Sağlık Yönetimi Dergisi

SDÜ Sağlık Yönetimi Dergisi, Yıl: 2025, Cilt: 7, Sayı: 1, 44-59.

SDU Healthcare Management Journal, Year: 2025, Volume: 7, No: 1, 44-59.

ISSN: 2757-5888

DERLEME MAKALESİ / REVIEW ARTICLE

IMPACT OF ARTIFICIAL INTELLIGENCE ON HEALTH ECONOMIC EVALUATIONS

SAĞLIK EKONOMİSİ DEĞERLENDİRMELERİNDE YAPAY ZEKANIN ETKİSİ

Dr. Tuba SAYGIN AVŞAR [1]

ABSTRACT

Artificial intelligence (AI) is revolutionising healthcare by enhancing diagnostic accuracy, personalising treatment, and improving operational efficiency. Economic evaluations, critical for assessing healthcare interventions, increasingly focus on AI-based solutions to balance costs and benefits effectively. High initial costs, uncertain long-term outcomes, and the evolving nature of adaptive algorithms challenge the economic evaluation of these technologies. AI also serves as a transformative analytical tool, enabling processing of large datasets, predictive modelling, and automation of systematic reviews. However, algorithmic bias, equity concerns, and resource constraints highlight the need for hybrid approaches, integrating AI with local expertise. The study provides an overview of the recent developments in health economics and outcomes research, focusing on the dual role of AI as both intervention and enabler. Ethical considerations, robust reporting, and capacity building are vital for leveraging AI's potential to optimise resource allocation, improve health outcomes, and ensure equitable access to advanced technologies.

Keywords: Health Economics, Artificial Intelligence, Economic Evaluation, Health Technology Assessment.

ÖZET

Yapay zeka (YZ), teşhislerin doğruluğunu artırarak, tedavileri kişiselleştirerek ve operasyonel verimliliği artırarak sağlık hizmetlerinde bir devrim yaratmaktadır. Ekonomik değerlendirmeler, sağlık hizmetleri müdahalelerini değerlendirmek için kritik olup, maliyetleri ve faydaları etkili bir şekilde dengelemek amacıyla giderek daha fazla yapay zeka tabanlı çözümlere odaklanmaktadır. Yüksek başlangıç maliyetleri, belirsiz uzun vadeli sonuçlar ve algoritmaların gelişen doğası bu teknolojilerin değerlendirilmesinde zorluklar ortaya çıkarmaktadır. YZ aynı zamanda büyük verilerin işlenmesine, ekonomik modelleme ve sistematik incelemelerin otomasyonu için dönüştürücü bir analitik araç olarak da hizmet veriyor. Ancak algoritmik önyargı, eşitlik endişeleri ve kaynak kısıtlamaları gibi zorluklar,YZ'yi yerel uzmanlıkla bütünleştiren yaklaşımlara olan ihtiyacı vurgulamaktadır. Çalışma, YZ'nin sağlık ekonomisi ve çıktıları yazınında hem bir müdahale hem de kolaylaştırıcı olarak kullanılmasını iceren ikili rolüne odaklanarak son gelişmelere kapsamlı bir genel bakış sunmaktadır. Etik analizler, sağlam raporlama ve kapasite geliştirme, YZ'nin etkili kaynak tahsisi, sağlık çıktılarını iyileştirme ve ileri teknolojilere adil erişim sağlama potansiyelinden yararlanmak için hayati öneme sahiptir.

Anahtar kelimeler: Sağlık ekonomisi, Yapay Zeka, Ekonomik Değerlendirme, Sağlık Teknolojisi Değerlendirmesi

[1] Institute for Global Health, University College London, tubasaygn@gmail.com, ORCID: 0000-0002-4143-3852

1. INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative technology with applications across diverse fields, driving innovation, efficiency, and new capabilities. By leveraging advanced algorithms, machine learning, and big data, AI systems can analyse vast amounts of information, identify patterns, and make predictions or decisions that were previously beyond human capabilities (Howell et al., 2024). In healthcare, AI-based interventions provide a wide range of benefits, including enhancing diagnostic accuracy and personalising treatment plans.

Economic evaluation in healthcare is a systematic approach for comparing the costs and outcomes of different interventions or programs to determine their value and guide resource allocation decisions (Drummond et al., 2015). Common types include cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA), each focusing on different metrics (e.g., life years gained, quality-adjusted life years) or monetary benefits. These evaluations are crucial in healthcare because resources are limited, and decision-makers must prioritise interventions that maximise health benefits for the population while minimising costs. For example, an economic evaluation might compare incremental cost per life saved when a new vaccine is provided compared to not providing it, helping policymakers allocate funding more effectively. By providing a structured approach to assessing value, economic evaluations ensure that investments in healthcare deliver the greatest possible impact, promote equity, and improve overall health system efficiency.

The interplay between AI and economic evaluations reflects a dual influence: the economic assessment of AI-based health interventions and the utilisation of AI as a tool to enhance traditional economic evaluation methods. Evaluating AI-based health interventions involves analysing their cost-effectiveness, scalability, and long-term impacts (Vithlani et al. 2023). Such evaluations consider the costs of development, implementation, and maintenance against measurable benefits like improved health outcomes, reduced error rates, or increased operational efficiency.

AI also revolutionises economic evaluations by acting as a powerful analytical tool (Fleurence et al., 2024a). It enables the processing of large datasets, improves predictive modelling, and supports scenario analyses, making it possible to capture complex, dynamic systems and personalised factors in economic studies. For instance, AI can simulate the economic impact of policy changes across diverse populations or predict the long-term cost implications of adopting AI-driven technologies. While AI thus enhances both the evaluation of interventions and the evaluation process itself, challenges such as ethical concerns, algorithmic bias, and resource constraints must be carefully managed to maximise its potential and ensure equitable outcomes. This study provides an overview of the recent literature, exploring how AI is shaping economic evaluations in healthcare.

2. CHALLENGES IN CONDUCTING ECONOMIC EVALUATIONS OF AI-BASED HEALTH INTERVENTIONS

As AI technologies continue to transform healthcare, assessing their economic impact remains a complex task. Unlike conventional medical interventions, AI-based solutions often function as decision-support tools rather than direct treatments, making their value difficult to quantify. Traditional health economic evaluation frameworks may not fully capture AI-based interventions' unique cost structures, evolving performance, and indirect benefits, such as workflow improvements and early disease detection.

Conducting economic evaluations of AI-based interventions poses unique challenges as evidenced in two recent systematic reviews (Voets et al. 2022; Vithlani et al. 2023). These include high upfront costs and uncertain long-term value, difficulties in measuring health outcomes, the adaptability of AI algorithms, bias and health equity concerns, and regulatory constraints affecting real-world implementation. Addressing these challenges is essential for ensuring that economic evaluations provide accurate and meaningful insights for policymakers, healthcare providers, and technology developers. The reviews also highlighted that most economic evaluations of AI-based interventions did not report important details especially regarding the specifics of algorithm development (Voets et al. 2022; Vithlani et al. 2023). Thus, recently CHEERS-AI was developed to provide reporting guidance for health economists, editors and policymakers (Elvidge et al. 2024). The following sections outline these challenges as demonstrated in recent studies.

2.1. High Upfront Costs and Uncertain Long-Term Value

Firstly, AI-based interventions usually accrue high upfront costs, and their long-term value is often uncertain. These initial expenditures are often compounded by the need for ongoing maintenance, updates, and integration into existing systems (Khanna et al. 2022). Costs not directly linked to the technology, such as increased staff time, physician training, or software updates, are usually overlooked in evaluations (Wolff et al. 2020). Excluding these important cost components would result with overestimating the cost-effectiveness of these interventions. Despite these substantial initial costs, the long-term value of AI solutions can be difficult to predict, as their success depends on various factors, including the quality of data, adaptability to changing environments, and alignment with organisational structures (Kelly et al. 2019). High initial investments and uncertain outcomes significantly complicate the economic evaluation of AI-based interventions, primarily due to challenges in accurately estimating costs, benefits, and risks over time.

Substantial upfront expenditures for the development, data processing, infrastructure, and training, can skew cost-effectiveness analyses toward higher short-term costs. The long-term value of AI systems depends on factors such as changes in operational contexts and the quality of implementation, which are difficult to predict. Assessing the long-term value of AI-based interventions in health economic evaluations presents several methodological challenges, as demonstrated by a systematic review of 29 economic evaluations (Kastrup et al. 2024). First, while most evaluations defined research questions clearly, only 28% explicitly quantify economic importance, often using vague terms like "costly" instead of concrete cost data. Second, although 66% described methods for estimating costs, less than half reported resource use separately, making it difficult to determine long-term costs. Finally, none of the studies accounted for crucial long-term factors such as the learning curve, incremental innovation, dynamic pricing, or organisational impact, which are essential for evaluating the sustainability of AI-driven interventions.

2.2. Challenges in Measuring Outcomes

Measuring outcomes of AI-based interventions can be challenging due to the indirect nature of their benefits. Unlike traditional medical treatments, where direct clinical outcomes such as recovery rates or reduced mortality can be measured, many AI tools function as decision-support systems or workflow enhancers, making their impact on patient outcomes more difficult to quantify. This might lead health economists to use a proxy measure. For instance, the study by Hassan et al. (2023) used detection rates as an outcome measure in their cost-

ISSN: 2757-5888

effectiveness analysis. This approach can overestimate the effect of the technology and underestimate the Incremental Cost-Effectiveness Ratio (ICER), leading to potential misallocation of healthcare resources. Similarly, in another economic evaluation (Schwendicke et al. 2021), which assessed the cost-effectiveness of using artificial intelligence for proximal caries detection on bitewing radiographs, the primary outcome measure was tooth retention time, which serves as a proxy for the overall effectiveness of the AI technology in dental care. This approach can lead to challenges in accurately capturing the full impact of the intervention on patient health outcomes and cost-effectiveness.

Most AI-based interventions are reported to provide improvements in efficiency, accuracy, or accessibility, which may ultimately contribute to better health outcomes but are not always straightforward to measure. For example, it was shown that estimating the value of an algorithm that converts electronic documents into an online format that can be read and analysed by users can be challenging (Bongurala et al., 2024). This is because an algorithm that converts electronic documents into an online format may improve workflow efficiency and data accessibility, but translating these improvements into measurable health or economic outcomes is complex. A study demonstrated how an AI tool that assists radiologists in identifying early signs of lung cancer on imaging could reduce diagnostic delays, indirectly improving patient survival rates by enabling earlier treatment (Gandhi et al., 2023). In another economic evaluation, an algorithm was found to reduce the number of people being screened for atrial fibrillation, compared to traditional screening methods, and the authors concluded that this would consequently result in more cost-effective use of healthcare resources (Hill et al., 2020). However, it is not always possible to estimate these indirect benefits. Thus, for an AI intervention that is not itself a treatment, economic evaluations should explain the indirect mechanism by which it generates an effect on health outcomes (Elvidge et al. 2024).

2.3. Algorithmic Adaptability and Its Impact on Economic Evaluations

The nature of the algorithm itself can significantly complicate economic evaluations. A static algorithm, which does not change or adapt after deployment, would have relatively predictable performance and cost-effectiveness over time. In contrast, an intervention that incorporates a learning or adaptive AI component introduces additional complexity, as its effectiveness can evolve by continuously learning from data collected during its use (Yin et al., 2021). This ongoing learning process may lead to improved accuracy, better outcomes, or more efficient resource utilisation, which in turn can significantly influence the intervention's cost-effectiveness (Kastrup et al. 2024). For instance, in the economic evaluation by Hill et al. (2020), the machine learning model's ability to learn from new data improved its effectiveness in screening for atrial fibrillation. The AI described in that study was adaptive in the sense that it used machine learning techniques to continuously improve its predictive performance based on both static and dynamic risk factors. This adaptability enhanced the effectiveness of the screening process by accurately identifying high-risk individuals, optimising the use of healthcare resources.

However, this adaptability also introduces uncertainty into economic evaluations (Aung et al. 2021). Traditional economic evaluation methods, such as CEA and CUA, usually rely on static assumptions about the performance and costs of a technology. These methods typically assume that the technology's effectiveness and costs remain constant over the evaluation period. When evaluating adaptive AI, economic models need to account for assumptions about how the learning process will progress, how rapidly the AI system will improve, and whether these improvements plateau over time. Moreover, the cost of ongoing data collection, computational resources for retraining the model, and potential regulatory updates must also be considered as part of the economic evaluation. To ensure robust and transparent evaluations, economic models should clearly document any assumptions about how the learning process is expected to influence outcomes and costs (Elvidge et al., 2024). This includes detailing the anticipated timeline of improvements, the conditions under which learning might fail, and the broader system-level effects, such as changes in clinical workflows or patient behaviours. Transparent reporting of these assumptions is essential for stakeholders to understand the potential risks and benefits of implementing learning-based AI systems in healthcare.

2.4. Algorithmic Bias and Health Inequality Risks

Another important consideration is the risk for algorithmic bias and the potential implications for health inequalities. Algorithmic bias happens when a machine learning algorithm produces outcomes that are systematically unfair or prejudiced against certain groups or individuals, often due to biased data, flawed design choices, or unintended consequences of the algorithm's decision-making processes (Panch et al. 2019; Ratwani et al. 2024). AI models can inadvertently reinforce biases present in their training data, leading to unequal outcomes across different patient groups. Economic evaluations should consider these biases and their implications. For example, an AI-based tool developed for detecting skin cancers might be more effective in people with lighter skin tones compared to those with darker tones (Adamson et al. 2018). In one study, AI was able to identify racial information based on only patients' vital signs through a mechanism that researchers are yet able to explain, which might lead to biased clinical decision-making (Velichkovska et al., 2023). Economic evaluations should discuss these potential biases and consider their impact on health inequalities, especially if the intervention might worsen disparities. However, in the systematic review by Kastrup et al. (2024) only eight economic evaluations reported details of the subjects from whom valuation was obtained. This lack of detailed reporting can obscure potential biases in the data, which may affect the generalisability and fairness of the AI-based interventions.

In the economic evaluation of targeted screening for atrial fibrillation using a machine learning risk prediction algorithm, Hill et al. (2020) addressed the issue of algorithmic bias by estimating cost-effectiveness in different demographic groups. The findings highlight the importance of considering demographic variability to avoid exacerbating health inequalities. In another economic evaluation, which assessed automated retinal image analysis system for diabetic retinopathy screening in the UK, the screening performance varied with patient's age and ethnicity (Tufail et al., 2016). This indicates that differences in retinal images across ethnic groups can affect the accuracy of AI models, potentially leading to biased outcomes and exacerbation of health inequalities.

2.5. Regulatory and Decision-Making Constraints

Although AI is now extensively used in healthcare, in many countries, legal regulations do not allow AI to act independently when providing healthcare services while the discussions continue on whether it is permissible for AI to make decisions itself when there is strong evidence on clinical and cost-effectiveness (Gerke et al., 2022; Mainz et al., 2024; Palaniappan et al., 2024). Therefore, they are used as decision support tools rather than decision-making tools. This means that the intervention arm of the economic evaluation should be a human working with AI not the AI itself, which might impact on AI's accuracy (Sele et al., 2024). Recognising this, in an economic evaluation of diabetic retinopathy screening methods, three

ISSN: 2757-5888

comparators were used: standard-of-care (Human Assessment Model), semi-automated model (combines AI and human expertise), and fully-automated model (AI as standalone grading tool) (Xie et. al. 2020).

The final decision typically rests with healthcare professionals who may integrate additional clinical evidence, contextual factors, or their expertise to accept or override the AI's suggestions. The interaction between AI and the human making final decisions makes it challenging to estimate the benefits of AI as the outcomes are influenced not only by the AI's performance but also by the way healthcare professionals use and interpret its outputs (Hua et al., 2024). For example, in an economic evaluation it was concluded that when clinicians adopt AI tools cautiously, they might use the diagnostic or predictive capabilities while ignoring the opportunity to reduce length of stay through earlier discharge recommendations (De Vos et al., 2022). While the AI system might still deliver clinical benefits (improved readmission and mortality rates), its cost-effectiveness could be compromised. Without capturing this benefit, the economic case for the AI intervention weakens substantially, even if clinical outcomes improve. This shows similarities with the evaluations of diagnostic tests and prediction models, where healthcare professionals can overrule the test and model outcomes based on other factors (Giessen et al. 2014). Thus, when evaluating AI systems, it is crucial to consider not only their technical performance (e.g., accuracy, sensitivity, specificity) but also how they are used in clinical practice.

3. UTILISING AI FOR CONDUCTING ECONOMIC EVALUATIONS

Although the utilisation of AI as a tool for economic evaluations is still at its infancy, the idea dates back to early 90s (Gottinger 1991). AI offers enormous potential in health economics and outcomes research from data analysis to creating novel economic models and interpretation of evaluation outcomes. This section explains the key uses of AI that contribute to economic evaluations and highlights relevant studies supporting these advancements.

3. 1. Enhancing Large Dataset Analysis in Economic Modelling

Generative AI, which is a more advanced version of AI, can significantly enhance the analysis of large datasets in economic modelling by improving the efficiency, accuracy, and depth of insights that can be drawn from complex and voluminous data. Healthcare systems generate vast amounts of data, including electronic health records, medical imaging, patient demographics, treatment outcomes, and operational data. Generative AI can help address challenges in analysing large datasets by using its powerful algorithms to process and analyse large datasets more efficiently, identifying trends that might be missed through conventional methods. One key advantage of generative AI in large dataset analysis is its ability to model complex relationships and predict outcomes based on high-dimensional data (Morgernstern et al. 2020). For example, a large language model was able to predict key patient outcomes (e.g.30-day all cause readmission, and length of stay) better than traditional models (Jiang et al., 2023). There are many studies that demonstrate how generative AI aid in data imputation by filling in missing or incomplete data points, which are common in healthcare datasets (Liu et al., 2023). This helps create a more comprehensive and accurate dataset for economic modelling. Overall, generative AI provides a more scalable, dynamic, and nuanced approach to analysing large datasets in economic modelling, which might ultimately lead to more accurate cost-effectiveness analyses and better-informed policy decisions.

ISSN: 2757-5888

3.2. AI for Systematic Literature Reviews in Economic Evaluations

It is also possible to use AI for conducting systematic literature reviews that are often undertaken to identify relevant parameters needed for economic evaluations. Natural language processing algorithms can rapidly screen thousands of abstracts to identify relevant studies based on inclusion and exclusion criteria (Sundaram and Berleant, 2023). For instance, in the study by Robinson et al. (2023), the authors developed Bio-SIEVE, a family of instruction fine-tuned language learning models, which outperformed both ChatGPT and traditional approaches in inclusion/exclusion classification tasks.

Machine learning models can extract key data points and assess study quality with comparable accuracy to human reviewers (Jardim et al., 2022). For instance, a recent study conducted and empirical evaluation of GPT-4's capabilities as a primary screening instrument in healthcare evidence synthesis, analysing performance characteristics within information retrieval frameworks (Landschaft et al., 2024). The study found an almost perfect agreement between an AI-based reviewer and a human reviewer, measured by the Cohen's kappa coefficient. Text mining approaches can analyse patterns across large bodies of literature to identify emerging themes and research gaps that might be missed in traditional manual reviews (O'Mara-Eves et al. 2015). Recent advances in large language models have further enhanced these capabilities by improving the contextual understanding of scientific texts and enabling more nuanced categorisation of research findings (Luo et al., 2024).

Although there are promising examples of utilising AI for various tasks as part of systematic reviews, this is a developing area and studies showed some limitations. For example, in the study by Robinson et al. (2023), it was needed to adapt the model to safety-first scenarios, where the priority is to avoid excluding relevant studies. This indicates a need for further refinement to ensure high recall in critical applications. Additionally, Bio-SIEVE-Multi demonstrated limited success in multi-task training, which suggests that more effective methods are needed to enable the model to handle multiple tasks simultaneously without compromising performance. Overall, the current literature demonstrates that large language models can significantly enhance the efficiency of the systematic review process by automating the abstract screening phase. This can reduce the time and resources required for systematic reviews, making them more feasible in the face of exponentially increasing literature. However, studies report areas for further research and improvement.

3. 3. AI for Economic Model Development and Validation

Generative AI is a powerful tool for replicating existing economic models in healthcare by learning the underlying patterns and structures within these models and then generating new, similar models that retain the original's key characteristics. This is especially useful for validating or extending established models, as it can quickly simulate different scenarios or predict outcomes based on modified assumptions. For instance, Reason et al. (2024) demonstrated that Generative Pre-Trained Transformer 4 (GPT-4) can be utilised in partitioned survival models evaluating interventions in non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC). Once trained, the AI can generate synthetic datasets or model variations that mimic the behaviour of the original model, allowing researchers to test how changes in model assumptions could impact results (Gonzales et al., 2023).

AI can be transformative in updating existing models. For example, Pandey et al. (2024) developed a virtual assistant interface powered by the Claude-3-Opus large language model to interact with and customize Excel-based cost-effectiveness models. Researchers created a Python-based user interface allowing natural language interactions with the model, where the model processed user inputs to generate logic codes that could retrieve or modify Excel data without compromising privacy, successfully completing all 30 test prompts (10 for data retrieval and 20 for data updates) and accurately processing 20 distinct country-specific input sheets. The authors concluded that their AI assistant interface effectively adapts Excel models for country-specific healthcare needs, simplifies exploration of complex models for non-expert users, and could potentially serve as a unified platform for accessing various Excel models in the future. Similarly, Rowlinson et al. (2024) assessed GPT-4's ability to automate the adaptation of an Excel-based cost-effectiveness model for muscle-invasive urothelial carcinoma from a UK to Czech Republic healthcare setting. Researchers prepared the model with improved descriptive text, provided GPT-4 with natural language instructions and tabular data for country-specific adaptations, and found the AI completed 62 of 64 required updates with 100% accuracy on those changes, achieving an overall accuracy score of 97% across various cost categories in just 245 seconds. The authors concluded that large language models demonstrate technical feasibility for automating edits to Excel-based health economic models, showing promise for highly accurate input value modifications when models are clearly structured.

AI can also be used for creating de novo economic models, improving accuracy and efficiency. For instance, Chhatwal et al. 2024 evaluated the feasibility of using Generative AI (specifically GPT-4) to develop a de novo health economic Markov model for hepatitis C treatment through a custom platform called ValueGen.AI. Researchers created a multi-agent pipeline that successfully constructed model structures with 10-15 health states and over 22 transitions, estimated parameters including transition probabilities, costs, and utilities, all while maintaining face validity despite some variability across multiple experimental iterations. The authors concluded that while their work demonstrates the feasibility of AI-driven health economic model development for chronic diseases, further research is needed to reduce variability in the development process and to compare AI-generated outputs with published models.

3.4. AI for Interpreting Economic Evaluation Outcomes

Interpreting the findings of economic models can be complex, particularly when these models involve large datasets or incorporate advanced techniques such as machine learning or AI. Generative AI can be highly valuable in providing insights into the underlying drivers of model outcomes and helping stakeholders understand the implications of various assumptions and scenarios. AI-driven tools can analyse the results of an economic model and highlight the most influential factors contributing to cost-effectiveness or patient outcomes, making it easier for policymakers and healthcare professionals to interpret the findings in a practical context. For example, Swami and Srivastava (2024) demonstrated how generative AI can interpret economic evaluation outcomes for a specific audience such as modelers, clinicians, providers, payers, and public. This study investigated using ChatGPT 4.0 as a tool to translate complex health economic and outcomes research findings into language appropriate for various stakeholders with different levels of expertise. Researchers created a proof-of-concept system with virtual stakeholder profiles and found the AI effectively translated technical metrics like ICER into more accessible language for clinicians and non-technical audiences. The authors concluded that generative AI shows promise for bridging communication gaps in health economic and outcomes research, though further research is needed to refine these approaches and fully realise AI's potential for tailored information dissemination across diverse audiences.

SDÜ Sağlık Yönetimi Dergisi, Yıl: 2025, Cilt: 7, Sayı: 1, 44-59.
SDU Healthcare Management Journal, Year: 2025, Volume: 7, No: 1, 44-59.
ISSN: 2757-5888

3.5. Challenges in AI-Driven Economic Evaluations

Although the use of AI has considerable potential to improve economic evaluations, this is still a developing field and there are some challenges, such as explainability, scientific validity and reliability, bias, and adoption and integration challenges (Fleurence et al., 2024b). Generative AI models, such as large language models, are often described as "black boxes" due to their complex architectures and the vast amounts of data they process. Achieving full explainability in these models remains challenging (Jarke et al., 2024). The outputs generated by AI models, particularly in complex fields like health economics and outcomes research, can vary in terms of accuracy and reliability (Marey et al., 2024). AI-based tools' ability to generate scientifically valid outputs that do not require human oversight is limited, with studies reporting mixed results (Reason et al., 2024; Srivastava et al., 2024). Additionally, AI models can inadvertently perpetuate or even exacerbate existing biases present in their training data (Mittermaier et al., 2023). This includes systemic biases in healthcare systems or the underrepresentation of historically marginalised groups. Biased AI outputs can lead to unequal treatment or resource distribution among populations, potentially impacting health policies and patient outcomes. Finally, integrating AI into existing workflows can be technically and organisationally challenging. There are some additional challenges that must be overcome to integrate automation into economic model development processes. Health economists, health technology assessment (HTA) organisations and other stakeholders might be reluctant to enhance the use of AI for economic evaluations as evidenced in a survey of decision-makers by Heinz et al. (2024). Resistance to change from traditional methods, lack of expertise in AI, and the high costs associated with implementing and maintaining AI solutions can hinder their widespread adoption. The use of AI for conducting economic evaluations is a developing area and there are limited number of published studies (Dolin et al., 2024). Consequently, there a paucity of evidence on how these limitations may affect the results of economic evaluations.
Despite, these challenges, many organisations recognise the potential of AI to reduce the timelines of technology assessments, which would improve timely access to cost-effective health technologies. To guide health economists and other stakeholders who produce and use economic evaluations, the ISPOR taskforce for good practices in machine learning methods for health economics and outcomes research created a checklist including key considerations for evaluating the transparency of ML to stakeholders and decision makers (Padula et al., 2022). The position statement published by National Institute for Health and Care Excellence (NICE, 2024) in the United Kingdom outlines good practices, standards and guidelines to follow when using AI methods for evidence generation. Additionally, in a recent study, Zemplényi et al. (2023) addresses the challenges and proposes solutions for integrating AI-generated evidence into HTA processes, with inputs from HTA and reimbursement decision-makers in Central Europe. The study concludes that the potential of AI to support evidence generation and evaluation in HTA has not been fully utilised. Raising awareness of the benefits and consequences of AI-based methods and encouraging political commitment are necessary to upgrade the regulatory and infrastructural environment and knowledge base required for better integration of AI into HTA decision-making processes.

4. DISCUSSION

This study presents an overview of how AI impacts economic evaluation in health care, addressing AI both as an intervention and as an analytical tool. The existing literature indicates that economic evaluations of AI-based interventions present unique challenges that need to be addressed through methodological innovation. High initial costs of AI-based interventions, uncertain long-term effects, and additional costs related to integration into care processes are

ISSN: 2757-5888

the main factors that can affect the accuracy of such evaluations. In particular, algorithms that learn and adapt over time can lead to significant uncertainties in cost and effectiveness estimates. Economic models need to clearly define the effects of learning processes and potential cost and benefit dynamics. However, the current literature shows deficiencies in terms of such detailed reporting (Voets et al. 2022; Vithlani et al. 2023). For example, the lack of detailed information about algorithm development processes limits the reliability of economic evaluation results. Therefore, the development of the CHEERS-AI reporting guidelines is an important step for health economists, editors, and policymakers (Elvidge et al., 2024).

The use of AI as a tool in health economic evaluations offers significant opportunities in areas such as analysing large datasets, accelerating systematic reviews, and developing economic models (Fleurence et al., 2024b). Generative AI can make economic evaluation processes faster and more effective through their assistance in processes such as missing data completion, trend analysis, and model validation. By leveraging generative AI, health economists can create, replicate, test, and improve existing economic models more efficiently, ultimately leading to more reliable and adaptive models for evaluating healthcare interventions. However, ethical considerations such as impartiality and reliability should be prioritized when using these technologies. For example, the scientific validity of models created by generative AI based on assumptions should be carefully evaluated. ISPOR's good practice checklist for machine learning methods provides an important guide to increase transparency during the adoption of these technologies (Padula et al., 2022).

There were only a few economic evaluations of AI-based interventions from low- and middle-income countries (LMICs) identified in the recent systematic reviews, including studies from China, Malawi, Pakistan, and Turkey (Voets et al. 2022; Vithlani et al. 2023). This is consistent with the low number of studies reporting implementation of AI-based interventions in LMICs (Holmes et al., 2022). There are many challenges in adopting AI-based technologies in LIMIC, including the quality of existing data sources, training and modelling AI solutions based on contextual data; and implementing privacy, security and accountability policies (Lopez et al., 2022). Experience gained from economic evaluations in high-income countries can provide insights for health economists and policy makers in LMICs.

Using AI for economic evaluations in Turkey and other LMICs has several significant implications. AI can significantly reduce the resources needed to conduct economic evaluations by automating data collection and analysis, making these studies more feasible in resource-constrained settings. AI can help address the common problem of missing or incomplete data in LMICs through techniques like predictive modelling and pattern recognition to estimate missing values and identify trends. Machine learning algorithms can help adapt economic models to local contexts by incorporating region-specific variables and patterns, potentially leading to more accurate and relevant evaluations. However, significant challenges remain, such as bias and equity issues associated with the use of AI. Additionally, while AI can reduce some costs, it requires substantial digital infrastructure and technical expertise, which may be limited in LMICs. Therefore, a hybrid approach that combines AI-driven methods with local expertise and traditional evaluation techniques is likely to be the most effective and equitable solution for these settings.

5. CONCLUSIONS

This study highlights the significant role of AI both as an intervention and an analytical tool in health economic evaluations. The research was conducted to explore how AI is transforming economic evaluation processes in healthcare, addressing both opportunities and challenges. The findings indicate that AI-based interventions pose unique economic evaluation challenges, such as high initial costs, uncertain long-term effects, algorithmic biases, and the complexities arising from human-AI interaction. Additionally, there are important gaps in the current literature, particularly the lack of detailed reporting on algorithm development and implementation processes, which can impact the reliability of economic evaluations.

AI's potential to enhance economic evaluation methodologies is evident, particularly in improving data analysis, conducting systematic literature reviews, and developing economic models. Generative AI, in particular, offers promising advancements in automating complex tasks, identifying trends, and improving predictive modelling in healthcare economic evaluations. However, the integration of AI into economic evaluation must be approached cautiously, ensuring transparency, ethical considerations, and fairness in healthcare decision-making processes.

Future research should focus on refining methodological approaches for AI-driven economic evaluations, addressing biases, and developing standardised reporting frameworks such as CHEERS-AI. Additionally, expanding research on AI-based economic evaluations in low- and middle-income countries is crucial to understanding its feasibility and impact in diverse healthcare settings. A hybrid approach that combines AI-driven methods with traditional economic evaluation techniques and expert knowledge is recommended to ensure comprehensive, accurate, and equitable assessments.

Overall, while AI presents immense potential to improve economic evaluations in healthcare, it is imperative to address methodological, ethical, and regulatory challenges to maximise its benefits and ensure equitable healthcare outcomes globally.

DECLARATION OF THE AUTHORS

Contribution Rate Statement: The author contributes the study on her own

Statement of Support and Acknowledgment: No support is taken from any institution or organization.

Conflict Statement: There is no potential conflict of interest in the study.

6. REFERENCES

Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. JAMA Dermatology, 154(11), 1247-248.https://doi.org/10.1001/jamadermatol.2018.2347.

Aung, Y. Y., Wong, D. C., & Ting, D. S. (2021). The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. British Medical Bulletin, 139(1), 4–15. https://doi.org/10.1093/bmb/ldaa043.

Bongurala, A. R., Save, D., Virmani, A., & Kashyap, R. (2024). Transforming health care with artificial intelligence (AI): Redefining medical documentation. Mayo Clinic Proceedings: Digital Health. https://doi.org/10.1016/j.mcpdh.2024.100045.

Chhatwal, J., Yildirim, I. F., Samur, S., Bayraktar, E., Ermis, T., & Ayer, T. (2024). P28 Development of de novo health economic models using generative AI. Value in Health, 27(12), S7. https://doi.org/10.1016/j.jval.2024.08.013.

De Vos, J., Visser, L. A., De Beer, A. A., Fornasa, M., Thoral, P. J., & Elbers, P. W. G. (2022). The potential cost-effectiveness of a machine learning tool that can prevent untimely intensive care unit discharge. Value in Health, 25(3), 359–67.

Dolin, O., Lim, V., Chalmers, K., Hepworth, T., Gonçalves-Bradley, D., Langford, B., & Rinciog, C. (2024). Large Language Models for Health Economics and Health Technology Assessment: A Targeted Review. Value in Health, 27(12), S462.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). Methods for the economic evaluation of health care programmes (4th ed.). Oxford University Press.

Elvidge, J., Hawksworth, C., Avşar, T. S., Zemplenyi, A., Chalkidou, A., Petrou, S., & Wilson, E. (2024). Consolidated health economic evaluation reporting standards for interventions that use artificial intelligence (CHEERS-AI). Value in Health. https://doi.org/10.1016/j.jval.2024.09.007

Fleurence, R., Wang, X., Bian, J., Higashi, M. K., Ayer, T., Xu, H., & Chhatwal, J. (2024a). Generative AI in health economics and outcomes research: A taxonomy of key definitions and emerging applications—An ISPOR working group report. arXiv Preprint. https://doi.org/10.48550/arXiv.2410.20204

Fleurence, R. L., Bian, J., Wang, X., Xu, H., Dawoud, D., Higashi, M., & Chhatwal, J. (2024b). Generative AI for health technology assessment: Opportunities, challenges, and policy considerations — An ISPOR working group report. Value in Health. https://doi.org/10.1016/j.jval.2024.08.012

Gandhi, Z., Gurram, P., Amgai, B., Lekkala, S. P., Lokhandwala, A., Manne, S., & Surani, S. (2023). Artificial intelligence and lung cancer: Impact on improving patient outcomes. Cancers, 15(21), 5236. https://doi.org/10.3390/cancers15215236

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence- driven healthcare. In T. Cohen & G. Cohen (Eds.), Artificial intelligence in healthcare (pp. 295– 336). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00016-4

Giessen, A., Wilcher, B., Peters, J., Hyde, C., Moons, K. G., de Wit, G. A., & Koffijberg, H. (2014). Health economic evaluation of diagnostic and prognostic prediction models: A systematic review. Value in Health, 17(7), A560. https://doi.org/10.1016/j.jval.2014.08.888

Gottinger, H. W. (1991). Artificial intelligence and economic modelling. Expert Systems, 8(2), 99– 105. https://doi.org/10.1111/j.1468-0394.1991.tb00435.x

Hassan, C., Povero, M., Pradelli, L., Spadaccini, M., & Repici, A. (2023). Cost-utility analysis of real-time artificial intelligent-assisted colonoscopy in Italy. Endoscopy International Open. https://doi.org/10.1055/a-2136-3428

Heinz, S., Bondal, E., Kumari, C., and Castellano, G. (2024). HTA141 Artificial Intelligence: Is It a New Era in Payer Decision Making? Value in Health, 27(12), S381.

Hill, N. R., Sandler, B., Mokgokong, R., Lister, S., Ward, T., & Boyce, R. (2020). Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm. Journal of Medical Economics, 23(4), 386–93. https://doi.org/10.1080/13696998.2019.1706543

Holmes, T., Singh, R., Axt, M., Brenner, S., & Barteit, S. (2022). Artificial intelligence for strengthening healthcare systems in low-and middle-income countries: A systematic scoping review. NPJ Digital Medicine, 5(1), 162. https://doi.org/10.1038/s41746-022-00665-w

Howell, M. D., Corrado, G. S., & DeSalvo, K. B. (2024). Three epochs of artificial intelligence in healthcare. JAMA, 331(3), 242–244. https://doi.org/10.1001/jama.2023.25902

Hua, D., Petrina, N., Young, N., Cho, J. G., & Poon, S. K. (2024). Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. Artificial Intelligence in Medicine, 147, 102698. https://doi.org/10.1016/j.artmed.2023.102698

Jardim, P. S. J., Rose, C. J., Ames, H. M., Echavez, J. F. M., Van de Velde, S., & Muller, A. E. (2022). Automating risk of bias assessment in systematic reviews: A real-time mixed methods comparison of human researchers to a machine learning system. BMC Medical Research Methodology, 22(1), 167. https://doi.org/10.1186/s12874-022-01620-z

Jarke, J. & Heuer, H. (2024). Reassembling the black box of machine learning: Of monsters and the reversibility of foldings. In J. Jarke, B. Prietl, S. Egbert, Y. Boeva, H. Heuer, M. Arnold (Eds.). Algorithmic regimes. Methods, interactions, politics. University Press.

Jiang, L. Y., Liu, X. C., Nejatian, N. P., Nasir-Moin, M., Wang, D., Abidin, A., & Oermann, E. K. (2023). Health system-scale language models are all-purpose prediction engines. Nature, 619(7969), 357–362. https://doi.org/10.1038/s41586-023-06498-1

Kastrup, N., Holst-Kristensen, A. W., & Valentin, J. B. (2024). Landscape and challenges in economic evaluations of artificial intelligence in healthcare: A systematic review of methodology. BMC Digital Health, 2(1), 39. https://doi.org/10.1186/s44229-024-00068-7

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17, 1–9. https://doi.org/10.1186/s12916-019-1426-2

Khanna, N. N., Maindarkar, M. A., Viswanathan, V., Fernandes, J. F. E., Paul, S., Bhagawati, M., & Suri, J. S. (2022). Economics of artificial intelligence in healthcare: Diagnosis vs. treatment. In Healthcare (Vol. 10, No. 12, p. 2493). MDPI. https://doi.org/10.3390/healthcare10122493

Landschaft, A., Antweiler, D., Mackay, S., Kugler, S., Rüping, S., Wrobel, S., & Allende-Cid, H. (2024). Implementation and evaluation of an additional GPT-4-based reviewer in PRISMA-based medical systematic literature reviews. International Journal of Medical Informatics, 189, 105531. https://doi.org/10.1016/j.ijmedinf.2023.105531

Liu, M., Li, S., Yuan, H., Ong, M. E. H., Ning, Y., Xie, F., & Liu, N. (2023). Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques. Artificial Intelligence in Medicine, 142, 102587. https://doi.org/10.1016/j.artmed.2023.102587

López, D. M., Rico-Olarte, C., Blobel, B., & Hullin, C. (2022). Challenges and solutions for transforming health ecosystems in low-and middle-income countries through artificial intelligence. Frontiers in Medicine, 9, 958097. https://doi.org/10.3389/fmed.2022.958097

Luo, X., Chen, F., Zhu, D., Wang, L., Wang, Z., Liu, H., & Chen, Y. (2024). Potential roles of large language models in the production of systematic reviews and meta-analyses. Journal of Medical Internet Research, 26, e56780. https://doi.org/10.2196/56780

Mainz, J., Munch, L., & Bjerring, J. C. (2024). Cost-effectiveness and algorithmic decision-making. AI and Ethics, 1–13. https://doi.org/10.1007/s43681-024-00223-0

Marey, A., Arjmand, P., Alerab, A. D. S., Eslami, M. J., Saad, A. M., Sanchez, N., & Umair, M. (2024). Explainability, transparency and black box challenges of AI in radiology: Impact on patient care in cardiovascular radiology. Egyptian Journal of Radiology and Nuclear Medicine, 55(1), 183.

Mittermaier, M., Raza, M., & Kvedar, J. C. (2023). Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digital Medicine, 6(1), 113.

Morgenstern, J. D., Buajitti, E., O'Neill, M., Piggott, T., Goel, V., Fridman, D., & Rosella, L. C. (2020). Predicting population health with machine learning: A scoping review. BMJ Open, 10(10), e037860. https://doi.org/10.1136/bmjopen-2020-037860

National Institute for Health and Care Excellence [NICE]. (2024). Use of AI in evidence generation: NICE position statement v.1. Retrieved from https://www.nice.org.uk/about/what-we-do/our- research-work/use-of-ai-in-evidence-generation--nice-position-statement

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews, 4, 1-22. https://doi.org/10.1186/2046-4053-4-5

Palaniappan, K., Lin, E. Y. T., & Vogel, S. (2024). Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. Healthcare, 12(5), 562.

Pandey, S., Kaur, R., Teitsson, S., Malcolm, B., Rai, P., Singh, B., & Klijn, S. (2024). EE494 AI- Driven Virtual Assistance Interface for Excel-Based Economic Model. Value in Health, 27(12), S153.

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of Global Health, 9(2).

Ratwani, R. M., Sutton, K., & Galarraga, J. E. (2024). Addressing AI algorithmic bias in health care. JAMA, 332(13), 1051-1052.

Rawlinson, W., Klijn, S., Teitsson, S., Malcolm, B., Gimblett, A., & Reason, T. (2024). P48 Automating Economic Modelling: Potential of Generative AI for Updating Excel-Based Cost- Effectiveness Models. Value in Health, 27(6), S11.

Reason, T., Rawlinson, W., Langham, J., Gimblett, A., Malcolm, B., & Klijn, S. (2024). Artificial Intelligence to Automate Health Economic Modelling: A Case Study to Evaluate the Potential Application of Large Language Models. PharmacoEconomics-Open, 8(2), 191-203.

Robinson, A., Thorne, W., & Wu, B. P. (2023). Bio-sieve: Exploring instruction tuning large language models for systematic review automation. arXiv preprint arXiv:230806610.

Schwendicke, F., Rossi, J. G., Göstemeyer, G., Elhennawy, K., Cantu, A., & Gaudin, R. (2021). Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection. Journal of Dental Research, 100, 369–76.

Sele, D. & Chugunova, M. (2024). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. PLoS One, 19(2), e0298037.

Srivastava, T., Swami, S., & Tong, T. (2024). EE504 Leveraging Large Language Models for Conceptualizing Health Economic Models: A Feasibility Study in Oncology. Value in Health, 27(12), S155.

Sundaram, G. & Berleant, D. (2023). Automating Systematic Literature Reviews with Natural Language Processing and Text Mining: a Systematic Literature Review, in Proceedings of the Eighth International Congress on Information and Communication Technology (ICICT 2023), 1:73-92, Springer, https://doi.org/10.1007/978-981-99-3243-6

Swami, S. & Srivastava, T. (2024). Can Gen-AI Assist in Interpreting the Health Economic Model Results as Per Target Audience? Value in Health, 27(12), S483.

Tufail, A., Kapetanakis, V. V., & Salas-Vega, S. (2016). 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 Technology Assessment, 20(92), 1–72.

Velichkovska, B., Gjoreski, H., Denkovski, D., Kalendar, M., Mullan, I. D., Gichoya, J. W., & Osmani, V. (2023). AI learns racial information from the values of vital signs. medRxiv, 2023- 12.

Vithlani, J., Hawksworth, C., Elvidge, J., Ayiku, L., & Dawoud, D. (2023). Economic evaluations of artificial intelligence-based healthcare interventions: a systematic literature review of best practices in their conduct and reporting. Frontiers in Pharmacology, 14, 1220950.

Voets, M. M., Veltman, J., Slump, C. H., Siesling, S., & Koffijberg, H. (2022). Systematic review of health economic evaluations focused on artificial intelligence in healthcare: the tortoise and the cheetah. Value in Health, 25(3), 340-349.

Wolff, J., Pauling, J., Keck, A., & Baumbach, J. (2020). The economic impact of artificial intelligence in health care: systematic review. Journal of Medical Internet Research, 22(2), e16866.

Xie, Y., Nguyen, Q. D., Hamzah, H., Lim, G., Bellemo, V., & Gunasekeran, D. V. (2020). Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. The Lancet Digital Health, 2(5), e240–9.

Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of artificial intelligence applications in real-life clinical practice: systematic review. Journal of Medical Internet Research, 23(4), e25759.

Zemplényi, A., Tachkov, K., Balkanyi, L., Németh, B., Petykó, Z. I., Petrova, G., Czech, M., Dawoud, D., Goettsch, W., Gutierrez Ibarluzea, I., Hren, R., Knies, S., Lorenzovici, L., Maravic, Z., Piniazhko, O., Savova, A., Manova, M., Tesar, T., Zerovnik, S., & Kaló, Z. (2023). Recommendations to overcome barriers to the use of artificial intelligence-driven evidence in health technology assessment. Frontiers in Public Health, 11, 1088121. https://doi.org/10.3389/fpubh.2023.1088121.