ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Themed Section: Artificial Intelligence

Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah

Madelon M. Voets, MSc, Jeroen Veltman, MD, PhD, Cornelis H. Slump, PhD, Sabine Siesling, PhD, Hendrik Koffijberg, PhD

ABSTRACT

Objectives: This study aimed to systematically review recent health economic evaluations (HEEs) of artificial intelligence (AI) applications in healthcare. The aim was to discuss pertinent methods, reporting quality and challenges for future implementation of AI in healthcare, and additionally advise future HEEs.

Methods: A systematic literature review was conducted in 2 databases (PubMed and Scopus) for articles published in the last 5 years. Two reviewers performed independent screening, full-text inclusion, data extraction, and appraisal. The Consolidated Health Economic Evaluation Reporting Standards and Philips checklist were used for the quality assessment of included studies.

Results: A total of 884 unique studies were identified; 20 were included for full-text review, covering a wide range of medical specialties and care pathway phases. The most commonly evaluated type of AI was automated medical image analysis models (n = 9, 45%). The prevailing health economic analysis was cost minimization (n = 8, 40%) with the costs saved per case as preferred outcome measure. A total of 9 studies (45%) reported model-based HEEs, 4 of which applied a time horizon >1 year. The evidence supporting the chosen analytical methods, assessment of uncertainty, and model structures was underreported. The reporting quality of the articles was moderate as on average studies reported on 66% of Consolidated Health Economic Evaluation Reporting Standards items.

Conclusions: HEEs of AI in healthcare are limited and often focus on costs rather than health impact. Surprisingly, model-based long-term evaluations are just as uncommon as model-based short-term evaluations. Consequently, insight into the actual benefits offered by AI is lagging behind current technological developments.

Keywords: artificial intelligence, cost-effectiveness, health economic evaluation, impact, modeling, systematic review.

VALUE HEALTH. 2022; 25(3):340–349

Introduction

Within the healthcare sector, artificial intelligence (AI) has seen a substantial rise in development over the past years because of growing interest and its potential impact on healthcare delivery and effectiveness.[1] Advancements in computing power and algorithms together with the digitization of large volumes of health data have made AI-supported healthcare increasingly more common. The progression of AI is taking center stage in how healthcare is personalized and delivered to patients, leading to new opportunities and challenges in clinical practice.[1,2] A fundamental challenge in today's healthcare is the growth of digital health data quickly exceeding the human capacity to process and analyze it in routine clinical practice. The advancement of AI carries the potential to address this gap and simultaneously improve patient care in clinical practice.[1] Additionally, impeding healthcare staff shortages, aging populations, and increasing costs at narrowing budgets are asserting pressure on healthcare systems. Consequently, the healthcare industry is progressively and understandably resorting to AI to address these challenges. Nowadays, AI is growing in different domains of healthcare, from the automation of clinical workflows to the interpretation of clinical findings and the prediction of health outcomes, treatment response, and disease recurrence.[3] At the rate at which AI applications are being developed, augmented, and used, AI creates an opportunity for accessible and evidence-based decision making within the global health community.[4] In the areas of image processing and also electronic health record interpretation, medical decision support through text mining, and the analysis of medical time series data (ie, longitudinal data on blood pressure, electrocardiograms), the use of AI has shown promising results.[5-7] Therefore, processing digital health data with AI could support the delivery of effective and efficient healthcare.[8] Nonetheless, even though the advancement of AI carries much potential, what value AI can and will deliver in actual clinical practice remains a central question and proper implementation guidance is crucial. Although the number of publications describing AI applications in a healthcare setting has been growing rapidly over the past years,

the majority solely report on their accuracy and precision.[9] Nevertheless, neither excellent prediction accuracy nor clear explainable relations between patient or image characteristics and outcomes guarantee clinical effectiveness and adoption. Plus the commonly used area under the curve of a receiver operating characteristic of a detection task does not unquestionably reflect clinical applicability.[10] At the same time, evidence supporting clinical effectiveness, specifically comparative effectiveness, cost-effectiveness, or other formal health technology assessment (HTA), of AI in a clinical healthcare setting appears to be limited.[1] HTA serves an important purpose for stakeholders and decision makers as a method to establish policies making most efficient use of available health resources before being implemented in clinical practice. Regulatory policy is currently lacking, hindering the adoption of AI. The medical community is overwhelmed by the large number of developed AIs, yet the absence of clear guidelines makes it difficult for researchers, policy makers, and developers to determine when an AI is indeed qualified for clinical adoption.

Therefore, the aim was to systematically review recent health economic evaluations (HEEs) of AI applications in healthcare and to discuss their methods, outcomes, and reported challenges. This systematic review focuses specifically on formal HEEs, such as cost-effectiveness and cost-utility analyses, as one of the dimensions of HTA, generating evidence on (long-term) impact to support implementation and financing decisions.[11] In addition, the quality of the reported studies was assessed based on published checklists. Challenges for the future HEEs and implementation of AI in healthcare were identified and discussed.

Methods

A systematic literature review of health economic analysis of AI applications was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses and described in the following sections.[12]

Literature Search Strategy

Considering that AI is currently developing at an accelerated pace and novel AI applications will likely outperform older AI applications, the systematic literature search focused on studies published within the last 5 years between January 1, 2016, and April 1, 2021. The studies were extracted from the PubMed and Scopus databases separately using the search queries provided in Table 1. To identify potential studies regarding health economic analysis, relevant free-text terms included "health outcomes" or "effects," "economic," "cost," "budget," and "quality." These terms were required to be present in article titles only, because these were too broad when applied for abstract searches. Cost-related free-text terms included terms such as "reduction," "minimization," "benefit," and "sensitive." Free-text terms for AI were used in both title and abstract and included "artificial intelligence," "machine learning," "deep learning," "computer-aided," and "data-driven." Besides broad terms such as "artificial intelligence," "machine learning," and "deep learning," additionally more specialized terms such as "support vector machine," "neural network," or "random forest" in similar manner are apt within the AI paradigm. Medical Subject Headings terms and keywords describing AI and cost analysis methodologies were used to further limit results. On account of the presumption that any relevant studies would not only contain specialized terms, but at least 1 broader term too, a sensitivity analysis for "neural networks," "support vector machine," and "random forest" in both title or abstract and Medical Subject Headings terms was conducted. Databases and search strategies were discussed with information specialists and search strategies were pilot tested to ensure all studies previously identified by authors were captured. The final database searches were performed on May 10, 2021.

Table 1. Search queries for title, abstracts, and keywords in the database search conducted on May 10, 2021.

Database	Query
PubMed	(("artificial intelligence"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "computer aided"[Title/Abstract] OR "CAD"[Title/Abstract] OR "data driven"[Title/Abstract]) AND ("health outc*"[Title] OR "health eff*"[Title] OR "quality*"[Title] OR "econom*"[Title] OR "cost*"[Title] OR "budget*"[Title] OR "implement*"[Title])) AND 2016/01/01:3000/12/31[Date - Publication] AND (cost effectiveness[MeSH Terms] OR artificial intelligence[MeSH Terms])
Scopus	(TITLE-ABS("AI" OR "artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "computer aided" OR "CAD" OR "data driven") AND TITLE("Health outc*" OR "health eff*" OR "econom*" OR "quality*" OR "cost*" OR "budget*" OR "implement*") AND KEY(cost AND effectiveness)) AND PUBYEAR > 2015

CAD indicates computer aided diagnosis; MeSH, Medical Subject Headings.

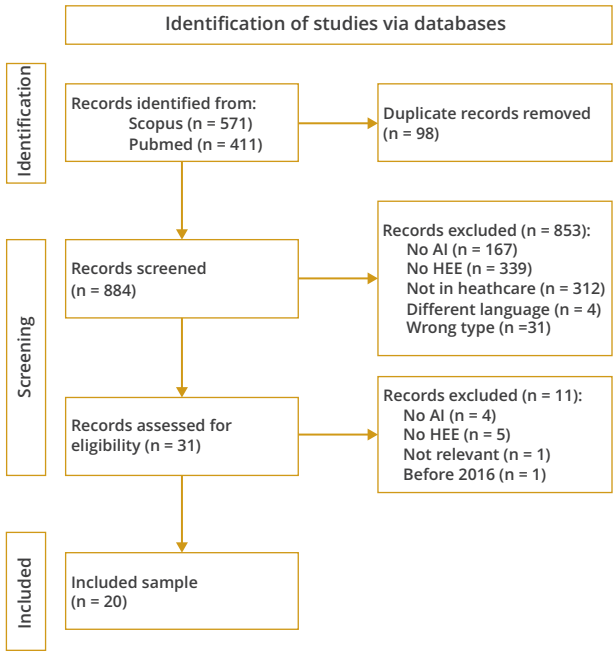
Inclusion and Exclusion Criteria

Studies that evaluated an AI application compared with standard care, other types of care, or another AI within the same healthcare setting and reported a quantified impact evaluation in terms of costs, health-related or process outcomes, or resources were included for analysis. Studies that did not comply with the inclusion criteria were excluded, as were those outside of healthcare, without any type of quantitative HEE of the AI application or not available in English. Moreover, studies of other types than "original research" or "systematic review," such as "commentary," "letter to the editor," or "editorial" were also excluded. Reviewers M.V. and R.K. independently screened the titles and abstracts of all identified records after duplicates were removed. Definitive inclusion or exclusion of the studies was concluded by the same 2 reviewers, who independently reviewed the full texts of the included studies. Persisting uncertainties or disagreements between reviewers during the screening or full-text review process were settled after consulting a third independent reviewer (H.K.).

Data Extraction

Relevant data of the included studies were extracted independently by the 2 reviewers (M.V. and R.K.). A data extraction form was designed and included several general aspects: year of publication, patient population, study location, and funding source. Specific information extracted regarding the subject AI included description of the AI, field of application, care pathway phase (prevention, diagnostics, intervention, prognosis, etc), and technological aspects (ie, pattern recognition, natural language processing, virtual reality, etc). Finally, extracted information related to the HEE involved type of health economic analysis, intervention and comparator, time horizon, perspective, health or cost outcomes, modeling technique, and sensitivity analysis.

Figure 1. PRISMA flowchart describing study selection and reasons for exclusion during full-text screening.



AI indicates artificial intelligence; HEE, health economic evaluation; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Quality Assessment

The methodological quality of the included studies was evaluated using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) and Philips checklist.[13,14] Although CHEERS was initially not developed as a quality assessment tool for health economic studies, but as a reporting checklist, its widespread use for evaluation of the design of health economic studies makes it an appropriated checklist for this systematic review.[15-17] CHEERS includes 24 items subdivided into 6 main categories to conduct a thorough reporting of HEEs, but does not include items directly related to model-based HEEs. The Philips checklist was specifically designed to analyze the quality of model-based HEEs and includes 3 pillars: modeling approach, model data, and assessment of uncertainty. The decision for a specific model type and justification of the model parameter values and handling of uncertainty are crucial to the perceived quality of the studies.[18] Therefore, the Philips checklist was applied in addition to the CHEERS checklist to model-based studies. Points were awarded for each of the criteria met. A point was withheld if the criterion was not completely met. A checklist score was derived for each included study based on the proportion of the 24 criteria met. The final included studies were independently reviewed by the same 2 reviewers (M.V. and R.K.).

Results

Search Results

The databases search identified 982 records in total, of which 98 were excluded as duplicates. In the remaining sample of 884 unique records, 853 records were excluded after title and abstract screening based on the exclusion criteria. The sensitivity analysis yielded 62, 13, and 1 additional articles, respectively. None were eligible for full-text screening after assessment of title and abstract. A total of 31 full-text studies were screened, after which another 11 studies were excluded because they did not meet the inclusion criteria. A flow diagram of records found, screened, selected, and excluded with corresponding exclusion criteria is shown in Figure 1. The remaining 20 studies were read full text.[19-38] The relevant data were extracted and the 2 reviewers independently further assessed the quality of the articles per the CHEERS and Philips checklists.

General Overview of the Included Studies

A general overview of the included studies, including the details of the reported AI applications, is provided in Table 2.[19-38] The majority of studies was published in 2019 or later. The studies were conducted in a range of medical specialties, yet ophthalmology was evidently the dominant field. A total of 4 of 20 articles (20%) involved ophthalmology, all of which evaluated the same AI application for diabetic retinopathy screening.[20,21,28,29] The most common type of AI was automated medical image analysis models, as 9 studies (45%) evaluated this type of AI.[19-22,27-30,32] A total of 6 studies (30%) evaluated pattern recognition in clinical data for predictive modeling.[24-26,36-38] Nevertheless, all phases of the care pathway from prevention to follow-up were supported by AI in at least 1 included study; screening and treatment monitoring (intervention) were the most prevalent phases of the care pathway in which AI was applied. Notably, 7 of 20 studies (35%) reported complete government funding, and 6 (30%) reported industry funding. One study (5%) reported that the industry funder participated in the analysis, data interpretation, and writing of the article.[36]

Health Economic Analysis

The primary health economic analysis was a cost-minimization analysis with the costs per case as primary outcome (n = 8, 40%). Most cost-minimization analysis adopted the hospital perspective (n = 4, 50%), and 2 others adopted the payer perspective. The remaining 2 studies applied a health system or patient perspective. A total of 3 studies incorporated the societal perspective in their evaluation, although the definition of this perspective varied.[32,34,37] One study defined the societal perspective by including healthcare utilization and productivity losses[34] and another by including reimbursement, opportunity, and additional hospitalization costs.[37] The third study did not explicitly elaborate on its definition of societal perspective.[32] The second most prevailing health economic analysis was incremental cost-effectiveness analysis (n = 6, 30%), with different health outcomes. One study reported improvements in life expectancy. A total of 3 studies performed incremental cost-utility analysis, using incremental quality-adjusted life-years as the health outcome. The remaining 3 studies reported the incremental effectiveness in context-specific outcome measures. Overall, the time horizons adopted by the studies ranged from 28 days to lifetime. A total of 6 studies (30%) reported a time horizon shorter than 12 months, 7 (35%) adopted a time horizon of 12 months, and 3 studies (15%) included patient lifetime as the time horizon. Of these studies, 4 (20%) reported the applied discount rates for health and economic outcomes. Two studies (10%) did not report a time horizon; nevertheless, 1 study evaluated incremental effectiveness on overall survival as health outcome.[23] A complete overview of the health economic methodological details of the included studies can be found in Table 3.[19-38]

Table 2. Characteristics of the included studies (N = 20).

Main author	Year	Patient population	Location	Description of AI application	Medical field	Care pathway phase	AI technology	CHEERS score, %	Funding
Philipsen et al[19]	2016	Self-referred Tuberculosis suspects	Africa	Automated digital chest radiography for tuberculosis detection*	Radiology	Screening	Automated image analysis	58	Government
Tufail et al[20]	2016	Consecutive patients with diabetes who attended a routine annual NHS DESP visit	UK	Automated retinal image analysis system for diabetic retinopathy screening*	Ophthalmology	Screening	Automated image analysis	96	Academic
Fernandez-Vicente et al[31]	2017	Patients in need of thumb orthosis	Europe	CAD software for 3D printing	Orthopedics	Intervention (treatment)	3D printing	50	Unknown
Takahashi et al[32]	2017	Women from 50 years old without breast cancer eligible for screening	Asia-Pacific	Mammography CAD and CT colonography CAD	Radiology	Screening	Automated image analysis	63	Government
Mervin et al[33]	2018	Geriatric residents, all 60 years or older and with documented dementia	Asia-Pacific	Therapeutic pet-type robot to reduce agitation and medication use in dementia	Geriatrics	Self-management	Robotics	83	Government
Bremer et al[34]	2018	Men and women aged 18 years or older with major depressive disorder, not of high suicidal risk, not treated for depression	Europe	Internet-based personalized psychology treatment recommendations	Psychology	Intervention (treatment)	Recommender system	67	EU
Gönel et al[35]	2018	No patient population	Turkey	Elimination method of unnecessary laboratory tests with 5 algorithms	Laboratory	Diagnostics	Recommender system	50	NR
Golas et al[36]	2018	Heart failure patients (men and women) who were discharged alive from an inpatient hospital admission	USA	Deep Unified Networks; a mesh-like network structure of deep learning designed to avoid overfitting	Heart failure	Intervention (treatment)	Pattern recognition	63	Industry
Padula et al[37]	2019	Hospitalized adults (men and women) with Braden scores	USA	Risk assessment for HAPI based on EHR	General	Prevention	Pattern recognition	83	Government
Lee et al[38]	2019	Adult patients (men and women) with total joint replacements	USA	ML prediction model with RUSBoost for total joint replacement readmission risk	Surgery	Intervention (treatment)	Pattern recognition	50	Government
Fuller et al[21]	2020	Adult patients (men and women) with diabetes	USA	Automated retinal image analysis system for diabetic retinopathy screening†	Ophthalmology	Screening	Automated image analysis	83	Industry and academic
Mansour et al[22]	2020	Patients (men and women) who presented with acute ischemic stroke	Africa	Automated cerebrovascular analysis on CT-imaging†	Neurology	Intervention (treatment)	Automated image analysis	38	NR
Murtojärvi et al[23]	2020	Male patients with metastatic castration-resistant prostate cancer	Europe	Survival prediction for patients with advanced prostate cancer	Urology	Prognosis	Feature Selection	54	Government
Eigner et al[24]	2020	Adult patients (men and women) hospitalized for a variety of major surgeries	Asia-Pacific	Prediction modeling for readmission risk	Surgery	Follow-up	Pattern recognition	25	NR
Hill et al[25]	2020	Patients (men and women) above 50 years old eligible for Atrial Fibrillation screening	UK	Targeted screening through risk prediction for atrial fibrillation	Cardiology	Screening	Pattern recognition	92	Industry
Rozenblum et al[26]	2020	Adult patients (men and women) who had at least 1 outpatient encounter in selected hospitals	USA	ML system to identify clinically valid medication error	General	Follow-up	Pattern recognition	67	Industry
Mori et al[27]	2020	Adult patients (men and women) with diminutive rectosigmoid polyps removed after assessment with AI	Asia-Pacific, Europe, USA	Support software for colorectal endoscopy to differentiate between neoplastic or non-neoplastic polyps‡	Gastroenterology	Diagnostics	Automated image analysis	58	Industry and government

continued on next page

Table 2. Continued

Main author	Year	Patient population	Location	Description of AI application	Medical field	Care pathway phase	AI technology	CHEERS score, %	Funding
Xie et al[28]	2020	Consecutive patients (men and women) with diabetes in a national diabetic retinopathy screening program	Asia-Pacific	Ensemble deep learning of 3 networks to detect referable diabetic retinopathy	Ophthalmology	Screening	Automated image analysis	75	Government
Wolf et al[29]	2020	Youths below 21 years old with Type 1 and Type 2 Diabetes	USA	Commercial AI, CNN as detector for diabetic retinopathy[†]	Ophthalmology	Screening	Automated image analysis	63	Industry
Schwendicke et al[30]	2021	Twelve years old individuals (men and women) with posterior permanent teeth	Europe	U-Net, a fully convolutional neural network	Dentistry	Diagnostics	Automated image analysis	96	None

3D indicates 3-dimensional; AI, artificial intelligence; CAD, computer aided diagnosis; CE, Conformité Européene; CHEERS, Consolidated Health Economic Evaluation Reporting Standards; CNN, convolutional neural network; CT, computed tomography; DESP, Diabetic Eye Screening Program; HER, electronic health record; EU, European Union; FDA, Food and Drug Administration; HAPI, hospital acquired pressure injury; ML, machine learning; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NR, not reported; UK, United Kingdom; USA, United States of America.
*CE approved.
[†]FDA approved.
[‡]NICE approved.

Quality Assessment

The methodological quality of studies was evaluated using the CHEERS and Philips checklists. A score was calculated as the percentage of criteria fulfilled for the analysis of the CHEERS results. As shown in Table 2,[19-38] the scores ranged from 25% to 96%, and the average score was 66%. The study with the highest score was an elaborate HTA study commissioned by the National Health Service in the United Kingdom.[20] A short article presumably containing preliminary results received the lowest score.[24] Additionally, Figure 2 specifies the number of studies that satisfied each checklist item, ranging from 1 to 20. All studies provided an explicit statement of the study context and the main question in the introduction. None but 1 study reported on handling variations among subgroups through conducting tests with different training and test cohorts.[23] A total of 6 studies (30%) described the analytical methods, 4 of which were modeling-based studies. Furthermore, 1 study evaluated an AI algorithm to maximize hospital profit based on retrospective data, but did not compare the outcomes to standard care profit.[24] This was the only study to not report a comparator. Finally, 11 (55%) and 13 of 20 studies (65%) explicitly mentioned study location and perspective, respectively.

The review included 9 modeling-based studies, and a summary of characteristics of these modeling-based studies is shown in Table 4.[20,21,25,28-30,32,37,38] The Philips checklist was used to report on the quality of the modeling-based studies.[14] Among the 9 modeling-based studies, merely 2 distinct model types were identified: decision trees and Markov models. A total of 4 of 9 modeling-based studies listed their specific reason for choosing a model type.[29,32,37,38] Markov modeling was particularly chosen because this allows incorporating recurring events[32] and time-dependent transitions.[37] Reasons listed for choosing decision trees were the trade-off between interpretability and accuracy[38] and the simplicity of the model.[29] Studies that did not list the specific rationale for the model structure used made it difficult to determine whether the modeling type was sufficiently reasoned (as reflected in Philips' items on structural assumptions and model type). The decision trees and the Markov models had simplistic structure. All decision trees had 2 or 3 arms, 1 for the AI application under evaluation and 1 or 2 for the comparator strategies, with similar events/states in each arm. The cycle lengths used in the Markov models ranged from 1 day to 2 years and were mostly determined based on healthcare protocols (ie, testing frequencies in practice) rather than the natural progression of disease. Regarding model data, only 3 of 9 studies reported to have used multiple studies or systematic reviews to synthesize model parameter values.[25,29,30] The other 6 studies (66%) reported a single data source, in 2 studies even from the same group or institution.[21,37] None of the 9 studies reported how heterogeneity and structural uncertainties were addressed. Methodological uncertainties were evaluated in 2 studies, both by means of sensitivity analysis with different clinical scenarios.[25,37] Nevertheless, all 9 studies reported on parameter uncertainty through sensitivity analysis. The majority of studies conducted 1-way or univariate and probabilistic analysis.[21,29,30,37] Two studies performed deterministic sensitivity analysis.[20,28] One study conducted 3-way sensitivity analysis.[38]

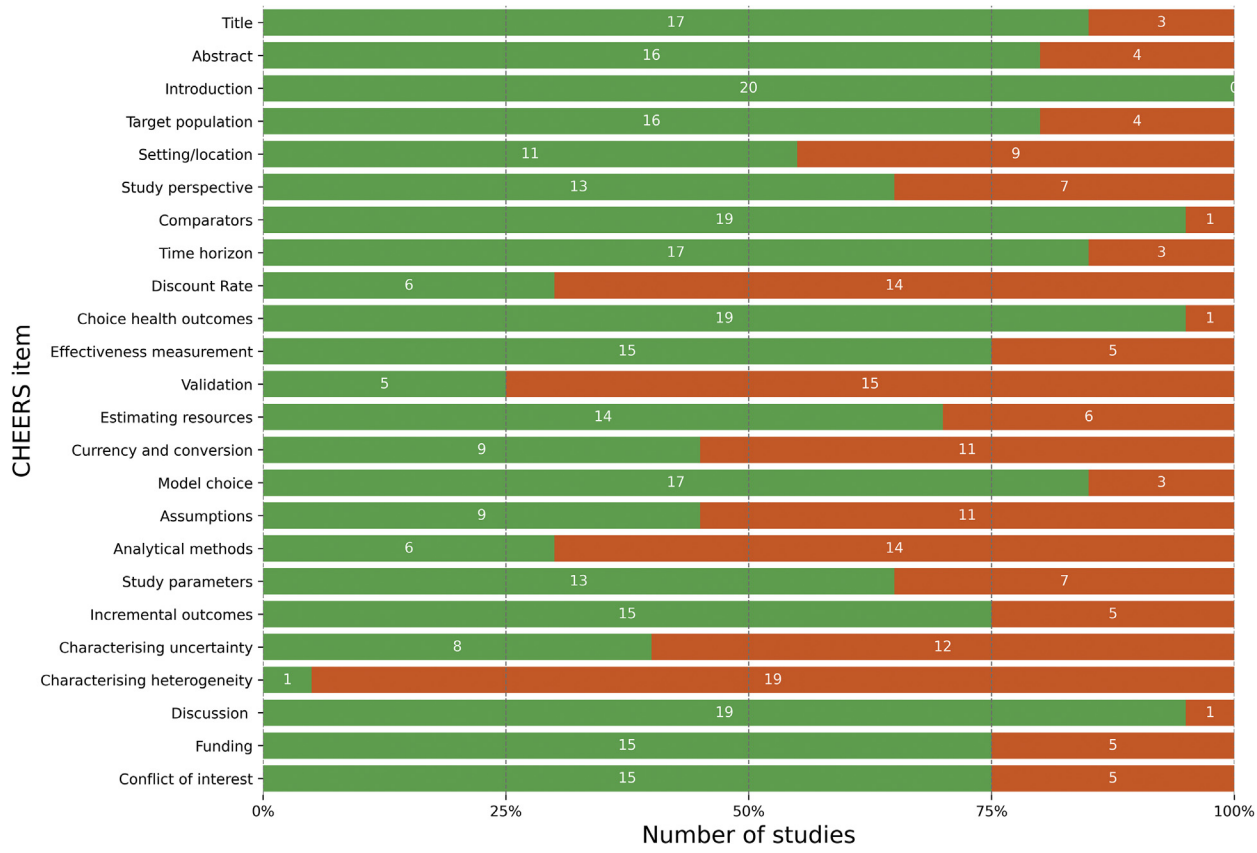
Discussion

In this systematic review, 20 studies were identified that reported on a HEE of an AI application in healthcare. Given the large total number of studies describing the development of AI (over 120 000 in 2019) or the number of HEEs (nearly 20 000 in 2016 alone), 20 included studies are quite a limited number.[39,40] The most common AI technology was automated medical image analysis in a variety of care pathway phases. The majority of studies (n = 17) compared an AI application with usual care and concluded the AI to be cost saving. Additionally, a large number of studies reported no details regarding characterizing uncertainty (n = 12), model assumptions (n = 11), and analytical methods (n = 14).

These limitations may be attributed to the choice of health economic method, because the included studies used relatively simple modeling methods, such as Markov models and decision trees. Therefore, one important finding of this study is that current HEEs of AI applications are unfortunately both quantitatively and qualitatively limited. The fact that only few assessments are published, often of suboptimal quality, may severely hinder the adoption of AI into clinical practice. Considerable challenges still need to be overcome to progress beyond the generally limited adoption in individual institutions and achieve its full potential.[9,41]

The continuous development of innovations is at the core of improving outcomes and affordability in healthcare. There is an

Figure 2. Overview of proportion of studies reporting CHEERS checklist items. Green, yes; red, no.



CHEERS indicates Consolidated Health Economic Evaluation Reporting Standards.

abundance of exiting AI innovations and inspiring initiatives, but clinical practice and policy makers are asking for suitable methodologies and outcomes measures relevant to assess (added) value and improve patient care. The availability and applicability of these outcomes measures and methodologies are the basis on which appropriate research can be performed and the collection of evidence can be further improved. Hence, continuous developments in AI require an accompanying regulatory framework. Although the Food and Drug Administration (FDA), the European Commission, and many European countries individually are developing strategies and policies to regulate the development of AI, the process is time consuming.[2,42] The disparity between AI development and implementation stems from the fact that AI is uniquely situated among healthcare innovations because of its ability to learn and improve performance from experience through retraining. Currently, the regulatory groundwork for AI applications as a medical device considers AI to be static models, meaning that alterations (ie, retraining) are difficult to regulate. To this end, the FDA is developing a framework to allow adjustments to AI applications and support a total product life cycle approach.[42] In Europe, since the introduction of the Medical Device Regulation in May 2020, approval for dynamic AI exists, yet in many situations still requires renewed risk assessment.

In these proposals, the FDA and European Commission expect AI applications to demonstrate analytical and clinical validation, yet validation guidelines have not been established.[43] Thus AI researchers do not know when AI performance is acceptable in the clinical validation setting or if the AI needs further adjustments. The quality of data from daily clinical practice may be much lower given that healthcare professionals do not always collect all the necessary clinical information in real time.[43] Therefore, an AI ready for clinical adoption should be able to manage low quality data adequately, but this is not explicitly addressed by the proposed frameworks. This is reflected in this systematic review, given that only 5 included studies mentioned validation of their AI[20,25,30,33,34] and data uncertainty (n = 8) and heterogeneity (n = 1) were greatly underreported. Thereupon, until regulatory policy is adopted for AI, the translation of AI toward clinical practice will remain fragmented and the incentive for qualitatively thorough evaluations remains low. Currently, no governing body has clear and definitive guidelines on the admission procedure for AI applications. The traditional paradigm of regulation of medical devices was not designed for adaptive AI technologies, resulting in inadequate high-quality evidence to support clinical implementation.[44,45] Several of the included studies affirmed the fragmented translation. They accentuated how AI implementation requires organizational development,[34] how evidence about long-term costs and benefits is incomplete,[21] and that they were the first to conduct a HEE in their field.[27]

Nonetheless, the number of Conformité Européene–marked and FDA-approved AI-based medical devices is increasing substantially, indicating that there is a degree of scientific evidence showing safety and effectiveness available.[46] Nevertheless, few of the currently approved AIs have yet been proven to be "value for money" from a societal perspective, given that this requires

Table 3. Health economic methodological details of included studies (N = 20).

Main author	HEE type	Intervention	Comparator	Perspective	Discount rate	Time horizon	Outcome measure
Philipsen et al[19]	CMA	Automated chest radiography	Human reading	Payer	*	*	CSS, CNTBC
Tufail et al[20†]	CEA	Replacing human graders with automated grading	Automated grading before human grading	Payer	3.5%	12 months	Outcome of cost per appropriate screening outcome
Fernandez-Vicente et al[31]	CMA	MCS with CAD 3D printing	Traditional MCS	Hospital	*	12 months	Cost reduction in MCS
Takahashi et al[32†]	CEA	CAD alone for mammography and computed tomographic colonography	2 human readers, 1 human reader + CAD	Societal	3%	Lifetime	Incremental expected life expectancy
Mervin et al[33]	CEA	Therapeutic pet-type robot	AI-disabled pet-robot, usual care	Hospital	*	10 weeks	Incremental effectiveness of unit improvement in agitation level
Bremer et al[34]	CEA	Combined internet-based personalized plus face-to-face treatment	Treatment as usual (face-to-face)	Societal	*	12 months	Incremental effectiveness of individual treatment recommendations
Gönel et al[35]	CMA	5 algorithms to eliminate unnecessary tests	Ordering tests without algorithms	Hospital	*	45 days	Cost savings of laboratory tests
Golas et al[36]	CMA	Deep learning prediction model	Traditional telemonitoring	Hospital	*	30 days	All-cause 30-day readmission
Padula et al[37†]	CUA	AI-based risk stratified prevention based on Branden scores	All-level risk assessment, standard care	Societal and hospital	3%	12 months	Incremental QALYs
Lee et al[38†]	Total cost	AI predictive model for readmission risk	Standard logistic regression	*	*	90 days	Total cost of TJR post discharge care and readmissions
Fuller et al[21†]	CUA	ARIAS	Current standard of care	Payer	3%	5 years	Incremental QALYs
Mansour et al[22]	CEA	Automated cerebrovascular CT image analysis	CT image scoring on mobile device	Hospital	*	90 days	Outcomes after IV thrombolysis
Murtojärvi et al[23]	BIA	Survival prediction feature selection with Greedy Cox	Feature selection with LASSO	Hospital	*	*	Incremental effectiveness overall survival
Eigner et al[24]	Profit	Decision algorithm for optimal time of patient discharge	*	Hospital	*	28 days	Prediction of readmission risk
Hill et al[25†]	CUA	Targeted screening for atrial fibrillation	Opportunistic, systematic screening	NHS (payer)	3.5%	Lifetime	Cost per QALY gained
Rozenblum et al[26]	CMA	AI-based CDS for medication errors	Standard CDS	Hospital	*	2 years	Cost savings of medication errors
Mori et al[27]	CMA	AI-supported colorectal endoscopy plus diagnose-and-leave	No use of AI, resect-all-polyps strategy	Payer	*	12 months	Cost reduction per colonoscopy and gross annual reimbursement
Xie et al[28†]	CMA	Screening for diabetic retinopathy with 1 of 2 DLS models	Standard screening program with manual grading	Health System	*	12 months	Annual cost savings
Wolf et al[29†]	CMA	Autonomous AI screening	Clinician-based screening	Patient	*	12 months	Reduction of patient out-of-pocket payments
Schwendicke et al[30†]	CEA	Bitewing radiographs with AI	Bitewing radiographs without AI	Payer	3%	Lifetime	Cost for tooth retention years

3D indicates 3-dimensional; AI, artificial intelligence; ARIAS, Automated Retinal Image Analysis System; BIA, budget-impact analysis; CAD, computer aided diagnosis; CDS, clinical decision system; CEA, cost-effectiveness analysis; CMA, cost-minimization analysis, CNTBC, cost per notified tuberculosis case; CSS, cost per screened subject; CT, computed tomography; CUA, cost-utility analysis; HEE, health economic evaluation; IV, intravenous; LASSO, least absolute shrinkage and selection operator; MCS, mould casting splinting; NHS, National Health Service; QALY, quality-adjusted life-year; TJR, total joint replacement.
*Item not reported.
†Model-based study, included in Table 4.[20,21,25,28-30,32,37,38]

Table 4. Summary of Philips' main items of health economic modeling studies (n = 9).

Main author	Model type	HEE type	Model states/ tree summary	Time horizon, cycle length	Sensitivity analysis	Outcome	Result
Tufail et al[20]	Decision tree	CEA	2 screening pathways	1 year, NA	Deterministic, 1-way	£4.51, £11.81 (strategy 1); £2.80, £9.71 (strategy 2)	Strategy 1 preferred
Takahashi et al[32]	Markov model	CEA	Diagnosis, treatment, follow-up; diagnosis, after detection, treatment	Lifetime, 2 years	Univariate	ICERs are ¥268,181/life and ¥163,971/life (BC) $18,980/life and $16,336/life (CRC)	Intervention dominant
Padula et al[37]	Markov model	CUA	State transition, 5 risk categories until discharge	1 year, 1 day	Univariate 1-way, 2-way and probabilistic	$2000/QALY (societal), $2142/QALY (hospital)	Intervention dominant
Lee et al[38]	Decision tree	Cost	High, Medium, Low risk	90 days, NA	3-way	Total costs $651 490, $1 994 654 and $963 550	Lowest total cost for high-risk patients with home-service
Fuller et al[21]	Markov model	CUA	Referred, not referred; adherent, nonadherent	5 years, NA	1-way and probabilistic	$258 721.81/QALY	Intervention dominant
Hill et al[25]	Decision tree + Markov model	CUA	AF, no AF. Targeted, systematic or opportunistic screening	Lifetime, NA	Univariate	£4847/QALY (systematic) and £5544/QALY (opportunistic)	Intervention dominant
Xie et al[28]	Decision tree	CMA	Fully automated, semiautomated, human assessment; screened/ unscreened	1 year, NA	Deterministic, 1-way	Total costs US$62 (semiauto), $66 (fully-auto) per patient per year. Annual savings $489 000	Semiautomated screening preferred
Wolf et al[29]	Decision tree	CMA	Autonomous AI, ECP; screening/ examination; diagnosed, not diagnosed	1 year, NA	1-way and probabilistic	Patient payment $8.52 for T1D and $10.85 for T2D (AI), $7.91 for T1D and $8.20 for T2D (ECP)	Autonomous AI screening preferred
Schwendicke et al[30]	Markov model	CEA	True- and false-positive and true- and false-negative detections	Lifetime, 1 year	Univariate and probabilistic	−13.9 Euros/year	Intervention dominant

AF indicates atrial fibrillation; AI, artificial intelligence; BC, breast cancer; CEA, cost-effectiveness analysis; CMA, cost-minimization analysis; CRC, colorectal cancer; CUA, cost-utility analysis; ECP, eye care professional; HEE, health economic evaluation; ICER, incremental cost-effectiveness ratio; NA, not applicable; QALY, quality-adjusted life-year; T1D, type 1 diabetes; T2D, type 2 diabetes; US, United States.

performing a formal HEE.[47] Of the 6 FDA- or Conformité Européene–approved AIs included in this review, only 1[29] is currently adopted in a clinical setting.[19-22,27] Between market approval and clinical adoption, the question on how AI applications can best be deployed as an integrated part of a healthcare system rises.[48] If the AI application needs to be reimbursed by health insurance, an economic evaluation becomes increasingly important to provide insight into the health gained and the timeframe in which costs are incurred and benefits are used. Unfortunately, the proposed regulatory frameworks for AI do not explicitly mention HEE, and therefore, it remains unclear whether HEE will become a necessary condition for health insurance reimbursement. A positive HEE also does not guarantee the adoption of AI in clinical practice. The current absence of regulatory frameworks could explain why so few HEEs are reported. Furthermore, a previous systematic review published by Wolff et al[49] in early 2020 evaluating economic impact of AI in healthcare, similarly reported a scarcity of publications conducting extensive and qualitatively sound economic impact evaluations. Comparable with findings in our study, this previous review concluded that the economic evaluations of AI in healthcare often focused on specific elements within health economics (eg, only included direct costs) rather than performing a comprehensive analysis.

Compared with many other interventions, the primary goal of an AI may not be to improve health outcomes but instead to improve other outcomes, such as shared decision making, well-being, or patient independency. This makes performing HEEs more challenging because such outcomes are notoriously difficult to value, for example, in monetary terms. Nevertheless, the society's expectations are growing toward reflecting the broader benefits of healthcare interventions, not captured in traditional HEEs. Therefore, future HEEs of AI need to include additional benefits other than in standard health dimensions, so that these benefits may weigh in as contextual factors in decision making.[50]

Furthermore, Guo et al[51] identified a gap in the evaluation methods used in practice for digital health innovations including AI-based software as a medical device. The wider use of different types of simulation approaches such as computational, clinical, and system simulation is needed to overcome limitations and support better decision making. Applying advanced simulation methods such as system dynamics, discrete event, and agent-

based simulation maintaining a complex system view could provide a methodology highly capable of modeling the effects of a complex intervention such as AI from a systems perspective.[52,53] In this systematic review, 9 articles included simulation-based modeling and many mentioned limitations of their simulation models. These limitations will continue to persist unless new and advanced simulation methods are used for the evaluation of advanced and complex technologies such as AI. Other limitations discussed were the underestimations of costs. Costs not directly linked to the assessed invention were often not considered in the evaluation. For example, costs incurred from increased staff time, physician training, or software updates were not included.[19,26] Equivalently, excluding future health benefits resulting from effective interventions leads to an underestimation of benefits.

Limitations

Despite the development of AI taking off in the last 2 decades, the number of scientific publications has been increasing even faster in recent years and justifies our search for publications since 2016.[9] Nevertheless, it is possible that relevant articles not written in English were missed. Many articles were excluded in this review based on title and abstract screening. Remarkably, we found that many of the articles in our initial search results claimed to describe a cost-effective AI application, yet did not conduct a HEE to justify those claims, leading to a high exclusion rate. Additionally, articles evaluating only patient health outcomes or hospital process outcomes were not included, even though such improved outcomes could lead to a reduction in costs, per the concepts of value-based healthcare.[54] Finally, following the CHEERS and Philips checklists can ensure economic evaluations include the appropriate components, but it does not necessarily reflect the correct implementation of the item. Although the computed score can be questioned for assuming equal weight for each checklist item, it does provide an estimation of the completeness of the evaluation per study.

Conclusions

This systematic review exposes an important gap in the methods used for HEE of AI applications in healthcare. In the context of health economics, the cheetah of AI innovation is only at a slow pace pursued by the tortoise (formal HEE). Currently, HEEs of AI are incapable of capturing the complexities and clinical applicability needed to support appropriate decision making. Unless this tortoise catches up, beneficial AI applications run the risk of not being adopted because of lack of proven health and economic benefits. Moreover, there is a risk of nonvaluable AI applications being adopted based on poor and limited evidence. Both situations lead to potential health loss and unnecessary costs and will likely persist until AI, with its seemingly endless possibilities, is recognized as an intervention that can and should be properly assessed. Therefore, further work to enhance health economic assessment of AI will likely be crucial to their future adoption into clinical practice.

Article and Author Information

Accepted for Publication: November 10, 2021

Published Online: December 16, 2021

doi: https://doi.org/10.1016/j.jval.2021.11.1362

Author Affiliations: Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, The Netherlands (Voets, Siesling, Koffijberg); Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands (Voets, Siesling); Multi-Modality Medical Imaging, Technical Medical Centre, University of Twente, Enschede, The Netherlands (Veltman); Department of Radiology, Ziekenhuisgroep Twente, Almelo, The Netherlands (Veltman); Department of Robotics and Mechatronics, Technical Medical Centre, University of Twente, Enschede, The Netherlands (Slump).

Correspondence: Hendrik Koffijberg, PhD, Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, P.O. Box 217, Enschede, The Netherlands 7500 AE. Email: h.koffijberg@utwente.nl

Author Contributions: *Concept and design:* Voets, Veltman, Siesling, Koffijberg
Acquisition of the data: Voets
Analysis and interpretation of data: Voets, Koffijberg
Drafting the manuscript: Voets, Veltman, Slump, Siesling, Koffijberg
Critical revision of paper for important intellectual content: Voets, Slump, Siesling, Koffijberg
Obtaining funding: Slump, Siesling
Supervision: Veltman, Koffijberg

Conflict of Interest disclosures: The authors reported no conflicts of interest.

Funding/Support: This research is supported by Dutch Cancer Society and the Dutch Research Council Domain Applied and Engineering Sciences as part of their joint strategic research program: Technology for Oncology II. The collaboration project is cofunded by the PPP Allowance made available by Health–Holland, Top Sector Life Sciences & Health, to stimulate public/private partnerships (grant number 17924).

Role of Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgment: Dr Rossen Kazakov contributed to abstract and full-text screening and data extraction of full-text articles.

REFERENCES

1. Wiegand T, Lee N, Pujari S, et al. White paper for the ITU/WHO Focus Group on artificial intelligence for health. ITU. https://www.itu.int/go/fgai4h. Accessed May 24, 2021.
2. Annoni A, Benczur P, Bertoldi P, et al. Artificial intelligence: a European perspective. Publications Office of European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC113826. Accessed September 7, 2021.
3. Pianykh OS, Langs G, Dewey M, et al. Continuous learning AI in radiology: implementation principles and early applications. *Radiology*. 2020;297(1):6–14.
4. Bell D, Gachuhi N, Assefi N. Dynamic clinical algorithms: digital technology can transform health care decision-making. *Am J Trop Med Hyg*. 2018;98(1):9–14.
5. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394(10201):861–867.
6. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15(141):20170387.
7. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
8. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94–98.
9. Gore JC. Artificial intelligence in medical imaging. *Magn Reson Imaging*. 2020;68:A1–A4.
10. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):1–9.
11. O'Rourke B, Oortwijn W, Schuller T, International Joint Task Group. The new definition of health technology assessment: a milestone in international collaboration. *Int J Technol Assess Health Care*. 2020;36(3):187–190.
12. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that

evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
13. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)–explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. *Value Health*. 2013;16(2):231–250.
14. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*. 2006;24(4):355–371.
15. Watts RD, Li IW. Use of checklists in reviews of health economic evaluations, 2010-2018. *Value Health*. 2019;22(3):377–382.
16. Rogers HJ, Rodd HD, Vermaire JH, et al. A systematic review of the quality and scope of economic evaluations in child oral health research. *BMC Oral Health*. 2019;19(1):1–15.
17. Mihalopoulos C, Chatterton ML. Economic evaluations of interventions designed to prevent mental disorders: a systematic review. *Early Interv Psychiatry*. 2015;9(2):85–92.
18. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15(12):1295–1310.
19. Philipsen RH, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. *Sci Rep*. 2016;5(1):1–8.
20. Tufail A, Kapetanakis VV, Salas-Vega S, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess*. 2016;20(92):1–72.
21. Fuller SD, Hu J, Liu JC, et al. Five-year cost-effectiveness modeling of primary care-based, nonmydriatic automated retinal image analysis screening among low-income patients with diabetes [published online October 20, 2020]. *J Diabetes Sci Technol*. https://doi.org/10.1177/1932296820967011.
22. Mansour OY, Ramadan I, Abdo A, et al. Deciding thrombolysis in ais based on automated versus on whatsapp interpreted ASPECTS, a reliability and cost-effectiveness analysis in developing system of care. *Front Neurol*. 2020;11:333.
23. Murtojärvi M, Halkola AS, Airola A, et al. Cost-effective survival prediction for patients with advanced prostate cancer using clinical trial and real-world hospital registry datasets. *Int J Med Inform*. 2020;133:104014.
24. Eigner I, Bodendorf F. Decision support for patient discharge in hospitals – analyzing the relationship between length of stay and readmission risk, cost, and profit. In: Ferreira JE, Palanisamy B, Ye K, Kantamneni S, Zhang LJ, eds. *World Congress on Services*. Cham, Switzerland: Springer International Publishing; 2020:77–84.
25. Hill NR, Sandler B, Mokgokong R, et al. Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm. *J Med Econ*. 2020;23(4):386–393.
26. Rozenblum R, Rodriguez-Monguio R, Volk LA, et al. Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. *Jt Comm J Qual Patient Saf*. 2020;46(1):3–10.
27. Mori Y, Kudo SE, East JE, et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). *Gastrointest Endosc*. 2020;92(4):905–911.e1.
28. 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.
29. 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.
30. Schwendicke F, Rossi JG, Göstemeyer G, et al. Cost-effectiveness of artificial intelligence for proximal caries detection. *J Dent Res*. 2021;100(4):369–376.
31. Fernandez-Vicente M, Chust AE, Conejero A. Low cost digital fabrication approach for thumb orthoses. *Rapid Prototyp J*. 2017;23(6):1020–1031.
32. Takahashi R, Kajikawa Y. Computer-aided detection: cost effectiveness analysis with learning model. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). IEEE Xplore. https://ieeexplore.ieee.org/document/8125306. Accessed June 27, 2021.
33. Mervin MC, Moyle W, Jones C, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. *J Am Med Dir Assoc*. 2018;19(7):619–622.e1.
34. Bremer V, Becker D, Kolovos S, et al. Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: data-driven analysis. *J Med Internet Res*. 2018;20(8):e10275.
35. Gönel A. Clinical biochemistry test eliminator providing cost-effectiveness with five algorithms. *Acta Clin Belg*. 2020;75(2):123–127.
36. 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):1–17.
37. Padula WV, Pronovost PJ, Makic MBF, et al. Value of hospital resources for effective pressure injury prevention: a cost-effectiveness analysis. *BMJ Qual Saf*. 2019;28(2):132–141.
38. Lee HK, Jin R, Feng Y, et al. An analytical framework for TJR readmission prediction and cost-effective intervention. *IEEE J Biomed Health Inform*. 2019;23(4):1760–1772.
39. Zhang D, Mishra S, Brynjolfsson E, et al. Artificial intelligence index report 2021. Stanford University Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf. Accessed July 9, 2021.
40. Jakovljevic MM, Pejcic AV. Growth of global publishing output of health economics in the twenty-first century: a bibliographic insight. *Front Public Health*. 2017;5:211.
41. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res*. 2017;19(11):e367.
42. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI / ML) -Based software as & medical device (SaMD) - discussion paper and request for feedback. Digital health regulatory policies. United States Food & Drug Administration. https://www.fda.gov/media/122535/download. Accessed June 27, 2021.
43. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res*. 2018;2(1):1–11.
44. Sendak MP, D'Arcy J, Kashyap S, et al. A path for translation of machine learning products into healthcare delivery. *EMJ Innov*. 2020;10:19–72.
45. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. *J Med Internet Res*. 2021;23(4):e25759.
46. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *Lancet Digit Health*. 2021;3(3):e195–e203.
47. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol*. 2021;31(6):3797–3804.
48. Gold HT, Pitrelli K, Hayes MK, Murphy MM. Decision to adopt medical technology: case study of breast cancer radiotherapy techniques. *Med Decis Making*. 2014;34(8):1006–1015.
49. 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.
50. Wester V, Huygens S, Versteegh M. Early HTA of the value of a decision-supporting AI-application in Multiple Sclerosis. *Institute for Medical Technology Assessment Erasmus University of Rotterdam*. Published online 2021: 1-38.
51. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. *NPJ Digit Med*. 2020;3(1):110.
52. Marshall DA, Burgos-Liz L, Ijzerman MJ, et al. Applying dynamic simulation modeling methods in health care delivery research - the SIMULATE checklist: report of the ISPOR simulation modeling emerging good practices task force. *Value Health*. 2015;18(1):5–16.
53. Marshall DA, Burgos-Liz L, Ijzerman MJ, et al. Selecting a dynamic simulation modeling method for health care delivery research - part 2: report of the ISPOR dynamic simulation modeling emerging good practices task force. *Value Health*. 2015;18(2):147–160.
54. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. *Acad Med*. 2020;95(5):682–685.