients with STDR without DME	Beta	0.628	0.052	[24]
Utility of patients with DME****	Beta	0.628	0.052	Assumed
Utility of DR patients with blindness	Beta	0.355	0.128	[24]

The relative risks of mortality among patients in the different health states were combined with age-specific all-cause mortality derived from the Thai population's life table data [25]. Direct medical costs of treatment for STDR and DME were estimated by assuming 75% of patients have the bilateral DR. This cost of image grading by DL per patient should cover practice expense of DL in Thailand considering the large population of patients screened in the national program. All incurred costs were converted to 2020 values using the consumer price index for Thailand and were converted to US$ using the exchange rate as of 1 July 2021 of 32.02 THB per US$ [26]

BB bilateral blindness, *CVD* cardiovascular disease, *DL* deep learning, *DR* diabetic retinopathy, *HG* human graders, *IPD* inpatient department, *IVB* intravitreal bevacizumab, *OCT* optical coherence tomography, *OP* outpatient, *OPD* outpatient department, *PDR* proliferative diabetic retinopathy, *PPV* Pars Plana vitrectomy, *PRP* panretinal photocoagulation, *STDR* sight-threatening diabetic retinopathy, *THB* Thai baht, *NHSO* National Health Security Office

*Adjusted from the monthly salary of trained human graders at 26,000 THB; **according to Thai National Health Security Office reimbursement policy, 1 vial of bevacizumab 100 mg/4 ml is divided into 30 doses; ***panretinal photocoagulation was divided to be performed twice in an eye as a standard practice; ****patients with STDR with or without DME were assumed to have the same utility

image interpretation. The first two costs were the same in both arms. The average direct medical cost for image interpretation by HG was estimated at 56 Thai baht (THB) (~ US$ 1.7) per patient. Additional costs included setting up certified training courses. The cost for image interpretation and other related practice expenses using DL was set at the price of 32 THB (~ US$ 1) per patient, a pricing estimate

comparable with HG interpretation. We probed this price per patient further in threshold analyses.

Total annual cost related to screening at primary health centers using DL, HG, and confirmation of referrals at tertiary health centers by retinal specialists were 375 (~ US$ 11.7), 399 (~ US$ 12.5), and 535 (~ US$ 16.7) THB, respectively. The details of these costs, including the cost of outpatient service, fundus

photography, and visual acuity measurement, are given in Fig. 2.

Cost of treatment for STDR without DME included the cost of pan-retinal photocoagulation (PRP), at the average of two sessions, at 7000 THB per patient. For PDR, the cost was the weighted average between a proportion of patients (two-third) treated with PRP and another proportion (one-third) treated with surgery [31], which was vitrectomy, at an average of 30,000 THB per patient. Patients with DME were treated with 17 doses of bevacizumab, the number of intravitreal injections during the first 5 years of treatment derived

from clinical trials [22]. Total cost of treatment for DME was 59,503 THB in the first year, 11,388 THB in the second year, and 619 THB for subsequent years. The cost of bevacizumab was used since this medication is listed on Thailand's National Essential List of Medicines and used as the first-line treatment for DME in the country.

Direct non-medical costs comprised costs of food, transportation, accommodation, equipment and facilities for patients, and productivity loss of caregivers, which was calculated based on gross national income per capita,
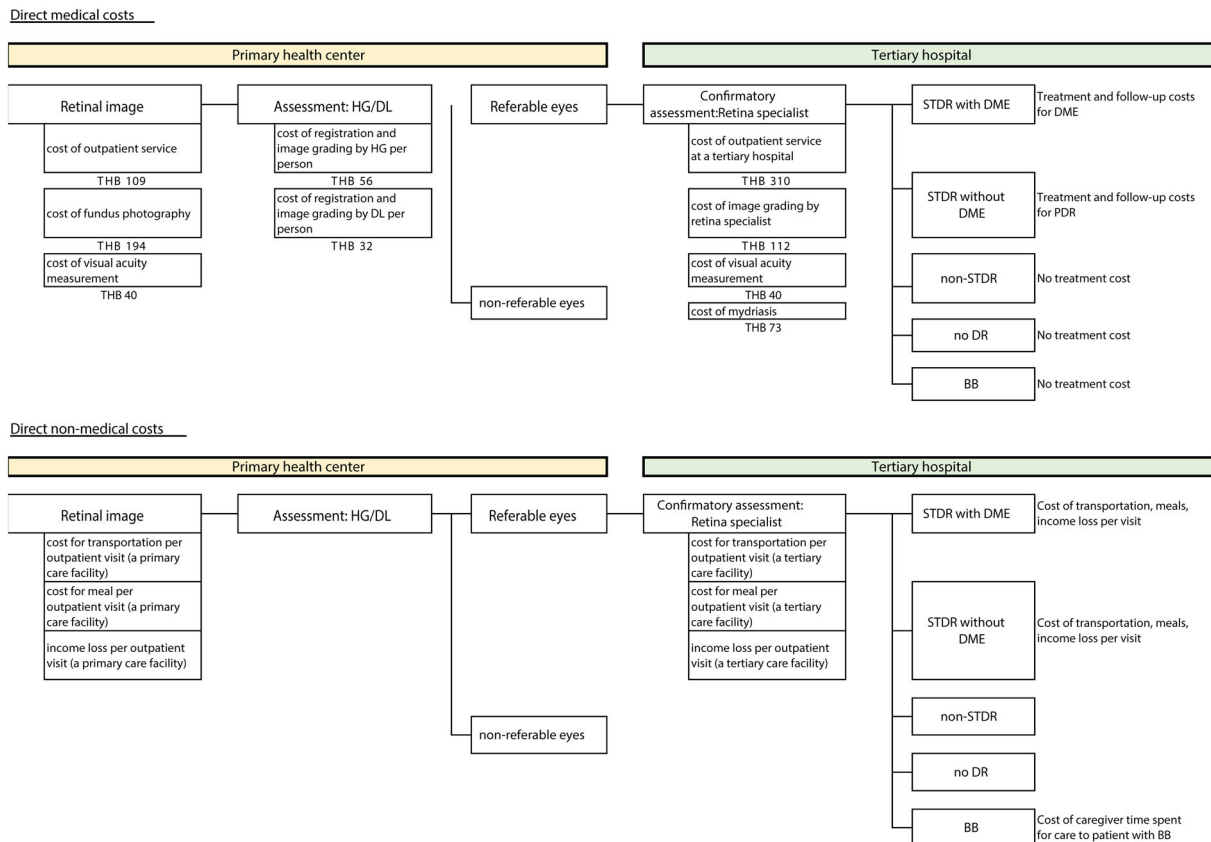

Fig. 2 Summary of details of screening and treatment costs. *BB* bilateral blindness, *DL* deep learning, *DR* diabetic retinopathy, *DME* diabetic macular edema, *HG* trained human graders, *STDR* sight-threatening diabetic retinopathy, *THB* Thai baht. Unit cost of treatment of DME includes cost of bevacizumab and intravitreal administration listed in Table 1. Unit cost of treatment of STDR includes cost of laser photocoagulation and vitrectomy (not shown in this figure but shown in Table 1). Treatment for DME includes cost of outpatient service, bevacizumab, intravitreal injection, and macular imaging by optical coherence tomography. Treatment for STDR in the first year covers the cost of outpatient service and cost of laser photocoagulation, or cost of vitrectomy and inpatient service

assuming caregiving time at 4 h per day. These costs per patient are the same for both arms.

Indirect costs were omitted to avoid double counting in the CUA according to the recommendation from the Thai Health Technology Assessment guideline. All incurred costs were converted to 2020 values using the consumer price index for Thailand and were converted to US\$ using the exchange rate as of 1 July 2021 of 32.02 THB per US\$. The cost of blindness is comprised primarily of visual rehabilitation aids and services, including residential and community care [32]. In Thailand, visual rehabilitation clinics are operated only in a few training centers with low patient volumes [33]. Therefore, this cost is minimal compared to the treatment costs and is not considered in the health provider perspective. Disaggregated costs from both societal and provider perspectives are shown in Fig. 3.

Health Utilities

Since the health utility weights for patients with DR in Thailand or Asia are not available, we applied those from the next closest match, i.e., those associated with patients with DR at a primary care service in Brazil, another MIC [34]. In that study, the utility values were estimated using the Brazilian EuroQol five dimension (EQ-5D) tariffs from patients without DR, with non-STDR, STDR, and bilateral blindness. Since DME was not considered in that model, we assumed that those who progressed into STDR with and without DME had the same utility values.

All these input parameters are shown in detail in Table 1 and Fig. 2.

Result Presentation

Total cost, life-years (LY), quality-adjusted life-years (QALY), and disaggregated costs for each screening strategy were reported. The incremental cost-effectiveness ratios (ICERs) per QALY gained in THB for DL and HG were calculated by dividing the incremental cost by incremental QALY. The Thai societal willingness-to-pay (WTP) threshold was 160,000 THB (∼ US\$ 4997) per QALY gained [24].
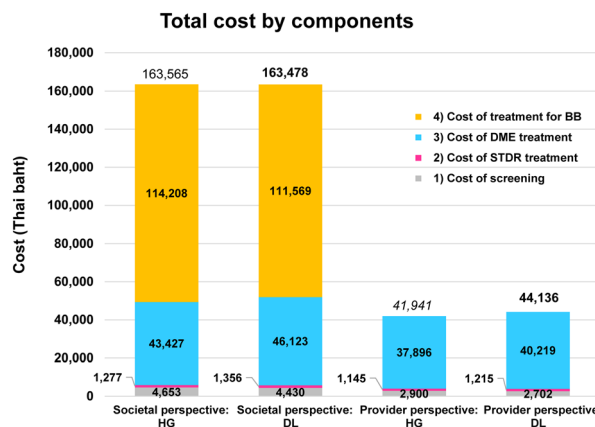

Fig. 3 Total and disaggregated costs for the two screening strategies (HG and DL). The costs are for the cost-utility analysis at base case from both societal and provider perspectives. *BB* bilateral blindness, *DL* deep learning, *DME* diabetic macular edema, *HG* trained human graders, *STDR* sight-threatening diabetic retinopathy. Costs of treatment of DME and STDR without DME are presented separately. We assumed no direct medical costs for bilateral blindness; all the values are Thai baht in 2020

Uncertainty Analyses

Parameter Uncertainty

To characterize the uncertainties of each input parameter, a one-way sensitivity analysis was performed and the percentage changes in the incremental net monetary benefit (iNMB) from the base case analysis were presented using a tornado diagram. Net monetary benefit of each screening method was calculated by multiplying the incremental QALY by the Thai societal WTP threshold at 160,000 THB/QALY and then subtracting the incremental cost. Parameter values were varied over the 95% confidence interval (CI) for the general parameters. The discount rates for costs and outcomes were varied from 0 to 6%.

The robustness of the results was evaluated by probabilistic sensitivity analyses (PSAs). All potential influencing parameters were varied simultaneously. The Monte Carlo simulation was run for 1000 iterations, and the results of PSA were shown as cost-effectiveness acceptability curves (CEACs) and a cost-effectiveness plane (CEP).

Table 2 Total costs, quality-adjusted life-years, and incremental cost-effectiveness ratios between the screening modalities from the base case analysis

Screening modality	Total cost (THB)	Total effectiveness		Incremental cost (THB)	Incremental effectiveness		ICER (THB per QALY gained)	iNMB (THB)
		LY	QALY		LY	QALY		
Societal perspective								
HG	163,565.04	18.5325	12.8574					
DL	163,478.16	18.5325	12.8617	− 86.88	0.0000	0.0043	Dominant	771.50
Provider perspective								
HG	41,941.10	18.5325	12.8574					
DL	44,135.95	18.5325	12.8617	2194.85	0.0000	0.0043	512,955.12	− 1510.24

BB bilateral blindness, *DL* deep learning, *DME* diabetic macular edema, *ICER* incremental cost-effectiveness ratio, *iNMB* incremental net monetary benefit, *LY* life year, *HG* trained human graders, *QALY* quality-adjusted life year, *STDR* sight-threatening diabetic retinopathy, *THB* Thai baht

Scenario Analysis

We conducted a threshold analysis varying the unit costs of DR screenings by both strategies at the Thai societal WTP. We also analyzed the effect of varying compliance to referrals for treatment in terms of iNMB between both strategies when the unit costs were fixed at base case. All uncertainty analyses were performed from a societal perspective over a lifetime horizon of patients.

RESULTS

From both perspectives, we found equal effectiveness in LY for HG and DL at 18.53, whereas QALYs were 12.857 and 12.862, respectively. From a societal perspective, DL was found to have a slightly lower incremental cost at 87 THB ($\sim$ US\$ 2.6) with an iNMB of 771.5 THB ($\sim$ US\$ 24.1) and being a dominant option compared to HG. From a provider perspective, DL was found to have higher incremental cost at 2195 THB ($\sim$ US\$ 66.7) and the ICER was 512,955 THB ($\sim$ US\$ 16,020) per QALY gained (Table 2). The CEP is shown in Fig. 4.

The reason for the difference between ICERs from a societal and provider perspective is presented in Fig. 3. The provider perspective did not incorporate the non-medical cost of bilateral blindness, which represents a substantial portion of lifetime costs (68.25% and 69.82% for DL and HG respectively from a societal perspective).

For parameter uncertainty, the tornado diagram (Fig. 5) demonstrates that the relative risk of DR progression in patients who received treatments was the most influential factor related to the change in iNMB. The PSA conducted from a societal perspective reinforces the robustness of the base case results. The CEAC (Fig. 6) illustrates that DL had the higher probability of being cost-effective at 84.6%, compared to HG at 15.4%, at the Thai WTP threshold.

Concerning the unit costs, in the base case, when the unit cost of image interpretation by HG was fixed at 56.11 THB ($\sim$ US\$ 1.7) per person, DL remained cost-effective from the societal perspective for its unit costs of image interpretation up to 150.5 THB ($\sim$ US\$ 4) per person at the Thai societal WTP threshold. We found that screening by DL could remain cost-effective even if compliance with referrals for treatment were lower than that of HG. The threshold of this compliance for treatment of screening by DL was 43.57%. This means screening by DL remained cost-effective when at least 44% of referrals detected by DL received
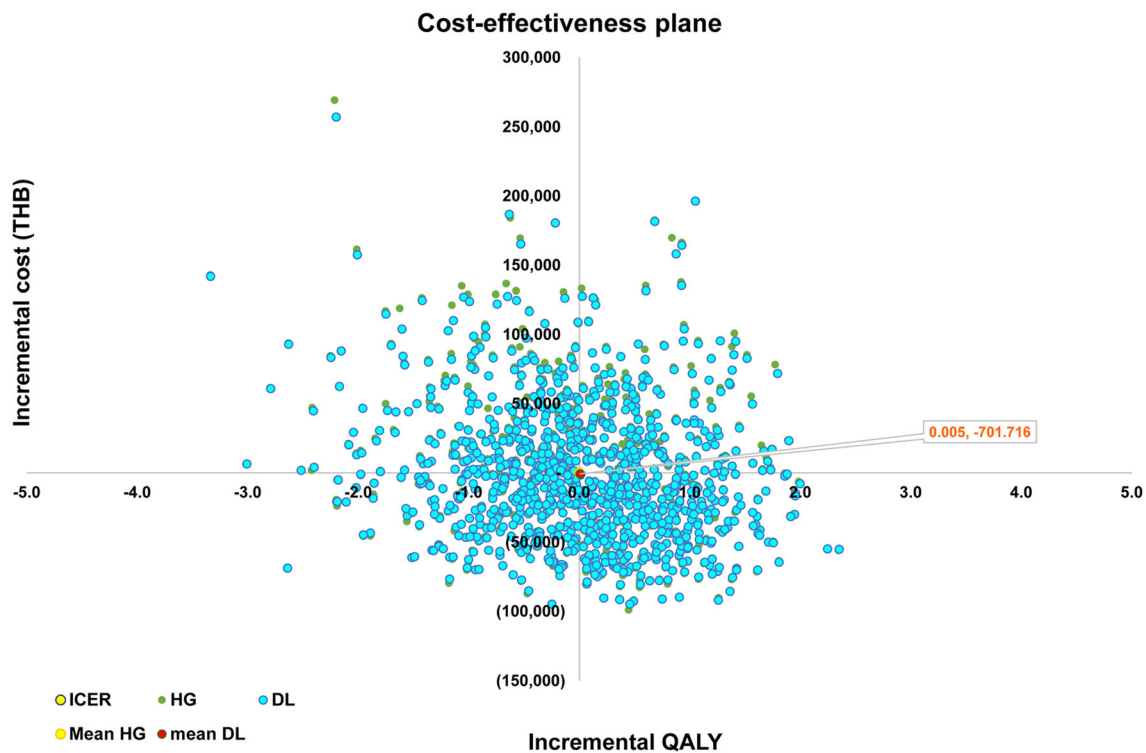
Cost-effectiveness plane

Fig. 4 Cost-effectiveness plane (CEP). *DL* deep learning, *HG* human graders, *ICER* incremental cost-effectiveness ratio, *QALY* quality-adjusted life-year, *THB* Thai baht

treatment compared to 60% compliance at base case for screening by HG. If compliance to referrals for treatment is even higher, screening using either strategy could achieve greater cost-savings with higher compliance to referrals for treatment as shown in Fig. 7.

DISCUSSION

In this DR screening program, we found DL had lower incremental costs compared to HG from a societal perspective. From a provider perspective, DL could detect more referral cases and therefore result in greater costs incurred from treatments. However, when the societal cost incurred from bilateral blindness, which imposes both humanistic and economic burden [32], was included, this model showed DL as a cost-saving screening tool. The saving of this societal cost, which mainly covers the costs of visual rehabilitation service by caregivers and

community care, is hidden from the provider perspective.

In settings like Thailand, where a well-established screening program and allocation of hardware equipment already exist, we anticipate that costs for adding DL software into the screening workflow is minimal. Since the labor cost is low and patients with STDR usually comprise the minority of patients screened, the additional labor of retinal specialists to confirm the increased numbers of more positive cases of STDR detected by DL should incur minimal cost.

While the cost of AI for DR screening has not yet been standardized, Scotland et al. proposed that the cost should include costs of initial implementation and maintenance with a lifespan of 10 years. The authors estimated that the server-based AI software used in Scotland under universal healthcare should be charged at 0.14 pound (6.44 THB or ~ US$ 0.2) per patient annually in 2006 [35]. In our study, we set the price charge of DL, also server-based and used
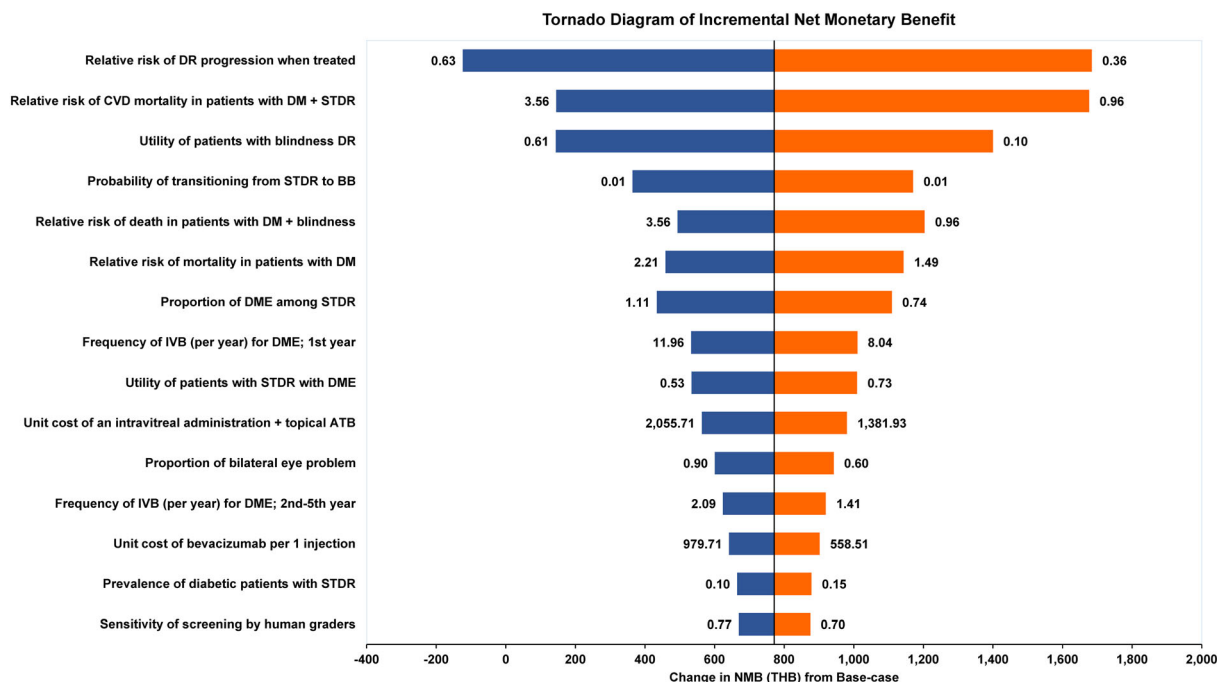
Fig. 5 One-way sensitivity analysis. Tornado diagram from one-way sensitivity analysis showing the percentage changes in incremental net monetary benefit (iNMB) of DL screening versus HG screening from the base case attributable to the change of each parameter. *ATB* antibiotics, *BB* bilateral blindness, *CVD* cardiovascular diseases, *DL* deep learning, *DM* diabetic mellitus, *DME* diabetic macular edema, *DR* diabetic retinopathy, *HG* human grader, *IVB* intravitreal bevacizumab, *STDR* sight-threatening diabetic retinopathy, *THB* Thai baht. Labels on the chart (next to bars) indicate input values (minimum and maximum) of each parameter; all the values of costs are Thai baht in 2020

under universal health coverage, at US$ 1 per patient annually. In a recently published CUA conducted in rural China, which found that AI-based DR screening was more cost-effective compared with ophthalmologist screening or no screening, the AI was priced similarly at US$ 1.447 per patient, based on market quotation of the supplier [36].

Our CUA supports the lower incremental costs of DL and extend the body of evidence in a cost-minimization analysis [37] and cost-effectiveness analyses [35, 38], which evaluated cost-savings over a year of screening. Our finding that greater patient compliance to referrals can potentially augment cost-effectiveness for DL is a motivation to explore additional strategies for boosting this compliance, such as digital notifications sent to patients. Ideally, a system for monitoring compliance to referrals should be built into a screening program to measure its

effectiveness. This compliance may be increased by the capability of DL to provide screening results immediately at the point of care [28]. Liu et al. found that DL could increase this compliance from 18.7% in the historical cohort to 55.4% [27]. Fuller et al. also found in a 5-year CUA that, with 54.9% compliance to referrals from pre-screening by DL compared to 11.5% compliance without pre-screening, DL could reduce costs by 23.3% [8].

A strength of our study is that our model evaluated cost-savings over a lifetime horizon of patients. Moreover, it was based on a nation-wide screening program with data representing all health regions. Another strength is the inclusion of the cost of treatment of DME, which has rarely been considered in previously published economic evaluations of DL for DR screening. This inclusion is pivotal because the treatment of DME was the primary cost burden
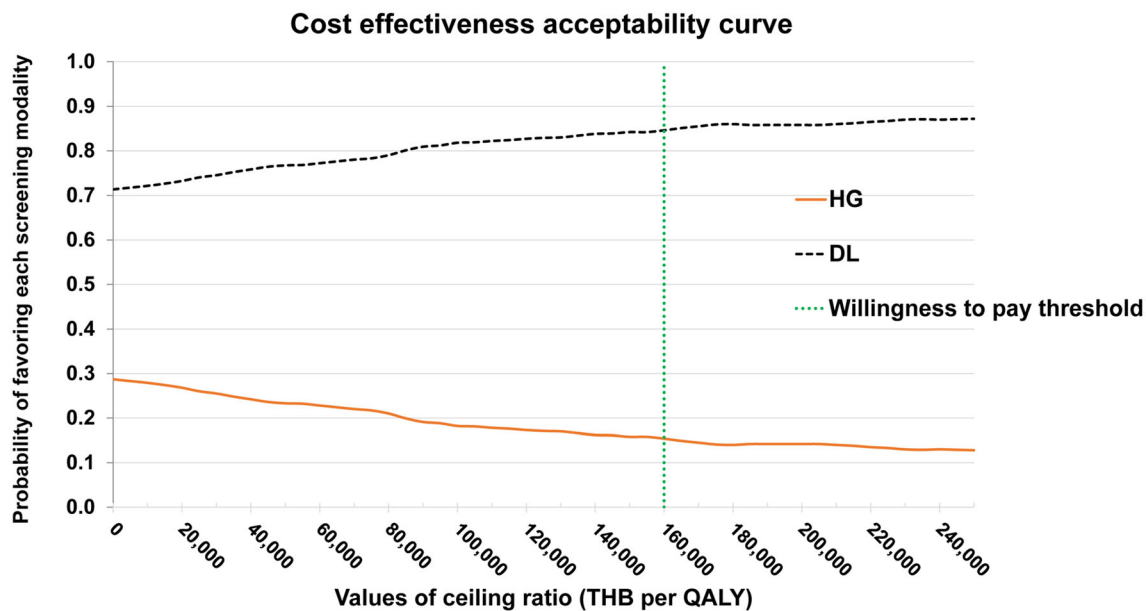
Cost effectiveness acceptability curve

Fig. 6 Cost-effectiveness acceptability curves. The curves compare the probabilities of being cost-effective at different willingness to pay of screening using HG and DL in the base case scenario; all the values are Thai baht in 2020. *DL* deep learning, *HG* human graders, *QALY* quality-adjusted life-year, *THB* Thai baht

of treatment of STDR in our model. In a US study, the cost of treatment of DME using ranibizumab, which is ten times more expensive than bevacizumab, was included in this 5-year CUA comparing pre-screening of DR using DL versus no pre-screening (refer all patients) in low-income patients. Even with this tremendous cost increase, DL for the pre-screening of DR was still more cost-saving than no pre-screening [8]. Another strength is our economic evaluation followed the Consolidated Health Economic Evaluation Standards.

There were some limitations to this study. Certain data are not available for inclusion in the analysis, such as national data on referral compliance and the health utility weights of patients with DR. Second, while the use of the Markov model for capturing changes in clinical health states and the use of health state utility values are traditionally applied for economic evaluations, these applications may have some limitations. Different countries might utilize different classifications of the clinical states of DR. The utility values of each state might also be varied among countries [39]. Therefore, the utility values derived from the Brazil study may

not exactly reflect the utilities of DR in Thailand.

Another limitation is that this model did not consider the possibility that DL may flag more ungradable patients than HG in real-world scenarios [40]. The patients with negative screening results (non-referrals) were not confirmed by experts; therefore, false negatives were not fully accounted for. DR screening programs in some countries implement a system for random sampling and review of 10% of negative screening results [2, 38]. According to our previous analysis [11], the proportion of false negatives by either DL or HG was a small fraction, and they would be identified in the next screening.

CONCLUSION

In conclusion, this CUA demonstrates that adoption of DL for a DR screening program in an MIC may result in lower societal costs over a lifetime of patients, providing economic rationale to expand DL-based DR screening in other MICs with similar prevalence of diabetes,

(A) Unit costs of DL and HG screenings

Unit cost of HG (THB) (% change from the cost at base-case analysis)	16 (-50%)	32 (0%)	64 (100%)	160 (400%)	224 (600%)	288 (800%)	352 (1000%)	416 (1200%)	480 (1400%)	544 (1600%)	608 (1800%)	672 (2000%)
1,178 (2000%)	8,363	8,259	8,050	7,425	7,008	6,592	6,175	5,758	5,341	4,924	4,508	4,091
1,066 (1800%)	7,614	7,510	7,301	6,676	6,260	5,843	5,426	5,009	4,593	4,176	3,759	3,342
954 (1600%)	6,865	6,761	6,553	5,928	5,511	5,094	4,677	4,261	3,844	3,427	3,010	2,594
842 (1400%)	6,117	6,012	5,804	5,179	4,762	4,345	3,929	3,512	3,095	2,678	2,262	1,845
729 (1200%)	5,368	5,264	5,055	4,430	4,013	3,597	3,180	2,763	2,346	1,930	1,513	1,096
617 (1000%)	4,619	4,515	4,307	3,682	3,265	2,848	2,431	2,014	1,598	1,181	764	347
505 (800%)	3,871	3,766	3,558	2,933	2,516	2,099	1,683	1,266	849	432	15	-401
393 (600%)	3,122	3,018	2,809	2,184	1,767	1,351	934	517	100	-316	-733	-1,150
281 (400%)	2,373	2,269	2,061	1,435	1,019	602	185	-232	-648	-1,065	-1,482	-1,899
112 (100%)	1,250	1,146	937	312	-104	-521	-938	-1,355	-1,711	-2,188	-2,605	-3,022
56 (0%)	876	771	563	-62	-479	-896	-1,312	-1,729	-2,146	-2,563	-2,979	-3,396
28 (-50%)	689	584	376	-249	-666	-1,083	-1,499	-1,916	-2,333	-2,750	-3,167	-3,583

Unit cost of DL (THB) (% change from the cost at base-case analysis)

(B) Changes in the iNMB when varying compliance to follow up for treatment of DL screening (given that the remaining parameters are constant)
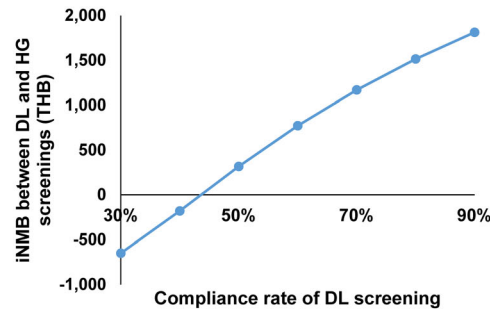
Chart: y-axis "iNMB between DL and HG screenings (THB)" ranging from -1,000 to 2,000; x-axis "Compliance rate of DL screening" ranging from 30% to 90%. The line rises from about -650 at 30% through 0 near 44% up to about 1,830 at 90%.

Fig. 7 Additional sensitivity analyses. For figure (**A**), the net monetary benefit (NMB) of each screening modality was calculated from total QALY of the intervention multiplied by the Thai willingness-to-pay threshold (160,000 THB per QALY) and then subtracting the total cost of the intervention. Incremental NMB (iNMB) was the difference between NMB of screening using DL and HG. Values in the table are iNMB of screening using DL and HG. Green cells represent the scenarios when the NMB of DL is greater than the NMB of HG (DL is preferred to HG for the screening). The uptake and compliance rates of HG screening were fixed as base case at 50% and 60%, respectively. Figure (**B**) shows iNMB between DL and HG screenings increases from zero when the compliance rate of DL screening is more than 44%

clinical specialist shortages, and low compliance to referrals for treatment. However, policy makers should be aware of budget impact of treating more patients with STDR with clinical deployment of DL. DR screening using DL or HG could achieve greater cost-savings with higher compliance to referrals for treatment. This compliance rate should be monitored as part of DR screening programs.

ACKNOWLEDGEMENTS

Tiwari, Fred Hersch, and Paisan Ruamviboonsuk were involved in the study concept and design. Attasit Srisubat, Kankamon Kittrongsiri, Sermsiri Sangroongruangsri, Prut Hanutsaha, Saowalak Turongkaravee, and Chalida Khemvaranan constructed the model. Attasit Srisubat, Kankamon Kittrongsiri, Sermsiri Sangroongruangsri, Paisan Ruamviboonsuk, Chalida Khemvaranan, Jacqueline Baras Shreibati, and Jack Ching did data acquisition and data analysis. Paisan Ruamviboonsuk, Kankamon Kittrongsiri, and Sermsiri Sangroongruangsri drafted the initial manuscript. Varis Ruamviboonsuk performed the literature search and reference citation. Sermsiri Sangroongruangsri, Richa Tiwari, Yun Liu, Rajiv Raman and Paisan Ruamviboonsuk supervised manuscript development and finalized the manuscript. All authors interpreted data, commented, and revised the manuscript critically. Paisan Ruamviboonsuk was involved in overall study supervision.

Disclosures. The authors report no conflicts of interest in this study. Paisan Ruamviboonsuk received consulting fees from and participated in Advisory Boards of Roche, Bayer, and Novartis; honoraria for lectures for Roche, Novartis, Bayer, and Topcon; support for meetings or traveling from Roche, Novartis, and TRB Medicals. Prut Hanutsaha received a research grant on macular degeneration from Novartis, participated in Advisory Boards of Bayer, Roche and Novartis, received honoraria for lectures for Afta, Bayer, and Novartis, and Zeiss, and received support for meetings or traveling from Bayer. John Hernandez, Jacqueline Baras Shreibati, Jack Ching, Richa Tiwari, Yun Liu, and Fred Hersch are Google employees and own stock at Alphabet. The authors received no financial support for conducting this research. Attasit Srisubat, Chalida Khemvaranan, Kankamon Kittrongsiri, Saowalak Turongkaravee, Sermsiri Sangroongruangsri, Varis Ruamviboonsuk, and Rajiv Raman have nothing to disclose.

Compliance with Ethics Guidelines. This study was approved by the Institutional Review Board of Rajavithi Hospital, Thailand (COA: 087/2562) and conducted according to the Declaration of Helsinki. Secondary data from previous studies and literature were used in this study.

Data Availability. Data sharing is not applicable to this article as no datasets were generated or analyzed during this model-based economic evaluation study.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSWW, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye. 2020;34:451–60.

2. Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. Br J Ophthalmol. 2021;105:723–8.

3. Zhang Y, Shi J, Peng Y, Zhao Z, Zheng Q, Wang Z, et al. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. BMJ Open Diabetes Res Care. 2020;8:e001596.

4. Ruamviboonsuk P, Chantra S, Seresirikachorn K, Ruamviboonsuk V, Sangroongruangsri S. Economic

evaluations of artificial intelligence in ophthalmology. Asia-Pacific J Ophthalmol. 2021;10: 307–16.

5. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proc 2020 CHI conf hum factors comput syst. New York: ACM; 2020. p. 1–12

6. Ruamviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. npj Digit Med. 2019;2:25.

7. Ruamviboonsuk P, Tiwari R, Sayres R, Nganthavee V, Hemarat K, Kongprayoon A, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. Lancet Digit Heal. 2022;4:e235–44.

8. Fuller SD, Hu J, Liu JC, Gibson E, Gregory M, Kuo J, 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. 2022;16:415–27.

9. Wilkinson C, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110: 1677–82.

10. Wong TY, Sun J, Kawasaki R, Ruamviboonsuk P, Gupta N, Lansingh VC, et al. Guidelines on diabetic eye care: the international council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. Ophthalmology. 2018;125:1608–22.

11. Limwattanayingyong J, Nganthavee V, Seresirikachorn K, Singalavanija T, Soonthornworasiri N, Ruamviboonsuk V, et al. Longitudinal screening for diabetic retinopathy in a nationwide screening program: comparing deep learning and human graders. J Diabetes Res. 2020;2020:1–8.

12. Ben ÂJ, Neyeloff JL, de Souza CF, Rosses APO, de Araujo AL, Szortika A, 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;18:57–68.

13. 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: 837–53.

14. 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;7:225–41.

15. Evans JR, Fau C, Virgili G. Laser photocoagulation for diabetic retinopathy. In: Evans JR, editor. Cochrane Database Syst Rev [Internet]. Chichester: Wiley; 2014. https://doi.org/10.1002/14651858. CD011234

16. Nwaneri C, Cooper H, Bowen-Jones D. Mortality in type 2 diabetes mellitus: magnitude of the evidence from a systematic review and meta-analysis. Br J Diabetes Vasc Dis. 2013;13:192–207.

17. Xu X-H, Sun B, Zhong S, Wei D-D, Hong Z, Dong A-Q. Diabetic retinopathy predicts cardiovascular mortality in diabetes: a meta-analysis. BMC Cardiovasc Disord. 2020;20:478.

18. Riewpaiboon A. Standard cost lists of health technology assesement. Nonthaburi: Health Intervention and Technology Assessment Program; 2011.

19. Lafuma A, Brezin A, Fagnani F, Mimaud V, Mesbah M, Berdeaux G. Nonmedical economic consequences attributable to visual impairment. Eur J Heal Econ. 2006;7:158–64.

20. Bressler NM, Beck RW, Ferris FL. Panretinal photocoagulation for proliferative diabetic retinopathy. N Engl J Med. 2011;365:1520–6.

21. Glassman AR, Wells JA, Josic K, Maguire MG, Antoszyk AN, Baker C, et al. Five-year outcomes after initial aflibercept, bevacizumab, or ranibizumab treatment for diabetic macular edema (protocol T extension study). Ophthalmology. 2020;127: 1201–10.

22. Wells JA, Glassman AR, Ayala AR, Jampol LM, Bressler NM, Bressler SB, et al. Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema: two-year results from a comparative effectiveness randomized clinical trial. Ophthalmology. 2016;123:1351–9.

23. The World Bank. GNI per capita, Atlas method (current US\$) - Thailand [Internet]. [cited 2021 Sep 2]. https://data.worldbank.org/indicator/NY.GNP. PCAP.CD?end=2019&locations=TH&start=1962.

24. Thavorncharoensap M, Sompitak S, Doungthipsirikul S. Policy brief: What is the value of a Quality-Adjusted Life Year and how much is it? 2015 Jul.

25. World Health Organization. Life tables by country: Thailand 2020 [Internet]. World Health Organization; 2021 [cited 2021 Sep 25]. https://apps.who.int/gho/data/view.main.61640?lang=en.

26. Bank of Thailand. Daily Foreign Exchange Rates [Internet]. 2021. https://www.bot.or.th/english/_layouts/application/exchangerate/exchangerate.aspx

27. Liu J, Gibson E, Ramchal S, Shankar V, Piggott K, Sychev Y, et al. Diabetic retinopathy screening with automated retinal image analysis in a primary care setting improves adherence to ophthalmic care. Ophthalmol Retina. 2021;5:71–7.

28. Pedersen ER, Cuadros J, Khan M, Fleischmann S, Wolff G, Hammel N, et al. Redesigning clinical pathways for immediate diabetic retinopathy screening results. NEJM Catal. 2021. https://doi.org/10.1056/CAT.21.0096.

29. Man REK, Fenwick EK, Sabanayagam C, Li L-J, Tey CS, Soon HJT, et al. differential impact of unilateral and bilateral classifications of diabetic retinopathy and diabetic macular edema on vision-related quality of life. Invest Opthalmol Vis Sci. 2016;57: 4655.

30. Wong TY, Sabanayagam C. Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence. Ophthalmologica. 2020;243:9–20.

31. Khan R, Surya J, Rajalakshmi R, Rani PK, Anantharaman G, Gopalakrishnan M, et al. Need for vitreous surgeries in proliferative diabetic retinopathy in 10-year follow-up: India retinal disease study group report no. 2. Ophthalmic Res. 2021;64:432–9.

32. Lopes N, Dias LLDS, Ávila M, Japiassú RM, Muccioli C, Bordon AF, et al. Humanistic and economic burden of blindness associated with retinal disorders in a Brazilian sample: a cross-sectional study. Adv Ther. 2021;38:4215–30.

33. Chotikavanich S, Chanvarapha N, Loket S, Yingyong R, Dongngam S, Nujoi W, et al. A 5-year retrospective record review of hospital-based low-vision rehabilitation in Thailand. Clin Optom. 2018;10:41–50.

34. Ben ÂJ, de Souza CF, Locatelli F, Rosses APO, Szortika A, de Araujo AL, et al. Health-related quality of life associated with diabetic retinopathy in patients at a public primary care service in southern Brazil. Arch Endocrinol Metab. 2020;64:575–83.

35. Scotland GS, McNamee P, Philip S, Fleming AD, Goatman KA, Prescott GJ, et al. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. Br J Ophthalmol. 2007;91: 1518–23.

36. Huang X-M, Yang B-F, Zheng W-L, Liu Q, Xiao F, Ouyang P-W, et al. Cost-effectiveness of artificial intelligence screening for diabetic retinopathy in rural China. BMC Health Serv Res. 2022;22:260.

37. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. Lancet Digit Heal. 2020;2: e240–9.

38. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. Ophthalmology. 2017;124:343–51.

39. Sampson CJ, Tosh JC, Cheyne CP, Broadbent D, James M. Health state utility values for diabetic retinopathy: protocol for a systematic review and meta-analysis. Syst Rev. 2015;4:15.

40. Tham Y-C, Anees A, Zhang L, Goh JHL, Rim TH, Nusinovici S, et al. Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study. Lancet Digit Heal. 2021;3:e29–40.