A Systematic Review of the Methods and Quality of Economic Evaluations for AI-assisted Cancer Screening or Diagnosis

EE35

Yuyanzi Zhang[1], Lei Wang[1], Yifang Liang[1], Annushiah Vasan Thakumar[2], Hongfei Hu[1], Yan Li[1], Aixia Ma[1,3], Hongchao Li[1,3], Luying Wang[1,3]

[1] School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing, China

[2] School of Pharmacy, Taylor's University, Subang Jaya, Malaysia

[3] Center for Pharmacoeconomics and Outcomes Research, China Pharmaceutical University, Nanjing, China

INTRODUCTION

- Artificial intelligence (AI) is rapidly transforming medical screening and diagnosis, primarily using image recognition to enhance the accuracy of clinical decisions for various cancers.
- Health economic evaluations (HEEs) of AI face distinct challenges, including complex cost measurement, uncertain population generalizability, and the dynamic evolution of AI performance over time.
- The **CHEERS-AI** checklist was established in 2024 to standardize reporting, yet no prior studies have adopted it.

OBJECTIVE

- This study aims to systematically review the **methods** and **reporting quality** of HEEs for AI-assisted cancer screening or diagnosis, with a secondary focus on summarizing health and economic outcomes.

METHOD

- The review protocol was registered with PROSPERO (Registration number: CRD42024625408) and conducted in accordance with PRISMA 2020 guidelines.
- Systematic Search: Five databases (Medline, Embase, Web of Science, Cochrane Library, International HTA Database) were searched from inception to December 2024.
- Eligibility: Studies were selected per PICOS framework, focusing on HEEs comparing AI-assisted versus conventional cancer screening/diagnosis.
- Study Selection & Data Extraction: Conducted independently by two reviewers; discrepancies were resolved by a third reviewer.
- Quality Assessment: Independently appraised using the **CHEERS-AI** and **Philips checklists**.
- Data Synthesis: A descriptive analysis was conducted, with findings presented in narrative summaries, tables, and figures.

RESULTS

- The systematic search yielded 2,564 records. After duplicate removal and screening, 17 studies met the inclusion criteria and were included for analysis. The selection process followed the PRISMA statement.

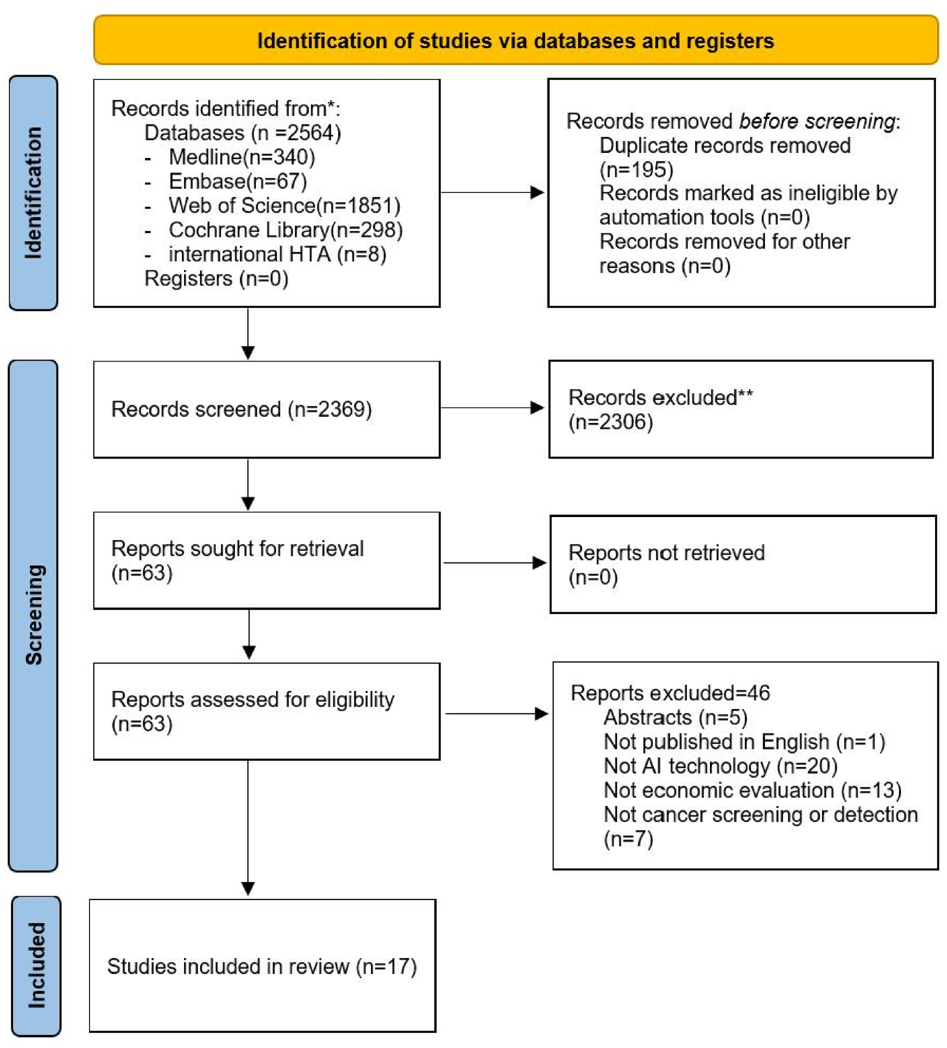

Fig1 PRISMA (2020) flow diagram.

Main Characteristics

- The 17 included studies were conducted across 9 countries, with the majority (15/17) published from 2022 to 2024.
- Studies evaluated AI-assisted strategies for eight cancer types.
- Cost-utility analysis was the predominant method, with QALYs serving as the primary outcome measure in most studies.

AI Interventions

- AI enhanced screening/diagnosis accuracy through image recognition.
- Mostly served as diagnostic support tools; rarely used for risk prediction.
- Enabled real-time guidance in colonoscopy, boosting adenoma detection.

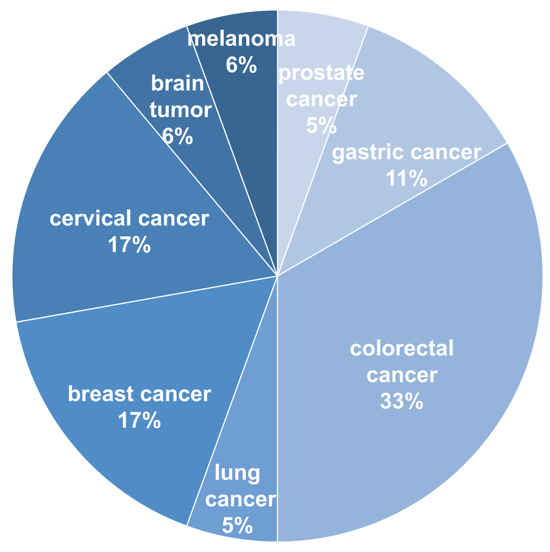

Fig2 Distribution of Included Studies by Cancer Type

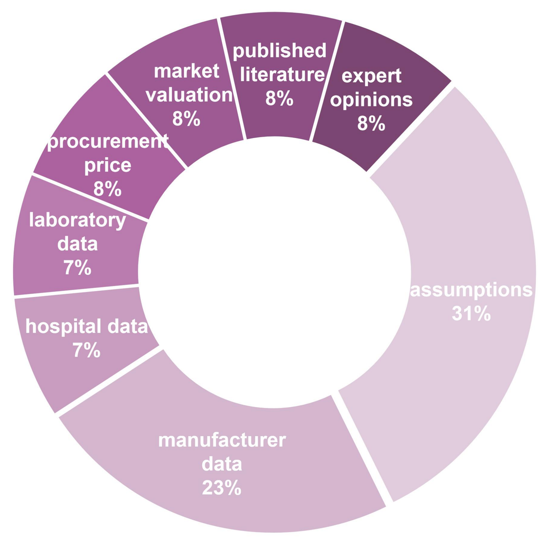

Fig3 Reported Sources of AI Costs

AI Costs

- AI cost sources were diverse but poorly reported.
- Few studies detailed cost components (setup, maintenance, training).

Model & Sensitivity Analysis

- Markov models predominated, with limited use of DES model.
- AI cost and accuracy (sensitivity/specificity) were the most influential parameters in DSA.

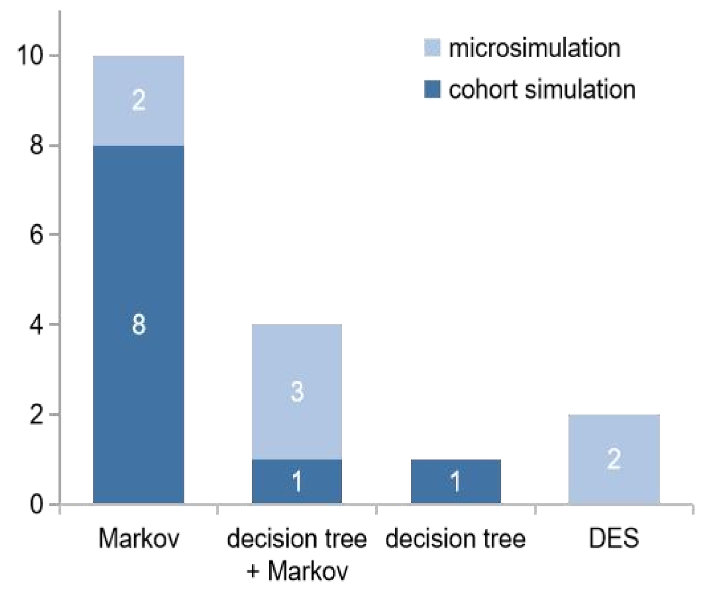

Fig4 Proportion of Studies by Model Type

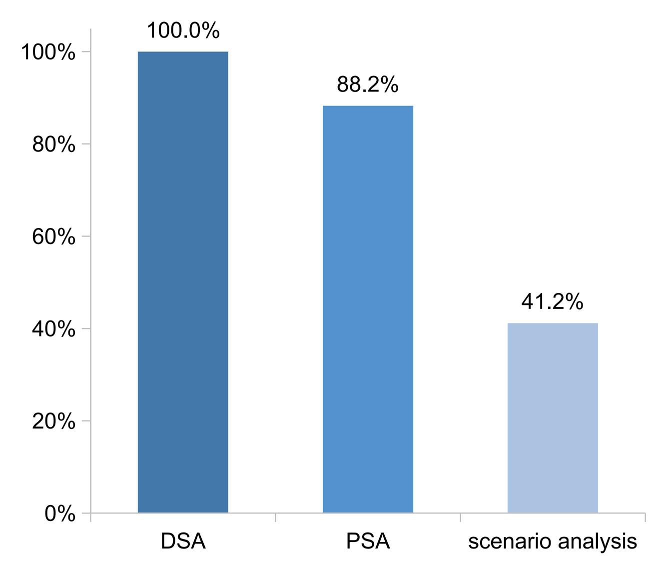

Fig5 Proportion of Sensitivity Analyses Applied

Results of HEEs

- AI improved health outcomes (QALYs, LYs) through enhanced diagnostic accuracy.
- Most studies found AI reduced overall costs through early detection and treatment.
- Most studies assumed full clinician adherence, overlooking compliance issues.

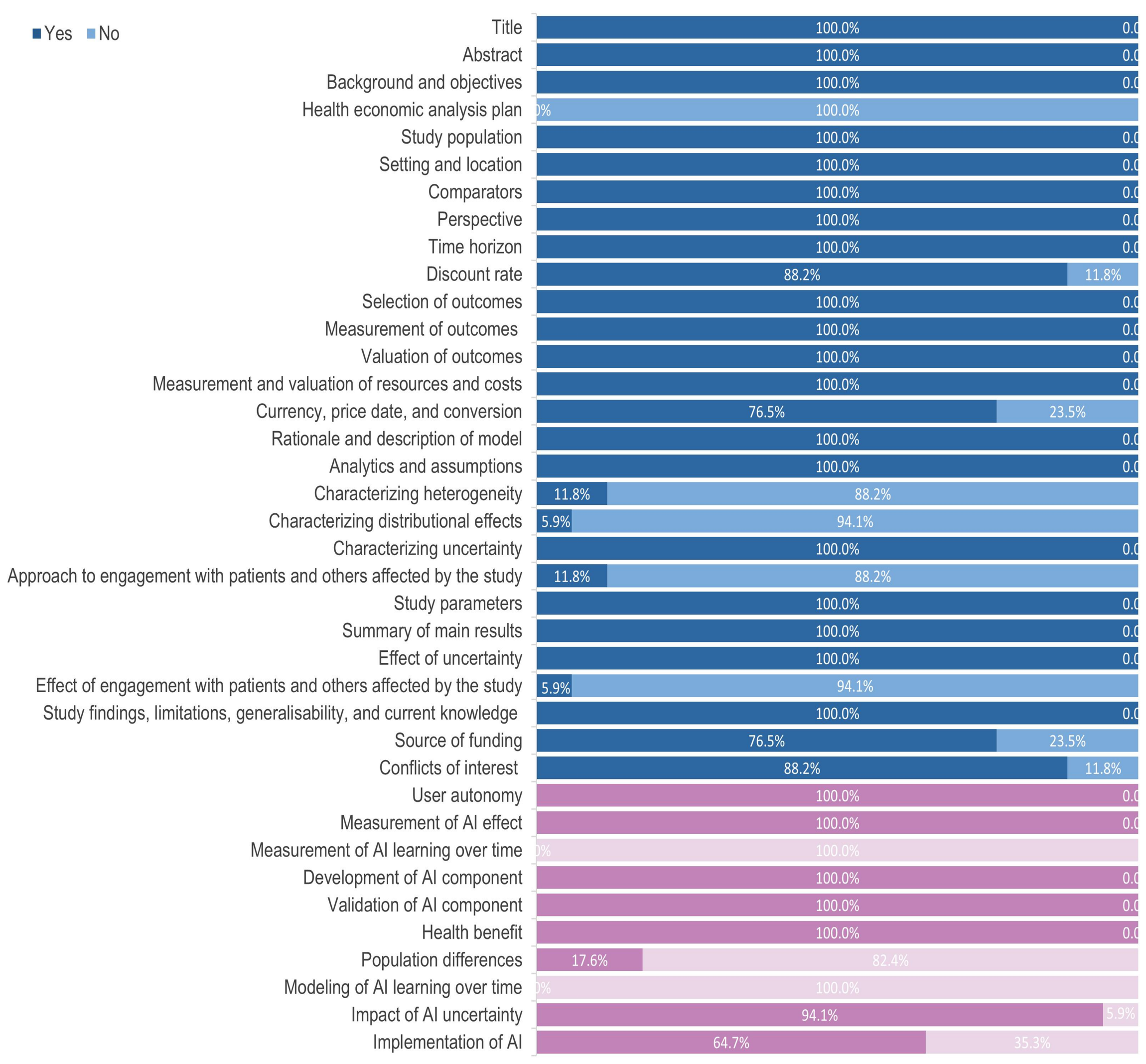

Fig6 Overview of the proportion of studies reporting CHEERS-AI checklist items (The purple bars indicate AI-specific items)

Quality Assessment

- CHEERS-AI:Critical underreporting of AI-specific items was observed, with **learning over time** (AI3/AI8) and **population differences** (AI7) being particularly neglected.
- Philips Checklist: half-cycle correction and data quality assessment.

DISCUSSION

- **Static AI Performance Modeling**: Current models treat AI performance as static, ignoring its potential for improvement through continuous learning. This likely underestimates long-term effectiveness and cost-effectiveness.
- **Inadequate Reporting of AI Costs**: AI cost reporting—especially for implementation and maintenance—is often incomplete and assumption-based, compromising the reliability and transparency of economic conclusions.
- **Limited Population Generalizability**: The diagnostic performance of AI models is highly dependent on their training data, yet many tools are trained on non-representative populations, limiting the applicability of HEE findings across diverse racial, ethnic, and socioeconomic groups.

CONCLUSIONS

This review reveals methodological and reporting limitations in current HEEs of AI-assisted cancer screening. Incomplete reporting of AI-specific details compromises the reliability of cost-effectiveness findings, underscoring the need for more comprehensive and transparent practices in future studies.

CONTACT INFORMATION

Zhang Yuyanzi, zhangyuyanzi@stu.cpu.edu.cn

Luying Wang, luying_w@cpu.edu.cn